IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 21, 2021, accepted March 6, 2021, date of publication March 10, 2021, date of current version March 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064962

Rolling Bearing Sub-Health Recognition via Extreme Learning Machine Based on Deep Belief Network Optimized by Improved Fireworks

HAO LUO, CHAO HE, JIANING ZHOU, AND LI ZHANG
College of Information, Liaoning University, Shenyang 110036, China

Corresponding author: Li Zhang (zhang_li@lnu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant No. 5170413 and in part by the Science and Technology Research Projects of Educational Department of Liaoning Province under Grant No. LQN201910.

ABSTRACT Rolling bearings, as the main components of the large industrial rotating equipment, usually work under complex conditions and are prone to break down. It can provide a certain theoretical basis for identifying the sub-health state of the industrial equipment by the analysis from the incipient weak signals. Thus, a sub-health recognition offline algorithm based on Refined Composite Multiscale Dispersion Entropy (RCMDE) and Deep Belief Network-Extreme Learning Machine (DBN-ELM) optimized by Improved Firework Algorithm (IFWA) is proposed. First of all, in light of the drawbacks that it is easy to fall into local optima and cross the boundary for exploding fireworks in Firework Algorithm (FWA), Cauchy mutation and adaptive dynamic explosion radius factor coefficient is introduced into IFWA. Secondly, Maximum Correlation Kurtosis Deconvolution (MCKD) optimized by the improved parameters is used to process the incipient vibration signals with nonlinearity, nonstationary, and IFWA is used to adaptively adjust to the period T and the filter length L in MCKD(IFWA-MCKD). Then, each sequence of signals is further extracted the feature—RCMDE to rich sample diversity. Finally, combining the powerful unsupervised learning capability from DBN and the generalization capability from ELM, DBN-ELM can be established. What's more, in order to avoid the interference of human on the parameters, IFWA is used to optimize the number of hidden nodes in DBN-ELM, and the IFWA-DBN-ELM is established. It shows that the algorithm has the higher sub-health recognition accuracy, better robustness and generalization, which has a better industrial application prospect.

INDEX TERMS Sub-health recognition, DBN, MCKD, RCMDE, improved firework algorithm.

I. INTRODUCTION

In recent years, with the development of large-scale industrial equipment towards automation and intelligence, sub-health monitoring for the large-scale equipment has attracted much attention. ''Sub-health'' is a diseased state which is an incipient state with minor fault. In this state, mechanical equipment can still run with illness. However, if it isn't detected in time, it may bring out large-scale malfunctioning, resulting in making mechanical equipment shut down, which can not only reduce efficiency, but also seriously threaten the lives of workers. Therefore, it is necessary to monitor whether the mechanical equipment falls into the sub-healthy state [1]. It shows that about 30% of the industrial equipment malfunctioning can be caused by the rolling bearings. Obviously, incipient fault diagnosis for the rolling bearings can provide an indispensable gist for the equipment sub-health status identification.

Incipient fault vibration signals of the rolling bearings have the characteristics with weak, non-linearity, non-stationary, low SNR, micro-amplitude, high dimension. Therefore, in terms of the signal processing in fault diagnosis, the accelerometers are usually used to measure the acceleration as the vibration signals. Enhanced Singular Spectrum Decomposition [2] (ESSD), Spectral Theory of Multidimensional Matrix [3] (STMM), Variational Mode Decomposition [4] (VMD) and others are used to increase the pulse numbers. Although the previous signal denoising algorithms have achieved good results, there are also some disadvantages. In VMD, if the signals decompose too many, it will be intermittent and lack of regularity. Unfortunately, MED can only deconvolve a single pulse or a group of pulses and can't deal

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

with the necessary periodic pulses that occur repeatedly during fault state. Luckily, McDonald *et al.* proposed Maximum Correlation Kurtosis Deconvolution [5]. MCKD is used the fault periodicity and the pulse-like vibration associated with the most faults to process the signal, which can not only improve performance and enhance periodic pulses, but also perform concurrent detection. However, the deconvolution effect by MCKD is severely interfered by the period T, the filter length L and the shift K. Any improper parameter can be difficult to achieve the ideal deconvolution effect.

After obtaining the desired deconvolution effect, aiming at the difficulty that how to extract the fault features further, some scholars introduce the nonlinear dynamic analysis method into the field of fault diagnosis. Dispersion Entropy [6] (DE)which overcame the weaknesses of Sample Entropy [7] (SampEn) that had the high time complexity and was easy to mutate for similarity measure as well as Permutation Entropy [8] (PE) without considering the interaction of signal amplitude, had an excellent information extraction capability. On this basis, Azami *et al.* proposed Refined Composite Multiscale Dispersion Entropy (RCMDE). The feature has stronger robustness and can reflect the complexity change for the time series signals. It has achieved good effect in the biomedical signal analysis [9] and the fault diagnosis by shallow networks such as Support Vector Machine (SVM) [10].

Many scholars applied SVM, GAN, AE, LSTM and so on to the fault detection [11]–[13]. However, the incipient signals are complex and changeable, and the models are a little inadequate. At the same time, we notice that DBN with strong unsupervised feature learning capability not only has a better generalization and mapping capability, making it to fit complex nonlinear data well, but also is a non-parametric data-driven model, without making assumptions about data generation. ELM has a good generalization and the idea of kernel or sigmoid function can be used to map nonlinear samples to higher dimensional space, which shows unique advantages in classification. [14]. DBN and ELM have achieved good results in the sub-health identification respectively [15]–[17].

In view of ELM, DBN-ELM is established. DBN-ELM has been successful in Intrusion Detection, Soft Measurement and short-term Energy Consumption Prediction. Dai *et al.* added many classifiers at each intermediate level in DBN and compared with the classification effect through ELM to achieve classification by majority voting and comprehensive calculation. The improved DBN-ELM achieved good results on the NSL-KDD intrusion detection data set [18]. Later, Wang *et al.* introduced DBN-ELM into the field of the soft measurement, and it showed that this method had higher comprehensive measurement accuracy in nutrient liquid composition for the soilless cultivated crop [19]. At the same time, Zhang *et al.* proposed a DBN-ELM short-term prediction based on Cyclic Feature (CF) about building energy consumption, which improved the prediction accuracy of the building energy consumption [20].

In this paper, DBN-ELM is introduced into the field of the sub-health recognition. In order to adjust the weights and bias of each layers adaptively in DBN, the softmax layer is added at the end of the RBM_3 and Adam [21] is used for back propagation. Then, ELM is used to complete classification. However, there is still a problem that DBN-ELM performance is restricted by the number of neurons in the hidden layers. Excessive neurons will cause the model to be bloated, increasing the training time and reducing the efficiency. Nevertheless, too little can't achieve the desired the sub-health recognition effect.

As an emerging computing technology, swarm intelligence algorithm has attracted much attention due to its simplicity and global search capability [22]. Shao *et al.* proposed a method based on combining Particle Swarm Optimization (PSO) to optimize the number of the hidden layer neurons, learning rate and the momentum in DBN, which was applied to analyze experimental signal of the rolling bearings, and PSO-DBN could recognize fault state more intuitively and accurately [23]. Du *et al.* proposed a landslide displacement prediction method based on EEMD and optimizing connection weight and threshold of ELM by PSO, and it showed that PSO-ELM could predict landslide displacements with stepped curves effectively in this area [24].

Based on the related researches from domestic and foreign countries, this paper proposes a new improved fireworks algorithm to optimize DBN-ELM sub-health identification model. Firstly, inspired by the swarm algorithm, IFWA that is introduced Cauchy variation and adaptive dynamic explosion radius is applied to adaptively select the parameters T and L in MCKD called IFWA-MCKD. After that, extract the feature RCMDE further. Secondly, ELM is used as a classifier to improve the model's generalization performance. DBN-ELM is introduced into the field of the sub-health identification. IFWA is used to optimize the number of the hidden nodes in DBN-ELM. Finally, instead of single feature, RCMDE, the processed signals by IFWA-MCKD and the waveform factor, pulse factor, gravity frequency, mean square frequency, frequency variance and other 16 kinds of time-frequency domain features constitute multi-feature data set. Train the DBN-ELM and classify the testing set, and verify the generalization capability under different loads.

II. IMPROVE FIREWORKS ALGORITHM

In order to maximize the performance for MCKD and DBN-ELM, IFWA which can adaptively adjust the hyperparameters of the models is introduced.

