Integrated approach for ball mill load forecasting based on improved EWT, refined composite multi-scale dispersion entropy and fireworks algorithm optimized SVM

Jiacheng Cai¹, Lirong Yang², Changxi Zeng² and Yongkang Chen²

Abstract
Shell vibration signals generated during grinding have useful information related to ball mill load, while usually contaminated by noises. It is a challenge to recognize load parameters with these signals. In this paper, a novel approach is proposed based on the improved empirical wavelet transform (EWT), refined composite multi-scale dispersion entropy (RCMDE) and fireworks algorithm (FWA) optimized SVM. Firstly, vibration signals are denoised by improved EWT, which uses cubic spline interpolation to calculate envelope spectrum for segmentation. Then, RCMDEs of the denoised signals are calculated as feature vectors. The vectors’ dimensionalities are reduced by principal component analysis (PCA). Finally, a mill load prediction model is established based on the FWA optimized SVM. The reduced feature vectors are fed to the model, thus material-to-ball ratio and filling rate being outputs. Grinding experiments show that the extracted features by RCMDE can effectively distinguish three load states. Meanwhile, experiments also show that FWA reduces the forecasting errors of material-to-balls ratio and filling rate by 1.9% and 2.9% compared with genetic algorithm (GA), as well as by 1.92% and 4.21% compared with particle swarm optimization (PSO) algorithm. It demonstrates that the proposed approach for ball mill load forecasting has high accuracy and stability.

Keywords
Mill load, empirical wavelet transform, refined composite multi-scale dispersion entropy, firework algorithm, support vector machine

Date received: 26 August 2020; accepted: 8 January 2021

Handling Editor: James Baldwin

Highlights
1. An improved empirical wavelet transform (EWT) denoising algorithm is proposed.
2. A refined composite multi-scale dispersion entropy (RCMDE) is proposed to extract the vibration features.
3. A fireworks algorithm (FWA) is proposed to optimize support vector machine (SVM).

¹Department of Machine Design/Mechatronics, KTH Royal Institute of Technology, Stockholm, Sweden
²School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, Ganzhou, Jiangxi, China

Corresponding author:
Lirong Yang, School of Mechanical and Electrical Engineering, Jiangxi University of Science and Technology, No.86, Hongqi Road, Zhanggong District, Ganzhou, Jiangxi 341000, China.
Email: candy_yang@js@163.com

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
4. The optimized SVM is applied to forecast the ball mill load with the extracted features.
5. Comprehensive experiments are designed to evaluate the effectiveness of the proposed methods.

Introduction

Ball mill is the key equipment of mineral production process, while its disadvantages include low working efficiency and high energy consumption, which influence the optimization control and energy saving of mineral processing enterprises. The work efficiency is highest while the ball mill load is in the best state. However, the ball mill load is difficult to measure directly. Therefore, the accurate recognition of ball mill load has become the focus of researchers.

In the grinding process, the vibration signal generated by ball mill contains useful information related to internal load parameters, which can be used to determine mill load parameters indirectly. However, due to the large amount of noises contained in the acquired vibration signals and their characteristics of non-linear and non-stationary, it is a challenge to remove the noises, extract the latent features, and construct an recognition model to forecast mill load parameters based on these signals. Many traditional and improved denoising algorithms, feature extraction algorithms and pattern recognition algorithms have been used in ball mill vibration signal processing and load recognition, but the processing effect and recognition accuracy still need further research.

In processing nonlinear signals, the empirical mode decomposition (EMD) and wavelet transform (WT) have certain advantages, but there are also problems such as mode mixing and inappropriate selection of wavelet basis function. Chen et al. used a wavelet packet to decompose the vibration signal of the rolling bearing into several sub-signals containing fault information. The wavelet packet transform is an extension of the WT that can select an arbitrary frequency resolution, which makes it relatively flexible. In 2013, Gills presented a new method called empirical wavelet transform. Its essence is to generate adaptive wavelets for extracting feature information from the original signals. In recent studies, the EWT decomposition technique has been applied in various fields. Chegini et al. proposed a new vibration signal denoising method based on the EWT for rolling bearings fault diagnose. The results show that the EWT-based denoising method is superior to the EMD denoising technique. In, Liu et al. EWT is used to decompose the original data for the wind speed forecasting. Singh and Sunkaria presented a novel method for removing the power-line interference and refining the baseline wander in electrocardiogram signal utilizing the EWT method. Li et al. proposed a novel technique for down-hole micro-seismic denoising using the EWT method and an adaptive threshold. Their results show that the performance of the EWT denoising method is better than the WT, the complete ensemble EMD and synchro-queezed WT. However, there are some problems in the EWT, such as inaccurate spectrum segmentation, which leads to an inaccurate decomposition effect. Overall, further optimization is needed.