A. FIREWORKS ALGORITHM

Ying Tan *et al.* proposed FWA [25]. The explosive fireworks at k dimensions are generated, and the fitness values are calculated, where the explosion radius R_i and the explosive fireworks number S_i are defined, as shown in Eq.1∼2:

$$R_i = Er \times \frac{f(x_i) - y_{min} + c}{\sum_{i=1}^{N} (f(x_i) - y_{min}) + c} \tag{1}$$

$$S_i = En \times \frac{y_{max} - f(x_i) + eps}{\sum_{i=1}^{N}(y_{max} - f(x_i)) + eps} \tag{2}$$

Er, En are respectively defined as explosion radius factor and number factor, and eps is constant. $f(x_i)$ is fitness function. In order to avoid the overwhelming influence from explosive fireworks, constraints are added as shown in Eq.3 to Eq.2:

$$s_i^{'} = \begin{cases} round(p * En), & S_i < pEn \\ round(q * En), & S_i > pEn, p < q < 1 \\ round(S_i), & otherwise \end{cases} \tag{3}$$

Among them, p and q are constants. Firework x_i is randomly selected from the fireworks populations to perform Gaussian variation, and the variation firework $x_{ik}^{'}$ is obtained, as shown in Eq.4:

$$x_{ik}^{'} = x_{ik} + (x_{min,k} - x_{ik}) \times e \tag{4}$$

$x_{min,k}$ is the minimum firework position, $e \sim N(0, 1)$. The newly generated fireworks may cross the boundary. If x_i exceeds the range $[L_k, U_k]$, the firework x_i needs to be mapped to a new position, as shown in Eq.5:

$$x_{ik} = x_{l,k} + rand(0, 1) \cdot (x_{U,k} - x_{L,k}) \tag{5}$$

$x_{U,k}, x_{L,k}$ are the upper boundary and lower boundary.

After explosion, select x_i with the smallest fitness value in the current population to inherit to the next generation, and then the roulette strategy is used to select other x_i to the next generation population. If there are many candidate fireworks around x_i, reduce the probability to select x_i. When the next generation of population is formed, calculate the fitness values, and perform mutation and selection strategies in turn until the iteration ends or the specified threshold is reached.

B. IMPROVE FIREWORKS ALGORITHM

The explosion radius can balance the local and global search capability. On one hand, a worse firework produces a larger explosion radius to enhance the algorithm's exploration capability and breaks out of the local optimal. On the other hand, a better firework produces a smaller explosion radius to enhance the algorithm's mining and the local search capability. The size of explosion radius R_i is related to the radius factor Er, but FWA adopts the fixed Er, which makes the algorithm easily fall into local optimal. Therefore, a new dynamic adaptive explosion radius function is proposed, as shown in Eq.6:

$$Er = A_\mu + 0.6A_\mu \sin\left[\frac{\pi \cdot (2t - T_{max})}{2 \cdot T_{max}} + \pi\right] \tag{6}$$

where, t is the current iteration number and T_{max} is the maximum iteration. In Eq.6, as the iteration increases, the explosion radius changes from large to small, and IFWA algorithm firstly has a strong global search capability and then has a strong local search capability. Where A_μ is the explosive radius factor coefficient that is related to the feasible region $\emptyset$.

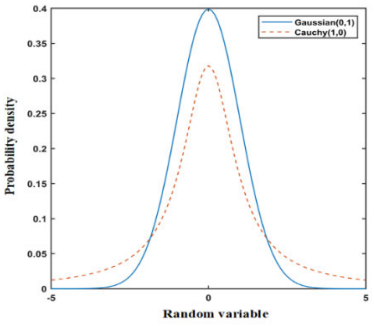

FIGURE 1. Standard cauchy and gaussian probability density functions.

Inspired by the Literature [26], the setting rules for A_μ are as shown in Eq.7∼9:

$$A_{max} = (U_k - L_k) * 0.02 \tag{7}$$

$$A_{min} = (U_k - L_k) * 0.005 \tag{8}$$

$$A_\mu = \frac{A_{max} + A_{min}}{2} \tag{9}$$

A_{max} and A_{min} are the upper and lower limits for Er respectively. With iterations, Er can traversal this region dynamically and randomly, so as to prevent not only too fast convergence to fall into the local minimum value, but also too slow convergence to search for the local optimal value.

Inspired that Cauchy variation can produce more extensive random numbers, it is introduced into IFWA. Cauchy mutation operator and Gaussian mutation operator are used alternatively under certain conditions. The standard Cauchy and Gaussian probability density functions are shown in Fig.1. Compared to the Gaussian distribution, the Cauchy distribution is high at both ends and low in the middle. As a mutation operator, Cauchy mutation operator can generate a wider range of random numbers. In the incipient iteration, through Cauchy mutation, it can produce more extensive variation fireworks, so that the fireworks escape from the local optimum. In the later stage, the Gaussian mutation is used to search the global optimum precisely near the optimal area.

The construction method is shown in Eq.10, instead of Eq.4:

$$x_{ik}^{'} = \begin{cases} x_{ik} + (x_{min,k} - x_{ik}) \times C(1, 0), & rand \le 1 - \frac{t}{T_{max}} \\ x_{ik} + (x_{min,k} - x_{ik}) \times N(0, 1), & otherwise \end{cases} \tag{10}$$

Among them, $rand$ is a random number between (0,1), $C(1, 0)$ is a standard Cauchy distribution random number, and $N(0, 1)$ is a standard Gaussian distribution random number.

C. IFWA PERFORMANCE EVALUATION

In order to verify the IFWA algorithm performance, GA, PSO, FWA and IFWA are used to carry out comparative simulation experiments on the test functions shown in Table 1.

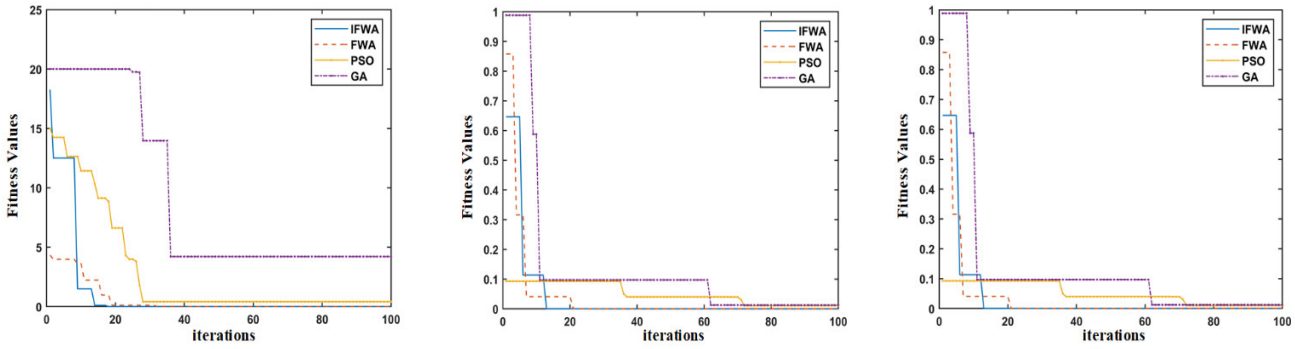

FIGURE 2. Fitness curves of the first three functions.

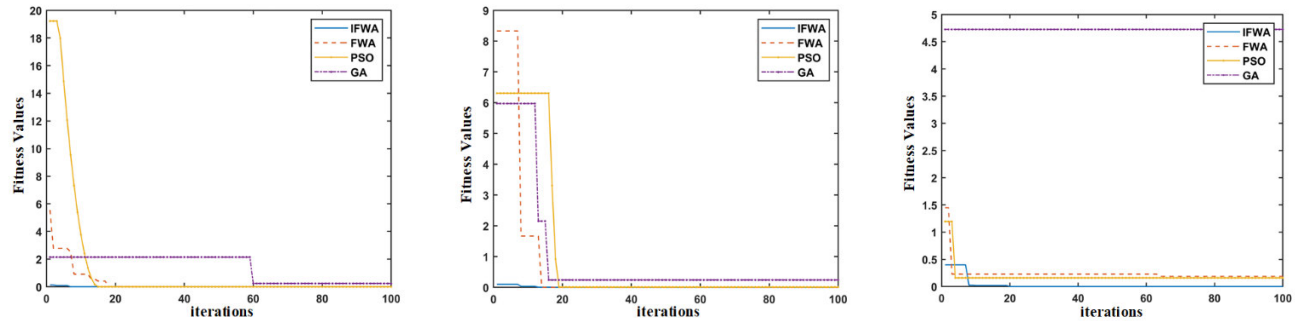

FIGURE 3. Fitness curves of the last three functions.

TABLE 1. Test functions.

Test functions	Range	f	P
$f_1(x) = -ae^{-b\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n} cx_i} + a + e$	$[-32.768, 32.768]$	0	M
$f_2(x) = 10d + \sum_{i=1}^{d} x_i^2 - 10cos(2\pi x_i)$	$[-5.12, 5.12]$	0	M
$f_3(x) = \frac{1}{4000}\sum_{i=1}^{d} x_i^2 - \prod_{i=1}^{d} cos\frac{x_i}{\sqrt{i}} + 1$	$[-600, 600]$	0	M
$f_4(x) = \sum_{i=1}^{d} ix_i^2$	$[-5.12, 5.12]$	0	S
$f_5(x) = \sum_{i=1}^{d} x_i^2$	$[-5.12, 5.12]$	0	S
$f_6(x) = 418.9829d - \sum_{i=1}^{d} x_i sin(\sqrt{\lvert x_i \rvert})$	$[-500, 500]$	0	M

The algorithms' parameters are set as follows:

The population size is 30. The maximum iteration is 100, and the dimension is five. Variation spark number is 5. p =0.3, q =0.6, En =5. Besides, Er =6 in traditional FWA; After running the four algorithms at 20 times respectively, the results of the test functions obtained show in Fig.2∼3.