In terms of feature extraction, nonlinear dynamics methods are favored. Zhang and Shang used single-scale permutation entropy and four-scale permutation entropy indicators to detect milling flutter online. Experiments demonstrate that this method can accurately and effectively identify milling flutter. And the characteristic entropy is insensitive to noises and parameter changes, which has good practicality. Mehta et al. investigated the time series complexity of the motor vibration signals. He applied the multi-scale fuzzy entropy (MFE) to fault detection of rolling bearings. Results show that this method is better than SVM. Besides, Cai et al. used adaptive frequency windows EWT-MFE and adaptive evolution particle swarm optimization probabilistic neural network (AEPSO_PNN) to identify and classify the load status of ball mills. The results showed that the MFE features can better distinguish three different operating conditions, and the AEPSO_PNN method has higher recognition accuracy compared with PNN and BP neural network. Among the above methods, MFE has certain advantages in the field of feature extraction. It can effectively solve the entropy catastrophe problem of multi-scale entropy (MSE) and has a good diagnosis effect. However, MFE tends to ignore the interrelationship of various time series during signal processing. In order to overcome the above shortcomings, Rostaghi and Azami reported a new irregularity index named dispersion entropy (DE). This fast algorithm considers the relationship between the signal amplitudes. And it is less affected by a sudden change signal. In addition, Azami et al. proposed the concept of RCMDE on this basis, and compared with other MSE algorithms, the results showed that RCMDE has unique advantages in feature extraction. Conzhi et al. proposed a new method for rolling bearing fault diagnosis based on RCMDE and support vector machine (SVM). Experiments show that, RCMDE has higher fault recognition rate compared with the multi-scale dispersion entropy and MSE.

The forecasting accuracy is the top priority in the proposed method. SVM solves the several problems such as unclear network structure and local minimum in neural network algorithms. SVM has a good advantage in dealing with nonlinear problems in small-sample identification and prediction and have been widely
used. However, the SVM classification prediction accuracy is affected by the smoothing parameters $C$ and $\sigma$. If $\alpha$ is not selected properly, great errors easily occur. Nowadays, a lot of improved SVM methods are presented. Qi et al. developed combination weights and proposed adaptive weighted elastic net SVM, which performs well in imbalanced data classification problem and presents to be more robustly than the standard SVM and other existing weighted SVMs. Li et al. utilized GA to optimize the parameters of SVM to improve the classification accuracy of SVM. By comparing with SVM, GA-SVM can recognize and classify the flip chips rapidly with high accuracy. Xiaokai et al. proposed a prediction and classification method about maximum tensile-shear strength of spot-welded joints based on the ultrasonic detection signal feature extraction and PSO-SVM classifier. The test results show that the accuracy of PSO-SVM classifier is higher than BP classifier. The GA and PSO algorithm have shown good results in SVM parameter optimization but have certain defects. For example, the GA has a long iterative convergence time in the calculation process. The PSO is susceptible to the local particle optima, and it appears to end iteration too early. Besides, the FWA can automatically optimize the smoothing parameters in an SVM. Min et al. FWA model was used to optimize the SVM parameters for the rolling bearing fault diagnosis. The experiments show that the estimation accuracy is higher than those of GA and PSO algorithm.

In summary, the improved EWT, RCMDE, and FWA-SVM were combined in this study to improve the accuracy of ball mill load prediction. First, the improved EWT algorithm is used to decompose the vibration signal, and the sensitive modal component is selected by the correlation coefficient method to reconstruct the signal. Then calculate the value of RCMDE to judge the load state of the ball mill. Finally, FWA-SVM is used to forecast the load parameters of ball mill. The proposed method provides a new approach of ball mill load forecasting.

**Improved EWT's refined composite multi-scale dispersion entropy (RCMDE)**

**Improved EWT**

EWT is a time-frequency analysis method to extract significant modes of signals. A more detailed introduction to EWT, including the construction of related functions and parameter selection, can be found in. The essential part of this method is segmenting the Fourier spectrum. Relying on the spectrum segmentation method based on the local maximum will lead to the problem of over-segmentation or under-segmentation. In view of this, a more reliable spectrum segmentation method is proposed, and its main process is as follows.

1. The time-domain discrete signal $f(n)$ is obtained. Due to improper selection of sample length or other external reasons (zero-point drift of sensor, signal waveform deviation caused by base movement, etc.) when recording is intercepted, the trend term is generated. Therefore, the least square method is used to eliminate the trend term of the signal.
2. According to the Fourier transform, the frequency spectrum $F(\omega)$ of the input signal can be obtained and normalized to $[0, \pi]$.
3. Then use the cubic spline interpolation method to obtain the spectral envelope of the signal. All local maxima in the signal spectrum are searched, and the maxima are sorted in descending order ($A_1 \geq A_2 \geq \cdots \geq A_m$). Then, the signal is clipped with $r = A_1 - \alpha (A_1 - A_m)$ as the threshold, where $\alpha$ is relative amplitude ratio and it is set as $\alpha = 0.9$ in this application. Lastly, both the intersection of the threshold and the envelope, and the minimum value of the envelope are taken as the spectrum segmentation point $w_a$.
4. Wavelet function $\psi_c(w)$ and scale function $\varphi_c(w)$ are constructed by using the segmented spectrum. The computing details are the same as EWT.
5. By using the inverse Fourier transform to compute $F(w) \times \psi_c(w)$ and $F(w) \times \varphi_c(w)$, the time-domain representation of each component can be obtained. The computing details are the same as EWT.

**Refined composite multi-scale dispersion entropy (RCMDE)**

Dispersion entropy (DE) is an algorithm that characterizes the complexity and irregularity of time series. The application of multi-scale dispersion entropy can further distinguish the characteristic difference. Compared with the multi-scale dispersion entropy algorithm, the RCMDE signal processing method can effectively reduce the loss of statistical information in the coarse-graining process, and prevent the influence of the calculation result from initial point shifting by averaging the multiple initial points, thereby reducing the calculation error. Based on the above analysis, this paper uses RCMDE to extract the characteristics of the vibration signals of the ball mill. The main calculation steps of RCMDE are as follows.

1. For a given original data $X = \{x_1, x_2, \ldots, x_N\}$ with a length of $N$, firstly, it is divided into $\lfloor N/\tau \rfloor$ non-overlapping segments with length $\tau$ from starting point $x_k$. Then, the average of
each segment is computed to derive the coarse-grained time series. This process is named coarse graining and the \( k \)-th coarse-grained time series is \( y_{k}^{\tau} = \{y_{k,1}^{\tau}, y_{k,2}^{\tau}, \ldots, y_{k,N/\tau}^{\tau}\} \), in which

\[
y_{k,j} = \frac{1}{\tau} \sum_{i=j-\tau + 1}^{j} x_{i}
\]

where \( N \) denotes the length of the time series; \( \tau \) is the scale factor; \( k \) represents the serial number of coarse-grained series with the starting point of \( x_{k} \); \( y_{k,j}^{\tau} \) represents the average value of \( j \)-th segment of the \( k \)-th coarse-grained time series while scale factor is \( \tau \).

When the starting point different from \( x_{1} \) to \( x_{\tau} \), we have \( \tau \) different coarse-grained time series.

\[
\text{(2) For each } y_{k}^{\tau}(k = 1, 2, \ldots, \tau), \text{ the } p_{k}^{\tau}(\pi_{wv_{1} \cdots wv_{m-1}}) \text{ can be estimated according to steps (1) to (4) of DE in Appendix A. Next, } p(\pi_{wv_{1} \cdots wv_{m-1}}) = \frac{1}{\tau} \sum_{k=1}^{\tau} p_{k}^{\tau}(\pi_{wv_{1} \cdots wv_{m-1}}) \text{ is computed as the final average dispersion patterns at scale factor } \tau
\]

Finally, the RCMDE of \( X(x_{i}, i = 1, 2, \ldots, N) \) at the scale \( \tau \) is defined by

\[
RCMDE(X, m, c, d, \tau) = -\sum_{\pi_{wv_{1} \cdots wv_{m-1}}} c_{\pi_{wv_{1} \cdots wv_{m-1}}} p(\pi_{wv_{1} \cdots wv_{m-1}}) \ln(p(\pi_{wv_{1} \cdots wv_{m-1}}))
\]

where \( m \) is embedding dimension; \( c \) is the number of class; \( d \) is time delay.

In the RCMDE algorithm, the parameters including class \( c \), embedding dimension \( m \) and time delay \( d \) need to be preset. In practice, time delay is recommended to \( d = 1 \) because aliasing may occur when \( d > 1 \). Clearly, we need \( c > 1 \) in order to avoid the trivial case of having only one dispersion pattern. We always classify the ball mill load into 3 classes including underload, normal load and overload, so \( c = 3 \) is set according to the application. Besides, since too small \( m \) will cause RCMDE insensitive to the dynamic changes of the original signal while too large \( m \) will make RCMDE unable to detect small changes, \( m \) usually takes 2 or 3. Therefore, \( c = 3, m = 2 \) and \( d = 1 \) were set as recommended by Rostaghi and Azami.

**Establishment of mill load prediction model**

**FWA-SVM algorithm principle**

The FWA is a swarm intelligence algorithm obtained by observing the blooming pattern of fireworks. The feasible solution of each space is regarded as a firework. The best iteration times are designed by using an explosion operator, mutation operator, mapping rules, and selection strategy. The steps to optimize SVM by the firework algorithm are as follows.

Step 1 Set \( N \) fireworks to represent random \( N \) solutions \( u_{w} \).