Instruction: f: Optional, P: Peak, M: Multiple, S: Single.

From Fig.2∼3, it can be seen that the overall optimization capability in GA is poorest, and the optimization accuracy is extremely low for most test functions. Compared with IFWA, in PSO, the optimization accuracy is lower and convergence speed is slower. However, IFWA adopts the adaptive dynamic explosion radius coefficient. At the same time, the alternating strategy of Cauchy mutation operator and Gaussian mutation operator is adopted to enhance the optimal performance. As you can see that IFWA has the highest optimization accuracy in multimodal functions such as $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_6(x)$. For the single functions such as $f_4(x)$, $f_5(x)$, the purpose is to compare the convergence speed. It can be seen that IFWA is the earliest to achieve convergence so that the convergence speed is faster.

The average value of the functions can reflect the optimization capability for algorithms, and the median can reflect the distribution for algorithms. It can be seen from Table 2 that the average accuracy in IFWA is higher, so IFWA has stronger stability and optimization capability. Similarly, the median for IFWA also has highest accuracy, indicating that the optimal results distribute more compactly, and local and global search is more balanced.

Through the above analysis, it is fully demonstrated that IFWA has better optimization accuracy and convergence speed. And it can optimize the hyperparameters of algorithms. Of course, IFWA can maximize the signal processing effect by MCKD and the performance in DBN-ELM

TABLE 2. Mean and median values of test functions.

Test functions		IFWA	FWA	PSO	GA
$f_1(x)$	Mean	4.5918e-13	3.3291e-08	0.4513	4.8546
	Median	1.1546e-14	7.7583e-10	0.5446	4.2120
$f_2(x)$	Mean	1.7763e-15	1.3500e-13	3.3701e-05	0.1552
	Median	0	1.7763e-15	8.5353e-06	0.3998
$f_3(x)$	Mean	2.2204e-16	5.5511e-16	0.0238	0.0125
	Median	0	0	0.0017	0.0125
$f_4(x)$	Mean	2.7401e-19	1.3052e-16	4.4214e-05	0.0001
	Median	4.2471e-23	3.0821e-17	3.2008e-07	2.5049e-05
$f_5(x)$	Mean	1.2334e-29	1.1550e-20	2.2595e-06	0.0006
	Median	5.1591e-32	2.8974e-25	1.3429e-07	2.5049e-05
$f_6(x)$	Mean	1.3328e-05	0.0980	0.0077	2.7145
	Median	1.6581e-05	0.0053	0.0107	0.0869

sub-health recognition by looking for the best parameters, so as to obtain a higher recognition accuracy.

III. IFWA-MCKD

A. MAXIMUM CORRELATION KURTOSIS DECONVOLUTION

Look for the FIR filter and the real signal $\boldsymbol{p}$, as shown in Eq. 11:

$$p = f * y = \sum_{i=1}^{L} f_i y_{n-i+1} \tag{11}$$

$\boldsymbol{y}$ is the actual signal. $\boldsymbol{f}$ is the filter coefficient. L is filter length. n is the signal length, and $*$ is the convolution. Then $K-shift$ MCKD is shown in Eq.12:

$$\max_{f} CK_K(T) = \max_{f} \frac{\sum_{n=1}^{N} \left(\Pi_{k=1}^{K} p_{n-KT}\right)^2}{\left(\sum_{n=1}^{N} p_n^2\right)^{K+1}} \tag{12}$$

p_n is the evaluation signal. T is the deconvolution period, and K is the shift number. By combining Eq.11∼12, $dCK_K(T)/dfi$, the optimal $\boldsymbol{f}$ can be obtained as shown in Eq.13:

$$\boldsymbol{f} = \frac{\|p\|^2}{2\|\beta\|^2} \left(Y_0 Y_0^T\right)^{-1} \sum_{k=0}^{K} Y_{kT} a_k \tag{13}$$

The parameters in Eq.13 are shown in Eq.14∼17:

$$p = Y_0^T f \tag{14}$$

$$Y_q = \begin{bmatrix} y_{1-q} & y_{2-q} & \cdots & y_{N-q} \\ 0 & y_{1-q} & \cdots & y_{N-1-q} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & y_{N-L-1-q} \end{bmatrix}_{L\times N} \quad (q = 0, T, 2T, \ldots KT) \tag{15}$$

$$\alpha_k = \begin{bmatrix} p_{1-kT}^{-1}(p_1^2 p_{1-T}^2 \cdots p_{1-KT}^2) \\ p_{2-kT}^{-1}(p_2^2 p_{2-T}^2 \cdots p_{2-KT}^2) \\ \vdots \\ p_{N-kT}^{-1}(p_N^2 p_{N-T}^2 \cdots p_{N-KT}^2) \end{bmatrix}_{N\times 1} \tag{16}$$

$$\beta = \begin{bmatrix} p_1 p_{1-T} \cdots p_{1-KT} \\ p_2 p_{2-T} \cdots p_{2-KT} \\ \vdots \\ p_N p_{N-T} \cdots p_{N-KT} \end{bmatrix}_{N\times 1} \tag{17}$$

B. IFWA-MCKD

The sub-health identification needs to monitor w eak fault signals. MCKD which considers the impact signal continuity can enhance the correlation kurtosis of the original incipient small impact signals and the pulse numbers.

The deconvolution effect of MCKD is seriously affected by T and L [27] and restricted by K.

On one hand, T is the interval between two endpoints for processing signals. T and L influence each other. If only one parameter is considered unilaterally, the interaction between the two parameters will be ignored, causing that we only obtain the relative optimal parameters and can't obtain the overall optimal deconvolution signals. Eq.18 in Literature [28] is introduced to calculate the approximate T. However, $[T, L]$ adaptive optimization strategy is not mentioned in Literature [28].

$$T = \left| \frac{f_{\text{sample frequency}}}{f_{\text{characteristic frequency}}} \right| \tag{18}$$

On the other hand, the larger K, the more the signal pulse numbers, but the larger K will decrease the accuracy in MCKD. Its interaction with T and L isn't obvious.

TABLE 3. IFWA-MCKD algorithm steps.

IFWA-MCKD steps:
Step1: Obtain vibration data and calculate sampling frequency and characteristic frequency.
Step2: Set IFWA parameters such as population size N and the maximum iteration period T_{max}.
Step3: Calculate approximate the filtering cycle T according to Eq.9, and set the filtering cycle error T', $[T-T', T+T']$ as the optimization constraint interval for the fireworks x_{i1};
Step4: Set Lϵ[100,600] as the fireworks x_{i2} optimization constraint interval.
Step5: Set the shift number K=1. According to x_{ik} initial population $[T_i, L_i]$, process the signal by MCKD and calculate the signal envelope spectrum entropy after denoising according Eq.19.
Step6: With the minimum envelope spectrum entropy value as the optimal goal, perform mutation and selection operations on the initial population in IFWA and update the global optimal firework's position.
Step7: Repeat Step5~6 until the maximum iteration period or specified threshold, and output the best parameters $[T_0, L_0]$ and the corresponding envelope entropy spectrum value.

In order to solve the above problems, we propose IFWA-MCKD algorithm which can select parameter T and L dynamically and adaptively. After that, K is determinated by experiments. IFWA-MCKD algorithm explores the strategy of parameter interval optimization and avoids the human factor interference.

C. [T, L] OPTIMIZATION STRATEGY

Envelope spectrum entropy can reflect the uniform and periodic impact about the incipient weak fault vibration signals, which has achieved good results in fault identification [29], [30]. Therefore, envelope spectrum entropy is introduced as the fitness function in IFWA-MCKD algorithm. The lower value is, the more uniform signal is and the higher SNR is, which it is proved that the deconvolution effect is better. Envelope spectrum entropy is shown in Eq.19:

$$\begin{cases} E_p = -\sum_{j=1}^{N} p_j lg p_j \\ p_j = a(j) / \sum_{j=1}^{N} a(j) \end{cases} \tag{19}$$

p_j is the normalized form of $a(j)$; $a(j)$ is the envelope signal obtained by Hilbert Demodulation for $\boldsymbol{p}$.

The specific steps in IFWA-MCKD algorithm are shown in Table 3:

D. K OPTIMIZATION STRATEGY

According to Literature [5], K should be no more than 10. Otherwise the signal accuracy will be decreased, so we select $1 \leq K \leq 10$.

It is noticed that the greater kurtosis value, the richer pulse information contained which can reflect the fault characteristic information in signals are more concentrated [31]. Therefore, the kurtosis is selected as the evaluation index for the deconvolution effect of the processed signal under the shift number K. Calculate the kurtosis and the K corresponding to the maximum kurtosis is taken as the optimal shift number.

IV. REFINED COMPOSITE MULTISCALE DISPERSION ENTROPY

In order to further extract the fault features, consider the complex relationship between signals fully.