Step 2 Solve the explosive radius \( B_{n} \) and the number of explosive sparks \( Q_{n} \) of each firework.

\[
B_{n} = C \frac{f(u_{n}) - y_{\text{min}} + \epsilon}{\sum_{n=1}^{N}(f(u_{n}) - y_{\text{min}}) + \epsilon}
\]

\[
Q_{n} = H \frac{y_{\text{max}} - f(u_{n}) + \epsilon}{\sum_{n=1}^{N}(y_{\text{max}} - f(u_{n})) + \epsilon}
\]

where \( B_{n} \) is the explosive radius of the \( n \)-th individual; \( Q_{n} \) is the number of sparks generated by the \( n \)-th individual; \( y_{\text{min}} = \min\{f(u_{n})\} \) represents the best fitness value in the current population; \( y_{\text{max}} = \max\{f(u_{n})\} \) represents the worst fitness value in the current population; \( C \) is a constant used to adjust the explosion radius; \( H \) is also a constant used to adjust the amount of sparks produced by the explosion; \( f(u_{n}) \) is the fitness value of the \( n \)-th individual; \( \epsilon \) is the machine minimum to avoid division by zero.

In the iterative process, the number of sparks can be limited to prevent too many or too few sparks under the optimal or inferior adaptability value. The limit conditions are as follows.

\[
S_{n} = \begin{cases} 
\text{round}(a * H), & S_{n} < aH \\
\text{round}(b * H), & S_{n} > bH, a < b < 1 \\
\text{round}(S_{n}), & \text{others}
\end{cases}
\]

where \( S_{n} \) is the number of sparks; \( a \) and \( b \) are constants; \( \text{round}(\cdot) \) is a rounding function.

Step 3 To generate sparks, the dimension of the set \( DS \) is \( z \), and \( z = \text{round}(D \times \text{rand}(0,1)) \). \( D \) represents the dimension of spark \( u_{w} \), and \( T_{nk} \) can be saved in the spark population by overstepping the boundary for blasting operations under each dimension \( k \) in the set \( DS \).

\[
h = R_{a}\text{rand}(-1, 1)
\]

\[
T_{nk} = u_{nk} + h
\]

where \( h \) represents the offset, \( u_{nk} \) represents the position of the \( n \) spark on the \( k \) dimension, and \( T_{nk} \) indicates the position of the spark after the explosion.

Step 4 The mutation operator is calculated, and the Gaussian variation of each dimension is processed.
by crossing the boundary. The result is saved in the Gaussian population.

\[ M_{nk} = u_ne \]  \hspace{1cm} (8)

where \( e \) is the Gaussian noise with a mean of 1 and a variance of 1; \( M_{nk} \) is the Gaussian population of the \( n \) spark on the \( k \) dimension.

Step 5 The best population of all populations is selected, and then the remaining \( N-1 \) population by the wheel selection method is selected. The formula is as follows.

\[ P(u_n) = \frac{R(u_n)}{\sum_{u \in K} R(u_n)} \]  \hspace{1cm} (9)

\[ R(u_n) = \sum_{u_j \in K} d(u_n - u_j) = \sum_{u_j \in K} ||u_n - u_j|| \]  \hspace{1cm} (10)

where \( P(u_n) \) represents the selection probability of the individual \( u_n \); \( R(u_n) \) represents the sum of the Euclidean distance between the \( n \) individual and the other individuals; \( d(u_n-u_j) \) represents the Euclidean distance between individual \( u_n \) and individual \( u_j \); \( K \) is the set of all individuals.

Step 6 If the condition of iteration termination is met, the optimal SVM parameter is generated at the end of the iteration; otherwise, the iteration continues.

**Establishment of load identification model**

The ball mill load forecasting model based on the FWA-SVM algorithm is established. The concrete steps are as follows.

1. The collected vibration signals from ball mill shell under different load conditions are decomposed by the improved EWT, and multiple modal components are obtained.
2. The correlation analysis of each modal component is carried out (details see Cai et al.23), and the sensitive IMF components with a large correlation coefficient are selected to reconstruct the signal. Then, the RCMDEs of the sensitive IMF components are calculated, and the 20-dimensional feature vector set is constructed based on that.
3. PCA is used to reduce the dimensions of the multi-dimensional feature vector.
4. The feature vectors after the dimension reduction are entered into the ball mill load prediction model based on the FWA-SVM algorithm, and then the mill load parameters are acquired. The flow chart is shown in Figure 1.

**Experimental study**

**Data collection**

This experiment uses the \( \Phi 330 \text{mm} \times 330 \text{mm} \) Bond index of dry type ball mill in laboratory as the experimental object. A DH5922N dynamic data acquisition instrument, DH131 vibration acceleration sensor, DH5857-1 charge adapter, cylindrical universal bearing, and PC constitute the vibration signal acquisition system. The picture and schematic diagram of the acquisition system are shown in Figure 2. The sampling frequency is set to 20 kHz. The filling rate and material-to-ball ratio were uniformly changed to collect multiple samples of ball mill shell vibration signals under three different loads. The mill load is classified into three categories: the filling rate between 10% and 20% as underload, the filling rate between 20% and 40% as normal load, and the filling rate between 40% and 60% as overload.

**Decomposition of the shell vibration signal**

The ball mill shell vibration signals are selected in underload (material-to-ball ratio 0.30, filling rate 15%), normal load (material-to-ball ratio 0.50, filling rate 30%), and overload (material-to-ball ratio 0.70, filling rate 50%), respectively. The results are shown in Figure 3.

In Figure 3, it can be seen that there are a lot of noises in the original signals under three load conditions. Although there are some amplitude differences between three load conditions in the time domain, the distinction is not obvious. Therefore, the improved EWT algorithm is used to preprocess 10 groups of vibration signals under each load condition. Then, 10 intrinsic mode function (IMF) components are obtained from each group of signals, and the correlation coefficient between each IMF component and the original signal is calculated. Finally, the average correlation coefficient and average threshold within the 10
groups of signals are calculated under each load condition. The results are shown in Figure 4.

As shown in Figure 4, the average threshold of each load condition is 0.2742 (underload), 0.2431 (normal load), and 0.2399 (overload), respectively. According to the sensitive IMF component selection rules, the correlation coefficient between the component and the original signal is greater than the threshold, the component is retained as a sensitive IMF component. Otherwise, the component is removed as a false component. In the
end, IMF2, IMF7, IMF8, and IMF10 are selected as sensitive modal components in underload condition. IMF2, IMF3, IMF6, IMF7, and IMF10 are selected as sensitive modal components in normal load condition. IMF1, IMF2, IMF3, and IMF10 are determined as sensitive modal components in overload condition. The remaining modal components are considered as false components. After that, the selected sensitive IMF components are used to reconstruct the original signals as shown in Figure 5.

By comparing Figures 3 and 5, we can see that the high frequency noises are effectively removed from the reconstructed signals under different load condition, while the amplitudes and trend changes of the signals are mostly retained, so that the rich feature information of the original signals is preserved in the reconstructed signals. To further verify the superiority of the improved EWT algorithm, EMD and EWT algorithms were used to decompose the same original vibration signals under load condition 1. Then, the correlation coefficient method was also used to select the sensitive IMF components for reconstruction. The reconstructed signals by EMD and EWT are shown in Figure 6.

By comparing Figure 6 with Figures 3(a) and 5(a), it can be seen that the improved EWT algorithm can remove more high-frequency noise than EMD and EWT. Moreover, the reconstructed signal processed by improved EWT is smoother than that processed by EMD or EWT algorithm while the original signal amplitude is well preserved. In order to further quantify the denoising effect of three algorithms, their signal-to-noise ratios are calculated, the results are shown in Table 1.

Figure 4. Relation between the correlation coefficient and the sequence number of the IMF component.

Figure 5. Waveforms of the reconstructed vibration signals: (a) underload, (b) normal load, and (c) overload.
As seen in Table 1, the noises are significantly reduced by the improved EWT algorithm. Moreover, compared with the other algorithms, the denoising effect of the improved EWT algorithm is the best.

### Extraction Load Features

In order to highlight the superiority of the improved EWT-RCMDE feature extraction method, four hybrid methods including improved EWT-MSE, improved EWT-MDE, EWT-RCMDE, and improved EWT-RCMDE are examined by the ball mill shell vibration signals in underload, normal load and overload conditions. RCMDEs in the first 15 scales (i.e., $t_{max} = 15$) of 20 samples in each class (60 samples in total) are computed and taken as initial load features. The characteristic means and standard deviations of the three load states (20 samples in each class) are shown in Figure 7.

As shown in Figure 7, several phenomena are worth to be analyzed. First, all the entropy values have a decreasing trend with the increase of scale factor except for that of underload when $nr \leq 3$. This indicates that the entropy values evaluate the complexity of time series in different scale factors, and the complexity of the coarse-grained time series in each scale factor is different from each other. Second, comparing the improved EWT-RCMDE with EWT-RCMDE, the entropy values of three load states obtained by EWT-RCMDE method fluctuate greatly, and there is a certain fluctuation and overlap area. However, the improved EWT model acquires more useful information from original signal, the entropy values tend to be steadier and more reliable. Third, by comparing the improved EWT-RCMDE, improved EWT-MSE and improved EWT-MDE, it can be seen that the mean entropy values of the three methods have consistent overall trends, and they can distinguish three different load conditions clearly. But the standard deviation of improved EWT-RCMDE is smaller than those of improved EWT-MSE and improved EWT-MDE. The reason is that the RCMDE can reduce the calculation deviation. Therefore, the hybrid method of improved EWT-RCMDE is effective to extract the features from the ball mill shell vibration signal, and it has obvious advantages over others.