RCMDE makes further improvements to MDE. MDE doesn't consider the relationship between the data fully after segmentation during segmenting, resulting in missing some statistical information, and because of the different initial point positions, there is a certain deviation about results obtained. RCMDE overcomes above problems. Under most conditions, the entropy standard deviation is smaller, indicating that the algorithm has better stability and reliability.

The standard normal distribution function is used to map the signal x_j processed by MCKD into $[0, 1]$, and y_i is obtained as shown in Eq.20:

$$y_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt \tag{20}$$

where μ, σ is the mean value and variance of x_j. y_i is mapped to $[1, c]$ through linear transformation, as shown in Eq.21:

$$z_i^c = round\,(c \cdot y_i + 0.5) \tag{21}$$

Reconstruct the z^c phase space, and the embedding vector $z_i^{m,c}$ is shown in Eq.22:

$$z_i^{m,c} = \left\{ z_i^c, z_{i+d}^c, \ldots, z_{i+(m-1)d}^c \right\} \tag{22}$$

$i = 1, 2 \ldots, N-(m-1)\,d$, m is the embedding dimension. d is the time delay, and each $z_i^{m,c}$ is mapped to a scattered pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$, where $z_i^c = v_0, z_{i+d}^c = v_1, \ldots, z_{i+(m-1)d}^c = v_{m-1}$, the number of scatter patterns is c^m. Calculate the probability p in each dispersion mode $\pi_{v_0 v_1 \ldots v_{m-1}}$, as in Eq.23:

$$p = \frac{Num(\pi_{v_0 v_1 \ldots v_{m-1}})}{N-(m-1)\,d} \tag{23}$$

Among them, Num is the number of times that $z_i^{m,c}$ corresponds to $\pi_{v_0 v_1 \ldots v_{m-1}}$; the DE of the reconstructed signal x_i is shown in Eq.24:

$$DE\,(x, m, c, d) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 v_1 \ldots v_{m-1}}) \cdot ln p(\pi_{v_0 v_1 \ldots v_{m-1}}) \tag{24}$$

The RCMDE is shown in Eq.25 in each scale factor τ:

$$RCMDE\,(x, m, c, d, \tau) = \frac{1}{\tau} \sum_{i=1}^{\tau} DE\left(x^{\tau}, m, c, d\right) \tag{25}$$

There are four parameters that should be set manually in RCMDE: m, c, d, τ_{max}.

V. IFWA-DBN-ELM SUB-HEALTH RECOGNITION ALGORITHM

A. DBN

DBN is composed of multiple Restricted Boltzmann Machines (RBM), and the DBN is shown in Fig.4(a).

For the given state v, h, the joint distribution probability is specified by the energy function, as shown in Eq.26:

$$P(v, h) = \frac{1}{Z} e^{-E(v,h)} \tag{26}$$

The energy function and the normalization constant Z are shown by Eq.27∼28:

$$E(v, h) = -b^T v - c^T h - v^T W h \tag{27}$$

$$Z = \sum_{v} \sum_{h} e^{-E(v,h)} \tag{28}$$

Among them, b and h are biases, and W is the weight between layers.

B. ELM

ELM is shown in Fig.4(b). For samples (x_j, t_j), the hidden layer that has h_4 nodes of ELM, which can be expressed as Eq.29:

$$\sum_{i=1}^{h_4} \beta_i g\left(W_i \cdot x_j + b_i\right) = t_j \tag{29}$$

$g(x)$ is the activation function, W_i and β_i are the input and output weights of the hidden layer, b_i is offset, and $W_i \cdot x_j$ is the inner product. Once W_i and b_i are randomly determined, the hidden layer output matrix H is uniquely determined. The training can be transformed into Eq.30:

$$H \cdot \beta = T \tag{30}$$

where, T is the expected output matrix. The output weights $\acute{\beta}$ as shown in Eq.31:

$$\acute{\beta} = H^+ \cdot T \tag{31}$$

where, H^+ is the pseudo inverse generalized inverse of the matrix H.

C. IFWA-DBN-ELM SUB-HEALTH RECOGNITION ALGORITHM

Establish IFWA-DBN-ELM sub-health identification model, as shown in Fig.4:

As shown in Fig.4, the firework $x_i(x_{i1}, x_{i2}, x_{i3}, x_{i4})$ can be the neurons number in $h_1 \sim h_4$ respectively. DBN is stacked by three RBMs. The visual layer v is used to input the training set, and the neuron number is equal to the dimension of the training set. The hidden layer $h_1 \sim h_3$ is used to extract characteristics, and the training rule in RBM is shown in Eq.32. In the third RBM, h_3 is used as the input layer for ELM.

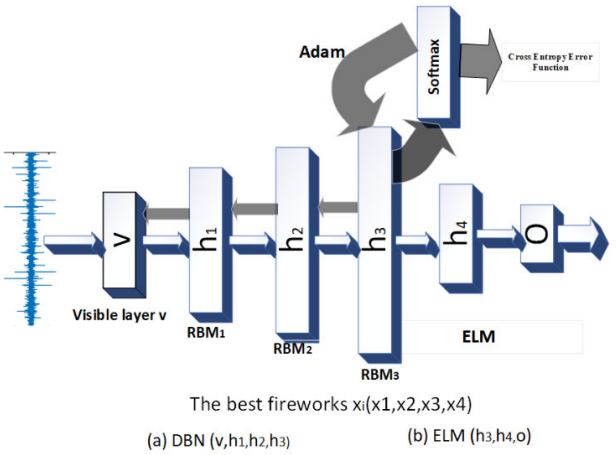

FIGURE 4. IFWA-DBN-ELM sub-health identification model.

$$\Delta w_{ij} = \gamma \left(\langle v_i, h_i \rangle_{data} - \langle v_i, h_i \rangle_{recon}\right) \tag{32}$$

$\langle v_i, h_i \rangle_{data}$, $\langle v_i, h_i \rangle_{recon}$ are respectively training data, refactoring expectations and learning rate γ. Adam is introduced to adjust the learning rate γ dynamically. According to Section V.B, the adjustment of parameters in ELM is different from the gradient learning algorithm. If ELM is directly used as the classification layer to train the model, DBN can't directly apply Adam algorithm to adjust the learning rate γ to match the best hyperparameters. For fitting the training set better, this paper adds *softmax* classification layer after h_3, and the cross entropy which is more suitable for the potential vector is chosen as the loss function. Contrastive Divergence (CD) is used to train the weights and bias in DBN layers.

Initialize IFWA related parameters, and the average accuracy error of the 10-fold cross-validation is used as the fitness function. Calculate model's fitness value, mutation, selection. After that, establish IFWA-DBN-ELM sub-health recognition model ($v \rightarrow h_1 \rightarrow h_2 \rightarrow h_3 \rightarrow h_4 \rightarrow o$).

In DBN, IFWA-MCKD, RCMDE, time-frequency domain features form the data set, and apply the CD-1+ Adam to update the weights and bias. Adopt unsupervised greedy layer-wise to train DBN until input the *softmax* layer to classify. After that, compared with the real labels, the cross-entropy loss is calculated, and Adam is applied to fine-tune the weight and bias by backpropagation. After completing, save the DBN hyperparameters.

In ELM, the output vector in h_3 is used as the input h_4. Eq.29 uses the sigmoid activation function, and calculates the output weight matrix $\acute{\beta}$ in h_4 according to Eq.32. After training, save the ELM hyperparameters.

VI. EXPERIMENTE

A. DATE SELECTION

In order to test the performance about IFWA-DBN-ELM sub-health recognition algorithm, the measured data of the drive rolling bearings from Case Western Reserve University

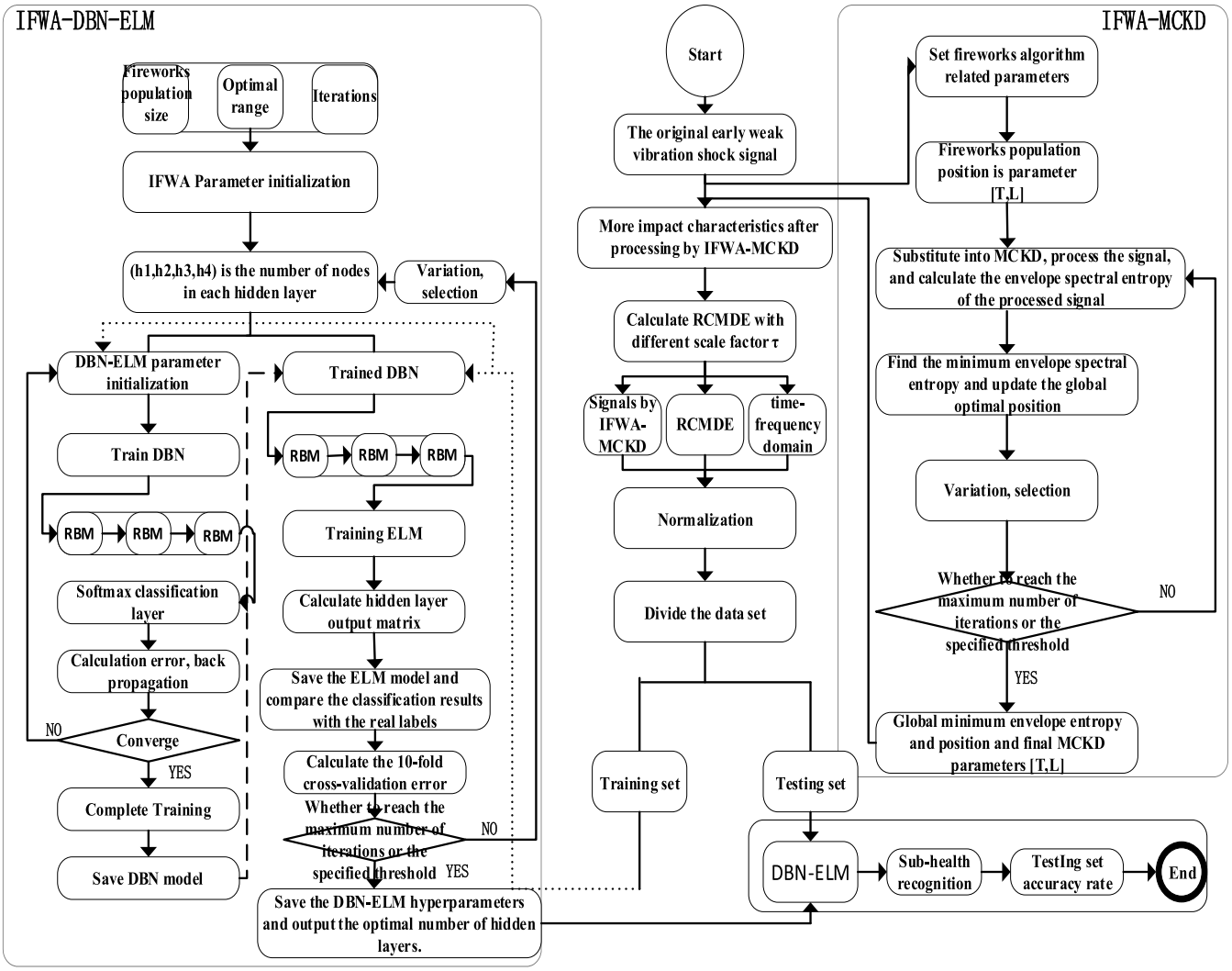

FIGURE 5. Flow of IFWA-DBN-ELM sub-health recognition model.

will be used. The sampling frequency is 12kHz, the load is 0∼2HP. The electric firework causes a single point of fault. The acceleration sensor is placed on the drive end of the motor shell to collect the acceleration data, as shown in Table 5.

B. IFWA-MCKD

To verify the improvement of the proposed method to MCKD algorithm, according to Literature [27], [32], [33], the IFWA and MCKD parameter settings are shown in Table 6. Experimental environment: Intel @Core @CPU i7-6700HQ, 8GB RAM, 64-bit operating system, MATLAB R2019a.

1) [T, L] OPTIMIZATION STRATEGY

Preliminarily stipulate that K=1. According to the related information, the fault characteristic frequencies of IR, Ball and OR are 162.2Hz, 141.1Hz and 107.3Hz respectively. Calculated from Eq.18, the possible values T. Therefore, T search space is set as $[69, 79]$, $[80, 90]$, $[107, 117]$. Taking the OR as an example, the effect of IFWA-MCKD is shown in Fig.6:

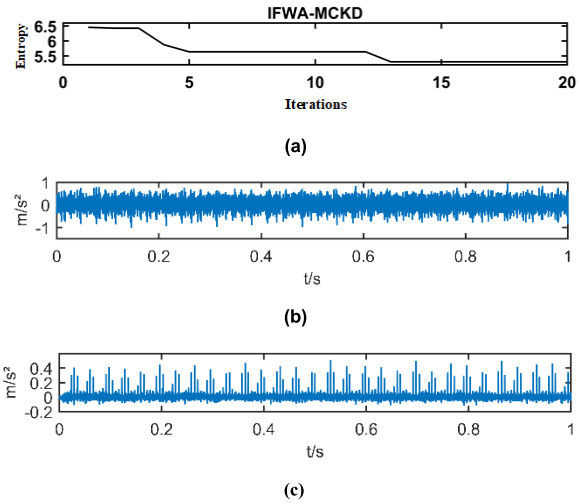

FIGURE 6. (a) Envelope entropy values under different periods for IR (b) The original signal and (c) Processing effect of IFWA-MCKD algorithm.

As can be seen from Fig.6(a), envelope spectrum entropy converges to 5.3027 at the 13th iteration, and the optimal

TABLE 4. IFWA-DBN-ELM algorithm.

IFWA-DBN-ELM sub-health recognition process:
Step1: Divided the training set and testing set.
Step2: Set relevant parameters in IFWA.
Step3: The locations in fireworks is mapped to each hidden layer neuron number in DBN-ELM.
Step4: Train DBN, and the $softmax$ layer is used as classifier first.
Step5: Calculate the cross-entropy loss and fine-tune DBN by backpropagation Adam. Save the adjusted DBN.
Step6: Input training set into the trained DBN. h_3 as input layer, h_4 as output layer in ELM, calculate weight output beta $\hat{\beta}$.
Step7: Calculate the average accuracy error of 10-fold-cross validation about classification labels and real labels in ELM, and take it as the IFWA fitness function.
Step8: Update the minimum fitness value. Fireworks population variation, selection.
Step9: Update the global optimal fireworks location and the number of the hidden layer nodes.
Step10: When the iteration ends or the specified threshold is reached, output the optimal fireworks location, and save relevant parameters. Otherwise, repeat steps 3 ~ 10.
Step11: Test the DBN-ELM performance and record training and test time.

TABLE 5. Fault description and sample distribution.

Fault state	label	training/testing	F. D
Normal(N)	1	120/30	0
Inner ring failure (IR)	2	120/30	0.007
Rolling body fault (Ball)	3	120/30	0.007
Outer ring failure (OR)	4	120/30	0.007

TABLE 6. Parameter settings of IFWA and MCKD.

N	MaxIt	TermIter	T'	L
30	20	30	5	[100,600]

result is [74,385]. Compared with Fig.6(b) and (c), the signal pulse components related to the fault can been seen increased significantly after processing. Therefore, the sparsity is enhanced, the SNR is improved, and the fault features are prominent, thus achieving the effect of noise reduction.

The same method is applied to other parts, and the relevant data will be shown in Table 7:

2) K OPTIMIZATION STRATEGY

Taking IR for example, explore the influence in different K, as shown in Fig.7. Obviously, with K increasing, the kurtosis generally decreases. When K=1, the kurtosis reaches the maximum value 37.2470. Although previous studies have

TABLE 7. Summary about IFWA-MCKD processed results.

Fault state	T	Optimal result	Entropy
N	-	[115,556]	5.7128
IR	74	[74,385]	5.3027
Ball	85	[84,298]	5.7722
OR	112	[112,447]	5.6700

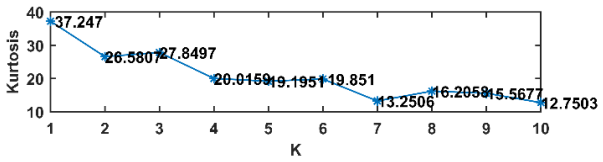

FIGURE 7. Corresponding kurtosis values at different shift numbers K.

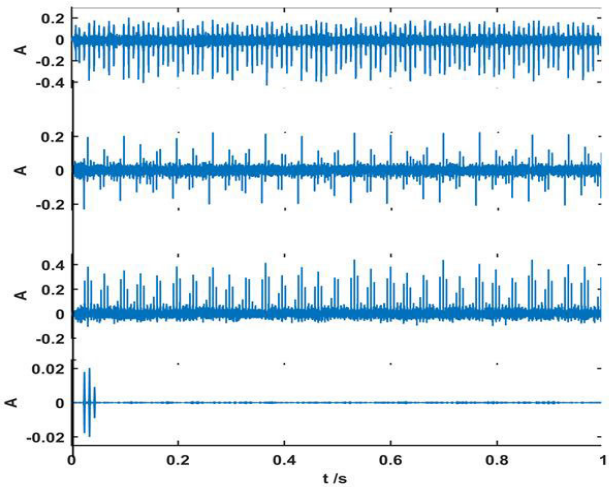

FIGURE 8. N, IR, Ball, OR by IFWA-MCKD.

shown that the greater K value, the more signal pulse composition. However, we find it that greater K can not only lead to accuracy decrease, but also cover some key signals easily, making the characteristic signal relatively fragmented, characteristics from the pulse signal less prominent. Eventually it leads to a decline in signal kurtosis, which is unfavorable to sub-health recognition for factory equipment.

To sum up, under the relevant settings, the final deconvolution signals can be obtained, as shown in Fig.8

C. RCMDE

In order to further extract the signal features, avoid adopting single attribute features and enrich sample feature types, extract RCMDE about processed signals by IFWA-MCKD.