### Training and testing

After normalizing the RCMDEs of reconstructed vibration signals into 20-dimensional feature vectors, four-dimensional feature vectors are finally obtained by PCA dimensionality reduction. The four-dimensional feature vectors are entered into the FWA-SVM mill load prediction model, and the material-to-ball ratio and filling rate are the outputs. About 30 groups of vibration signals under different load parameters were selected and processed by the improved EWT-RCMDE method, including 15 groups as training samples (five samples for each of the three loads), and 15 groups as test samples. In order to verify the superiority of FWA-SVM, PSO-SVM and GA-SVM were used for comparison. The initial parameters of the three algorithms are set as: the population number is 20, and the maximum number of iterations is 200. The parameters of SVM

![Figure 6. Vibration signal reconstruction of under working condition 1: (a) EMD reconstruction and (b) EWT reconstruction.](image-url)
model optimized by three algorithms and the recognition results are shown in Table 2 and Figure 8.

As can be seen in Table 2 and Figure 8, after the optimization of SVM parameters by FWA, the time to reach the maximum fitness value is shorter than that of PSO. Furthermore, FWA has higher recognition accuracy rate than PSO and GA (6.60% higher than GA-SVM and 6.60% higher than PSO-SVM). These findings indicate that the fireworks algorithm is more capable of obtaining optimal kernel function parameters than PSO and GA. The average iteration time of the SVM parameters optimized by FWA is only 13.80 s, which is 76.20 s shorter than that of GA. Meanwhile, the PSO has the longest iteration time but does not give the best fitness value. To sum up, the FWA can quickly and accurately optimize SVM parameters.

The above processed training samples and test samples were entered into these three different recognition models to conduct load prediction comparison, and the results are shown in Figures 9 and 10.

As can be seen in Figures 9 and 10, the material-to-ball ratio and filling rate predicted by the FWA-SVM model are close to the real value, and the accuracy of the mill load prediction is higher. In order to further determine the accuracy of mill load prediction, their mean absolute percentage error (MAPE) and root

![Figure 7. Curves of characteristic mean and standard deviation under three load conditions: (a) improved EWT-RCMDE, (b) EWT-RCMDE, (c) improved EWT-MSE, and (d) improved EWT-MDE.](image)

![Table 2. SVM parameters optimized by three algorithms and recognition accuracy of three algorithms.](table)

| Optimization algorithm | Kernel function C | Penalty factor $\sigma$ | Iteration time (s) | Overall recognition accuracy (%) |
|------------------------|-------------------|--------------------------|-------------------|----------------------------------|
| PSO                    | 43.67             | 11.21                    | 127.10            | 86.70                            |
| FWA                    | 36.78             | 9.13                     | 13.80             | 93.30                            |
| GA                     | 65.54             | 24.94                    | 89.60             | 86.70                            |
mean square error (RMSE) were calculated. The results are shown in Table 3.

As can be seen in Table 3, the MAPE of the material-to-ball ratio and filling rate predicted by the FWA-SVM model are 4.19% and 4.77%, respectively, which are lower than those predicted by GA-SVM and PSO-SVM models, respectively. Compared with the GA-SVM model and PSO-SVM model, both the FWA-SVM’s RMSE of the material-to-ball ratio and the filling rate are the smallest among the three models. This indicates that the FWA-SVM model prediction method has higher accuracy and stability.

Conclusion

(1) In view of the diversity and complexity of mill vibration signals, an improved EWT algorithm is proposed. By comparing the decomposition effect of improved EWT with those of EWT and EMD, the effectiveness of improved EWT method is proved. By using the improved EWT to decompose the vibration signal of ball mill and the correlation coefficient method to select the sensitive IMF components, the reconstructed signals can well retain the characteristic information. Moreover, the signal-to-noise ratio is improved by 14.03 dB compared with EMD.

(2) By calculating the average RCMDE of the reconstructed signal under three load states, the results show that the average RCMDE calculated in underload, normal load and overload conditions are quite different, which can better distinguish the different load states of the mill. Compared with MSE and MDE, RCMDE has significantly smaller errors and better recognition results.

(3) The experimental results show that the mill load prediction model based on the FWA-SVM model is more accurate and stable than that based on the GA-SVM and PSO-SVM. The FWA-SVM model has the smallest errors in predicting the material-to-ball ratio and filling rate among these three models while keeps the shortest time to optimize the SVM parameters.
Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was financially supported by the National Natural Science Foundation of China (Grant No. 51464017) and the Education Department of Jiangxi Province (Grant No. GJJ190452).