References [9], [10], setting the parameters of RCMDE: $m = 3, c = 6, d = 1, \tau = 20$; the corresponding RCMED is calculate under different scale factor τ as shown in Fig.9:

From Fig.9, it shows that the normal signal RCMDE is wavy. For fault signals, the trend of RCMDE generally decreases with the scale factors increasing and gradually becomes stable. Although there is a slight increase at some individual scales, it doesn't affect the overall trend. At the same time, it can be found that when the scale factor is

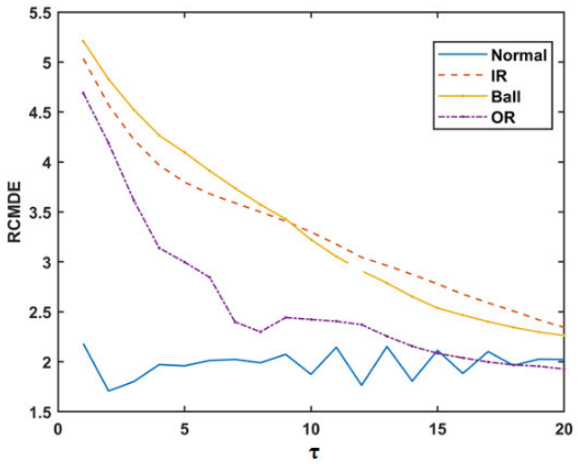

FIGURE 9. RCMDE of different scale factors τ.

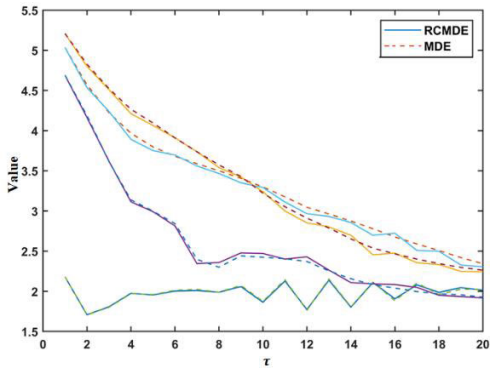

FIGURE 10. RCMDE and MDE of different scale factors τ.

TABLE 8. The standard deviations between RCMDE and MDE.

	N	IR	Ball	OR	Overall
RCMDE	0.1468	0.6369	0.7803	0.7766	0.5852
MDE	0.1517	0.6256	0.7859	0.7833	0.5866

large, the difference of RCMDE values is small and there is crossover and overlap. If more scale RCMDE is selected as the fault feature vector, although it can be distinguished, it will cause information redundancy, which is not good for classification and recognition. Based on the steady trend, RCMDE within $\tau \leq 14$ is taken as the signal feature. It is observed that when the scale factor is small, the various types are divided more obviously, but at larger scale factors, RCMDE values are close at times. Obviously, although RCMDE can save much characteristic information for the original signals, it also needs to use the IFWA-DBN-ELM to identify the fault.

Next, it is well known that RCMDE is put forward on the basis of MED. Compare the effects of RCMDE and MDE as shown in Fig.10:

In Fig.10, RCMDE and MDE are basically the same, and the overall trend of the results of them is similar. Therefore, compare the standard deviations between them in Table 8:

TABLE 9. Parameter settings of IFWA and DBN-ELM.

N	FmaxIt	Regularization coefficient	Batchsize
30	40	0.001	48
epoch	input	output	h
20	430	4	[100,500]

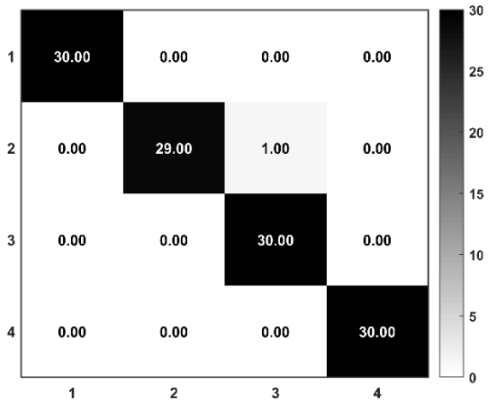

FIGURE 11. IFWA-DBN-ELM sub-health recognition confusion matrix.

As can be seen from Fig.10 and Table 8, although the curve trends are similar and the entropy value is not much different, RCMDE has a smaller overall standard deviation, which indicates that RCMDE has better stability and reliability.

D. IFWA-DBN-ELM SUB-HEALTH RECOGNITION ALGORITHM EXPERIMENT

Referring to the suggestion given by Literatures, Parameter settings about IFWA, DBN-ELM are shown in the Table 9.

The optimization limit about the hidden layer neuron number is [100, 500]. The neuron number in the visible layer is 430 (including 400-dimensional processed data, 16 time-frequency domain features, RCMDE under τ =1∼14). Train the model according to the process shown in Fig.5. The final topology is 430- **129-375-375-165**-4. The corresponding confusion matrix about testing set is shown in Fig.11.

It can be seen that the model has achieved accurate identification. For IR, only one testing sample is mistakenly identified as Ball, but the overall testing set recognition accuracy rate reached 99.17% (in tables, the accuracy is measured in %). It shows that the IFWA-DBN-ELM algorithm finds suitable hyperparameters through methods such as optimizing the number of hidden layer neurons and improving training ways. It gives full play to DBN-ELM powerful performance. It has achieved good results in the sub-health equipment recognition under certain working condition with less time which is shown as Table 11.

DBN-ELM performance is greatly affected by the number of neurons in the hidden layers. For this reason, explore DBN-ELM under different numbers of neurons.

TABLE 10. DBN-ELM hidden layer number change performance.

Structure	p_0	Kappa	Mi1	Ma1
500-200-100-500	59.17	45.15	59.17	NaN
500-200-200-500	60.83	46.99	71.67	61.03
500-300-200-500	62.50	49.55	62.50	62.02
500-200-200-10000	74.17	65.14	74.17	75.44
500-300-200-10000	86.67	81.76	86.67	88.43
500-300-200-20000	75.00	65.81	75.00	78.46

Based on confusion matrix, Kappa, Micro-F1(Mi1) and Macro-F1(Ma1) are defined as follows.

Kappa coefficient is used to measure the effect of classification, which is calculated as Eq.33∼34.

$$kappa = \frac{p_0 - p_e}{1 - p_e} \tag{33}$$

$$p_e = \frac{\sum_{i=1}^{n} a_i * b_i}{c^2} \tag{34}$$

Among these, p_0 is classification accuracy. a_i, b_i are respective the sum of the ith column, row.

F1-Score is harmonic mean of precision and recall, and is defined as Eq.35.

$$F_1 = 2 \times \frac{precision \cdot recall}{precision + recall} \tag{35}$$

Calculate the total precision and recall of all classes, and then calculate F1 called Micro-F1. After that, calculate the precision and recall of each class, and then calculate F1 to get the average for each class called Macro-F1.

Select the highest recognition effect in 10 experiments. Evaluate by Kappa, Mi1 and Ma1 values based on confusion matrix (10^{-2}), accuracy (p_0) as shown in Table 10:

In Table 10, it shows that when h_4 ≤10000, the more neurons is, the higher Kappa value and the better model consistency, but the sub-health recognition accuracy doesn't exceed 90%, because it doesn't match the optimal topology under the artificial setting. Later, when h_4 =20000, the model Mi1 decreases, indicating that DBN-ELM has been overfitting. In addition, when h_4 =10000, the model achieves the optimal effect 86.67%. However, after optimization by IFWA, h_4 is only 165, which it leads to a great waste of computing resources and illustrates the necessity of swarm algorithm optimization further.

GA is worse than other three smart algorithms as shown in Fig.2-3 and Table 2, so GA isn't worth considering. In order to verify the superiority by IFWA optimization effect, compare with PSO and FWA optimal DBN-ELM. The results including training time and testing time(s) are shown in Table 11.

The variation curve of the 10-fold cross-validation average errors are shown in Fig.12.

When the training accuracy is less than 0.01, we think the model converges, adopts Incipient Stopping, and saves hyperparameters. From Table 11 and Fig.12, it can be seen

TABLE 11. Recognition effect under different optimization algorithms.

Algorithm	p_0	Kappa	Mi1	Ma1	Train	Test
PSO	94.15	92.08	94.15	93.56	1.90	0.56
FWA	97.50	96.61	97.50	96.46	1.91	0.56
IFWA	99.17	98.87	99.17	97.88	1.89	0.55

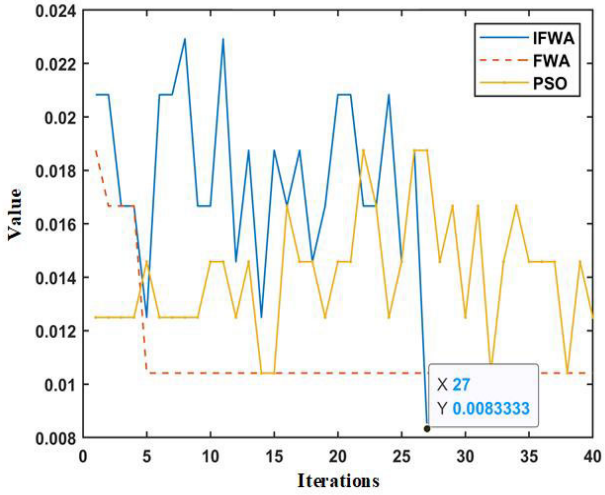

FIGURE 12. Changes in fitness values.