ORCID iDs
Jiacheng Cai https://orcid.org/0000-0003-1974-0781
Changxi Zeng https://orcid.org/0000-0003-4514-8298

References
1. Tang J, Qiao J, Liu Z, et al. Mechanism characteristic analysis and soft measuring method review for ball mill load based on mechanical vibration and acoustic signals in the grinding process. Miner Eng 2018; 128: 294–311.
2. Huang P, Jia MP and Zhong BL. Investigation on measuring the fill level of an industrial ball mill based on the vibration characteristics of the mill shell. Mater Eng 2009; 14: 1200–1208.
3. Tang J, Zhao L, Zhou J, et al. Experimental analysis of wet mill load based on vibration signals of laboratory-scale ball mill shell. Miner Eng 2010; 23: 720–730.
4. Gaipin C, Lu Z, Xiaoyan L, et al. Research on mill load forecasting based on CEMDAN-cloud model characteristic entropy and LSSVM. Vib Impact 2019; 38: 128–133.
5. Yonggui D, Sisi L, Gaowei Y, et al. Soft measurement of load parameters of wet ball mill based on Manifold Regularization domain. J Chem Eng 2018; 69: 1244–1251.
6. Jian T, Chai T, Wen Y, et al. Engineering modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information. IEEE Trans Auton Sci 2013; 10: 726–740.
7. Jian T, Wen Y, Chai T, et al. Selective ensemble modeling load parameters of ball mill based on multi-scale frequency spectral features and sphere criterion. Mech Syst Signal Proc 2015; 66–67: 485–504.
8. Guo Z, Zhao WG, Lu HY, et al. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012; 37: 241–249.
9. Jinglong C, Zipeng L, Jun P, et al. Wavelet transform based on inner product in fault diagnosis of rotating machinery. A review. Mech Syst Signal Process 2016; 70: 1–35.
10. Dandan P, Zhiliang L, Yaqiang J, et al. Improved empirical mode decomposition based on soft screening stopping criterion and its application in fault diagnosis of rotating machinery. J Mech Eng 2019; 55: 122–132.
11. Chen Y, Zhang T, Luo Z, et al. A novel rolling bearing fault diagnosis and severity analysis method. Appl Sci 2019; 9: 2356.
12. Gilles J. Empirical wavelet transform. IEEE Trans Signal Process 2013; 61: 3999–4010.
13. Chegini SN, Bagheri A and Najafi F. Application of a new EWT-based denoising technique in bearing fault diagnosis. Measurement 2019; 144: 275–297.
14. Liu H, Wu H and Li Y. Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction. Energy Convers Manag 2018; 161: 266–283.
15. Singh O and Sunkaria RK. ECG signal denoising via empirical wavelet transform. Australas Phys Eng Sci Med 2017; 40: 219–229.
16. Li J, Li Y, Li Y, et al. Downhole microseismic signal denoising via empirical wavelet transform and adaptive thresholding. J Geophys Eng 2018; 15: 2469–2480.
17. Humeau-Heurtier A. The multiscale entropy algorithm and its variants: a review. Entropy 2015; 17: 3110–3123.
18. Chi C, Wu J, Huang G, et al. Research on cycle slip detection based on difference morphology filter and singular value entropy. In: 2017 9th Chinese control and decision conference (CCDC), Chongqing, China, 28–30 May 2017, pp.5253–5257. New York: IEEE.
19. Yang W-C, Zhang P-L, Wang H-G, et al. Gear fault diagnosis based on multiscale fuzzy entropy of EEMD. J Vib Shock 2015; 34: 163–167, 187.
20. Sharma S, Tiwari SK and Singh S. Integrated approach based on flexible analytical wavelet transform and permutation entropy for fault detection in rotary machines. Measurement 2021; 169: 108389.
21. Zhang Y and Shang P. Permutation entropy analysis of financial time series based on Hill’s diversity number. Atmos Meas Tech Discuss 2017; 53: 1659–1671.
22. Mehta P, Gaikwad JA and Kulkarni JV. Application of multi-scale fuzzy entropy for roller bearing fault detection and fault classification based on VPMCD. In: 2016 IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), Bangalore, India, 20–21 May 2016, pp.256–261. New York: IEEE.
23. Cai GP, Liu X, Dai CC, et al. Load state identification method for ball mills based on improved EWT, multiscale fuzzy entropy and AEPSO_PNN classification. Processes 2019; 7: 725.
24. Rostaghi M and Azami H. Dispersion entropy: a measure for time series analysis. IEEE Signal Process Lett 2016; 23: 610–614.
25. Azami H, Rostaghi M, Abasolo D, et al. Refined composite multiscale dispersion entropy and its application to biomedical signals. IEEE Trans Biomed Eng 2017; 64: 2872–2879.
26. Conzhi L, Jinde Z, Haiyang P, et al. Rolling bearing fault diagnosis method based on fine composite multi-scale dispersion entropy and support vector machine. China Mech Eng 2019; 30: 1713–1719, 1726.
27. Gomez FR, Rajapakse AD, Annakkage UD, et al. Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. IEEE Trans Power Syst 2011; 26: 1474–1483.
28. Ruirui Z, Jiyin Z, Tingting Z, et al. Power transformer fault diagnosis based on genetic support vector machine and gray artificial immune algorithm. Proc Chinese Soc Electr Eng 2011; 31: 56–63.
29. Qi K, Yang H, Hu Q, et al. A new adaptive weighted imbalanced data classifier via improved support vector
machines with high-dimension nature. *Knowl Based Syst* 2019; 185: 104933.