TABLE 12. Recognition effect under different classifiers.

Structure	p_0	Kappa	Mi1	Ma1
IFWA-DBN-ELM	99.17	98.87	99.17	97.88
IFWA-DBN-SVM	98.33	97.73	98.34	97.51
IFWA-DBN-$softmax$	97.50	96.60	97.50	96.76

that the sub-health identification effect by IFWA is better than that by FWA and PSO. As the iteration increases, within the error range [0,0.024], the curve for IFWA changes as a jump, and the accuracy error reaches the minimum value 0.0083333 at the 27th generation. Observed at the same time, the fitness values by PSO and FWA fluctuation near 0.01. When searching, the algorithms sink into the local optimum, but IFWA enhance the local search capability, out of the local optimum, finding the better topology structure, so the accuracy of the test set has been improved.

Combined with Table 10∼11, compared to the traditional DBN-ELM, the sub-health identification accuracy is improved after using IFWA to optimize the parameters of the model, although it takes a long time (7427.2s) to search the optimal topology. At the same time, after determining the topology, train DBN-ELM. Although there is little time gap between the three algorithms, the topology structure optimized by IFWA has shortest training and testing time.

Furthermore, compare the effects under different classifiers, as shown in Table 12:

From Table 12, compared with the previously classifier model, with the same optimization method, it can be observed

TABLE 13. Comparison of data before and after processing.

Data	p_0	Kappa	Mi1	Ma1
After	99.17	98.87	99.17	97.88
Before	97.50	96.61	97.50	96.46

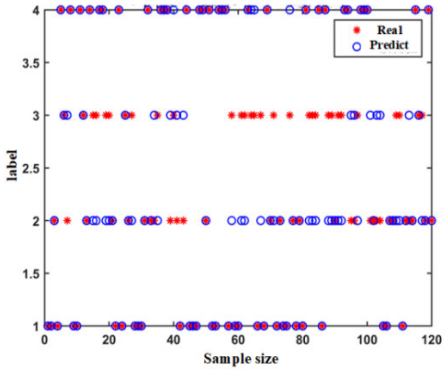

FIGURE 13. Unprocessed data recognition rate under load 2.

that there is no great difference in the recognition accuracy through three classifiers. The further analysis indicates that compared with SVM, ELM classifier has a relatively lower computational complexity due to random weights and other measures, and the training time is less. Besides, *Softmax* classifier is too simple, and it is relatively insufficient to improve overfitting in DBN.

In order to explore the effect of data signal processing on sub-health identification, the data before and after processed are compared in the same environment. The results are shown in Table 13.

The unprocessed data optimization topology is 430-**156-265-137-489**-4. The anti-noise capability for IFWA-DBN-ELM sub-health recognition algorithm can be tested by experiments before processing. Even if the data isn't processed, p_0 still reaches 97.50%, which shows that it has a strong capability to resist noise and has a strong robustness. After processing, because the noise is reduced and the samples are more abundant, DBN can better extract characteristics, and improve the overfitting. Mi1 is increased by 0.0167.

But it is found that the data processed for Transfer learning effect is very great. When the load changes from 0 to 1 and 2, the same model with the same topology accuracy for unprocessed signals will decrease to 60.00% $\sim$ 80.00%, including one experiment as shown in Fig.13, the main performance is that Ball and OR fault identification error is extremely high. It may be that their fault pulse features are alias. Although the normal signal recognition rate is still 100% at load 2.

Finally, compare other shallow models, as shown in Table 14:

E. TRANSFER LEARNING CAPABILITY

In order to verify the transfer learning capability for the proposed IFWA-DBN-ELM sub-health recognition algorithm,

TABLE 14. Performance comparison of different models.

Structure	p_0	Structure	p_0
GCSO-RBFNN [34]	94.19%	IROA-ELM [16]	95.34%
EFWA-BP [35]	94.93%	SVM [16]	82.33%

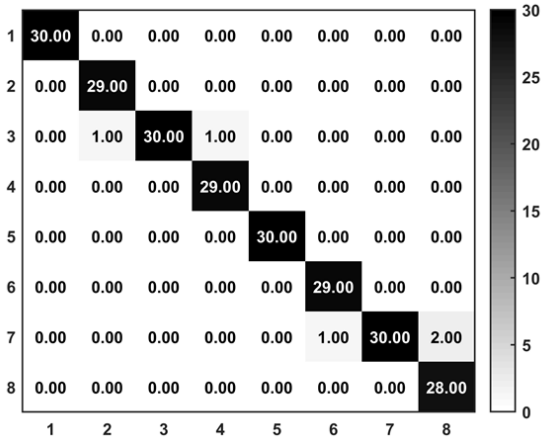

FIGURE 14. Confusion matrix under different loads.

signals under different loads 1 and 2 ($2 \times 4 = 8$) are selected. The confusion matrix as shown in Fig.14:

Labels 1$\sim$8 can be the signal types for four states under 1,2 load conditions respectively. It can be seen that only 5 samples are identified wrong, and the recognition accuracy reached 97.92%, which indicates that IFWA-DBN-ELM proposed in this paper has a good migration versatility fully. As can be seen from the confusion matrix in different load shown in Fig.14, the normal state recognition accuracy is 1. Only remaining misjudgments are errors between different fault categories, but the sub-health identification accuracy between normal and fault is 100%. The experimental results fully show that the proposed IFWA-DBN-ELM sub-health identification algorithm has the good mobility and generality, and it also has achieved better fault identification and fault location effect under different working conditions.

F. MULTICLASSFICATION CAPABILITY

To continue to verify the performance about IFWA-DBN-ELM. On the basis of Table 5, the damage diameters are increased by 0.014 and 0.021. Data can be shown in Table 15:

The data processing and the model training are described in VI. EXPERIMENTE. It takes 18568.63s to find the best topology. After optimization, the processing topology (18568.63s) is 430-**467-105-252-316**-10, the training error is 0.0041667, and the accuracy of the training set is 100%. The results of the test set are shown in Table 16:

The test set confusion matrix is shown in Fig.15:

From Fig.15, it can be seen that the optimized topology also has higher recognition accuracy, and the accuracy of normal and fault signal recognition is as high as 100%. Only part

TABLE 15. Comparison of data before and after processing.

Fault state	label	training/testing	F. D
Inner ring failure (IR)	5	120/30	0.014
Rolling body fault (Ball)	6	120/30	0.014
Outer ring failure (OR)	7	120/30	0.014
Inner ring failure (IR)	8	120/30	0.021
Rolling body fault (Ball)	9	120/30	0.021
Outer ring failure (OR)	10	120/30	0.021

TABLE 16. The results of the test set.

p_0	Kappa	Mi1	Ma1	Train	Test	Epoch
99.33	99.2555	99.3333	99.3325	7.70	0.54	30

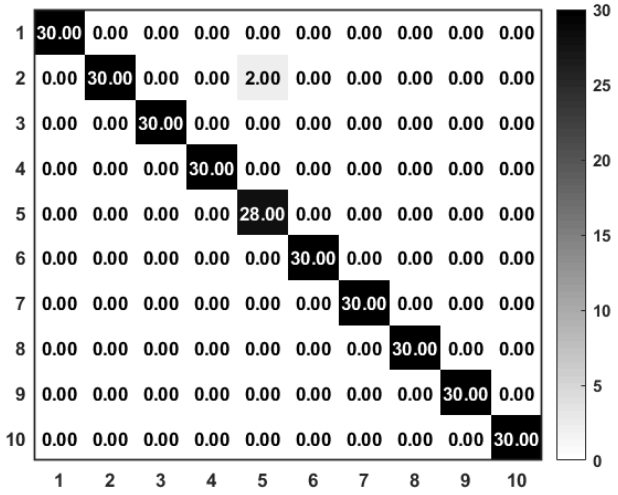

FIGURE 15. Confusion matrix under different damage diameters.

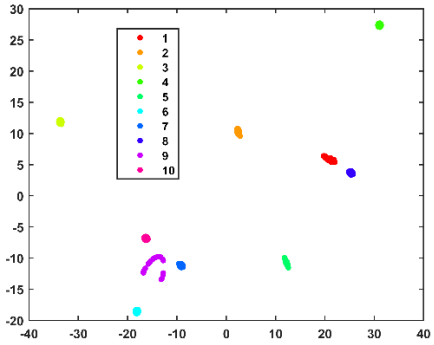

FIGURE 16. t-SNE design sketch.

of Label 5 is wrongly identified as 2, which it is possible that the model has slightly over-fitting for the learning feature of 2 or 5. It is gratifying that 99.33% is achieved with less epoch, and the running time is 8.24s, which has less running time.

In order to study the internal principle of IFWA-DBN-ELM, the data will be used for visual analysis in the future. Through t-SNE [36] dimension reduction, the characteristics of ELM output data are visualized, as shown in Fig.16.