30. Li K, Wang L, Wu JJ, et al. Using GA-SVM for defect inspection of flip chips based on vibration signals. *Micro-electron. Reliab* 2018; 81: 159–166.

31. Xiaokai W, Shanyue G, Lin H, et al. Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method. *Ultrasonics* 2019; 91: 161–169.

32. Min Z, Zhenyu C and Shanshan B. A fault diagnosis method for rolling bearing based on EEMD-Hilbert and FWA-SVM. *J Southwest Jiaotong Univ* 2019; 54: 633–639.

33. Xiaoyan L, Congcong D, Tiedong C, et al. Load identification method of ball mill based on improved EWT multi-scale entropy and KELM. *CIESC J* 2020; 71: 1264–1277.

### Appendix A: Reviews of DE

For a given univariate time series $X = \{x_1, x_2, \ldots, x_N\}$ with a length of $N$, the computation steps of DE are given as follows.

1. The original time series $X$ is mapped to $Y = \{y_1, y_2, \ldots, y_N\}$ by using normal cumulative distribution function (NCDF) shown as

$$
y_j = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{y_j} e^{-\frac{t^2}{2\sigma^2}} dt \tag{11}
$$

It is obvious that $y_j \in (0, 1)$ and $\sigma$ is standard deviation and $\mu$ is mean of $X$.

2. By using the linear transform, all elements of $Y$ ($y_j, j = 1, 2, \ldots, N$) are mapped to $c$ classes (which is an integer), that is,

$$
z_j = R(c \cdot y_j + 0.5) \tag{12}
$$

where $R$ represents the rounding operation and $z_j$ represents the $j$-th member of the classified time series. Although the step (2) is linear, the whole mapping way is nonlinear for the use of NCDF in step (1).

3. For the given time delay $\delta$ and embedding dimension $m$, $z^{m,c}_j$ can be obtained according to

$$
z^{m,c}_j = \{z^c_1, z^c_{1+\delta}, \ldots, z^c_{j+\delta}, \ldots, z^c_{j+(m-1)d}\}, \quad i = 1, 2, \ldots, N - (m - 1)d.
$$

Each time series $z^{m,c}_j$ can be mapped to a dispersion pattern $\pi^{m,c}_{y_j}$, where $z^c_0 = y_0, z^c_{1+\delta} = y_1, \ldots, z^c_{j+(m-1)d} = y_{m-1}$, The signal has $m$ members, and they all belong to the integer interval from 1 to $c$. The number of possible dispersion patterns is equal to $c^m$.

4. The relative frequency for each dispersion patterns $\pi^{m,c}_{y_j}$ can be estimated by

$$
p(\pi^{m,c}_{y_j}) = \frac{\text{Number of } y_i \text{ s.t. } i \leq N - (m - 1)d, z^{m,c}_j \text{ has type } \pi^{m,c}_{y_j}}{N - (m - 1)d} \tag{13}
$$

where $p(\pi^{m,c}_{y_j})$ stands for the number of dispersion patterns of $\pi^{m,c}_{y_j}$ assigned to $z^{m,c}_j$ divided by the total number of embedded signals for embedding dimension $m$.

5. The DE of $X$ is computed by:

$$
\text{DisEn}(X, m, c, d) = - \sum_{\pi^c = 1} c^m p(\pi^{m,c}_{y_j}) \ln(p(\pi^{m,c}_{y_j})) \tag{14}
$$

It can be found from the algorithm of DE that when all probabilities of distribution patterns $p(\pi^{m,c}_{y_j})$ are equal, DE gets the largest entropy value $\ln(c^m)$ and a typical example is Gaussian white noise. In contrast, when the probability of distribution pattern $p(\pi^{m,c}_{y_j})$ is unitary, that is, only one value is not equal to zero, DE get the smallest value, which indicates that the time series is a completely predictable data and a typical example is the periodic signal with low frequency.

### Appendix B: Abbreviations

| Abbreviation      | Description                                      |
|-------------------|--------------------------------------------------|
| EWT               | empirical wavelet transform                      |
| RCMDE             | refined composite multi-scale dispersion entropy  |
| FWA               | fireworks algorithm                              |
| SVM               | support vector machine                           |
| GA                | genetic algorithm                                |
| PSO               | particle swarm optimization                      |
| EMD               | empirical mode decomposition                     |
| WT                | wavelet transform                                |
| MFE               | multi-scale fuzzy entropy                         |
| AEPSO_PNN         | adaptive evolution particle swarm optimization probabilistic neural network |
| MSE               | multi-scale entropy                              |
| DE                | dispersion entropy                                |
| MF               | intrinsic mode function                          |