After DBN-ELM processing, the features are completely separated and gathered in the corresponding area. Because multi feature combination can obtain more information than single feature combination, the effect of feature visualization is more obvious.

VII. CONCLUSION

The bearing incipient signals are weak and noisy, causing fault identification is difficult. Therefore, an IFWA-DBN-ELM sub-health recognition algorithm based on IFWA-MCKD and RCMDE has been proposed. IFWA-MCKD can effectively improve SNR, enhance pulses. The fault features are extracted by RCMDE and the multi feature data set is composed of RCMDE, time-frequency domain and data denoised by MCKD. IFWA-DBN-ELM can adaptively select the number of neurons in the hidden layers. Moreover, it improves the training method to make the weight and bias more fit the signals and improve overfitting. Through experiments, IFWA-DBN-ELM has achieved better effect with less time, and the migration capability under different loads has achieved a higher accuracy. It is instructive for large equipment to identify sub-health state and deal with timely.

In addition, the method presented in this paper can also be used in the analysis of medical EEG signal and voltage signal of power electronic equipment [37], [38]. In the future, we will further try to explore and experiment on different signals in different fields.

REFERENCES

[1] L. Zhang, C. Qiu, and K. Zhang, "Sub-health recognition algorithm of improved capsule network optimized layered convolutional," *J. Frontiers Comput. Sci. Technol.*, early access, Sep. 2019, doi: 10.3778/j.issn.1673-9418.2004017.

[2] B. Pang, G. Tang, and T. Tian, "Enhanced singular spectrum decomposition and its application to rolling bearing fault diagnosis," *IEEE Access*, vol. 7, pp. 87769–87782, 2019.

[3] D. Ma, X. Hu, H. Zhang, Q. Sun, and X. Xie, "A hierarchical event detection method based on spectral theory of multidimensional matrix for power system," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Aug. 9, 2019, doi: 10.1109/TSMC.2019.2931316.

[4] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2013.

[5] G. L. McDonald, Q. Zhao, and M. J. Zuo, "Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection," *Mech. Syst. Signal Process.*, vol. 33, pp. 237–255, Nov. 2012.

[6] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016.

[7] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.

[8] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.

[9] H. Azami, M. Rostaghi, D. Abásolo, and J. Escudero, "Refined composite multiscale dispersion entropy and its application to biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2872–2879, Dec. 2017.

[10] H. Zhang, X. Hu, D. Ma, R. Wang, and X. Xie, "Insufficient data generative model for pipeline network leak detection using generative adversarial networks," *IEEE Trans. Cybern.*, early access, Dec. 9, 2020, doi: 10.1109/TCYB.2020.3035518.

[11] K. Shao, W. Fu, J. Tan, and K. Wang, "Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational harris hawks optimization-based SVM for fault diagnosis of rolling bearing," *Measurement*, vol. 173, Mar. 2021, Art. no. 108580.

[12] J. Li, X. Yao, X. Wang, Q. Yu, and Y. Zhang, "Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis," *Measurement*, vol. 153, Mar. 2020, Art. no. 107419.
[13] H. Shi, L. Guo, S. Tan, and X. Bai, "Rolling bearing initial fault detection using long short-term memory recurrent network," *IEEE Access*, vol. 7, pp. 171559–171569, 2019.
[14] J. Chorowski, J. Wang, and J. M. Zurada, "Review and performance comparison of SVM- and ELM-based classifiers," *Neurocomputing*, vol. 128, pp. 507–516, Mar. 2014.
[15] J. Yu and G. Liu, "Knowledge extraction and insertion to deep belief network for gearbox fault diagnosis," *Knowl.-Based Syst.*, vol. 197, Jun. 2020, Art. no. 105883.
[16] C. He, T. Wu, R. Gu, Z. Jin, R. Ma, and H. Qu, "Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm–extreme learning machine," *Measurement*, vol. 173, Mar. 2021, Art. no. 108636.
[17] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, Jan. 2018.
[18] D. Liang and P. Pan, "Research on intrusion detection based on improved DBN-ELM," in *Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE)*, Jul. 2019, pp. 495–499.
[19] X. Wang, W. Hu, K. Li, L. Song, and L. Song, "Modeling of soft sensor based on DBN-ELM and its application in measurement of nutrient solution composition for soilless culture," in *Proc. IEEE Int. Conf. Saf. Produce Informatization (IICSPI)*, Dec. 2018, pp. 93–97.
[20] G. Zhang, C. Tian, C. Li, J. J. Zhang, and W. Zuo, "Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature," *Energy*, vol. 201, Jun. 2020, Art. no. 117531.
[21] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980
[22] B. H. Nguyen, B. Xue, and M. Zhang, "A survey on swarm intelligence approaches to feature selection in data mining," *Swarm Evol. Comput.*, vol. 54, May 2020, Art. no. 100663.
[23] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Sep. 2015, Art. no. 115002.
[24] H. Du, D. Song, Z. Chen, H. Shu, and Z. Guo, "Prediction model oriented for landslide displacement with step-like curve by applying ensemble empirical mode decomposition and the PSO-ELM method," *J. Cleaner Prod.*, vol. 270, Oct. 2020, Art. no. 122248.
[25] T. Ying and Y. Zhu, "Fireworks Algorithm for Optimization," in *Proc. Int. Conf. swarm Intell.*, vol. 6145, Jun. 2010, pp. 355–364.
[26] J. Duan, Q. Qu, C. Gao, and X. Chen, "Adaptive dynamic enhanced fireworks algorithm based on Lévy flight," *Appl. Res. Comput.*, vol. 35, no. 10, pp. 3011–3015, 2018.
[27] G. Tang and X. Wang, "Adaptive maximum correlated kurtosis deconvolution method and its application on incipient fault diagnosis of bearing," *Proc. Chin. Soc. Electr. Eng.*, vol. 35, no. 6, pp. 1436–1444, 2015.
[28] G. Tang and X. Wang, "Diagnosis method for rolling bearing incipient faults based on sparsity of envelope spectrum and maximum correlated kurtosis deconvolution," *China Mech. Eng.*, vol. 26, no. 11, pp. 1450–1456, Nov. 2015.
[29] Y. Pan and J. Chen, "Application of envelope spectrum entropy in performance degradation assessment of rolling bearings," in *J. Shanghai Inst. Technol. Natural Sci.*, vol. 14, no. 3, pp. 220–223, 2014.
[30] L. Bo, Z. Tang, and G. Xu, "Bearing fault diagnosis based on variational modal decomposition combined with envelope of spectrum," *J. Vibrat., Meas. Diagnosis*, vol. 38, no. 3, pp. 597–602, Jun. 2018.
[31] H. Zhao and L. Li, "Wind turbine bearing fault diagnosis method based on maximum correlation kurtosis deconvolution and variational modal decomposition," *Acta Energiae Solaris Sinica*, vol. 39, no. 2, pp. 350–358, 2018.
[32] M. Majdouli, I. Rbouh, and S. Bougrine, "Fireworks algorithm framework for big data optimization," *Memetic Comput.*, vol. 8, no. 6, pp. 1–15, Jun. 2017.
[33] F. Zhu, D. Chen, and F. Zou, "A novel hybrid dynamic fireworks algorithm with particle swarm optimization," *Soft Comput.*, vol. 25, no. 3, pp. 2371–2398, Feb. 2021.
[34] W. Guo, C. Qiu, D. Zhang, Y. Wang, and L. Zhang, "Sub-health diagnosis model of RBF neural network optimized by growth chicken swarm algorithm," *J. Chin. Comput. Syst.*, vol. 41, no. 5, pp. 961–966, 2020.
[35] J. Zhou, *Research and Implementation of Fan Intelligent Diagnosis System Based on Improved Firework Algorithm*. Shenyang, China: Liaoning Univ., 2020.
[36] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, Nov. 2018.
[37] Q. Sun, R. Han, H. Zhang, J. Zhou, and J. M. Guerrero, "A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy Internet," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 3006–3019, Nov. 2015.
[38] R. Wang, Q. Sun, W. Hu, Y. Li, D. Ma, and P. Wang, "SoC-based droop coefficients stability region analysis of the battery for stand-alone supply systems with constant power loads," *IEEE Trans. Power Electron.*, vol. 36, no. 7, pp. 7866–7879, Jul. 2021, doi: 10.1109/TPEL.2021.3049241.

HAO LUO was born in 1987. He received the Ph.D. degree from Liaoning Technical University, in 2016. He is currently a Lecturer and a Ph.D. Supervisor with Liaoning University. His research interests include big data monitoring and early warning technology, and data mining of mining dynamic disaster.

CHAO HE was born in 1995. He is currently pursuing the M.S. degree with Liaoning University. His research interest includes fault diagnosis.

JIANING ZHOU was born in 1994. She is currently pursuing the M.S. degree with Liaoning University. Her research interest includes fault diagnosis.

LI ZHANG was born in 1971. He received the Ph.D. degree in rolling technology and rolling automation from Northeastern University, in 2000. He is currently a Professor and a M.S. Supervisor with Liaoning University. His research interests include sub-health status identification of equipment, intelligent computing, and so on.