

Article

A Recognition and Classification Method for Underground Acoustic Emission Signals Based on Improved CELMD and Swin Transformer Neural Networks

Xuebin Xie *  and Yunpeng Yang

School of Resources and Safety Engineering, Central South University, Changsha 410083, China;
215512090@csu.edu.cn

* Correspondence: xbxie@csu.edu.cn

Abstract: To address the challenges in processing and identifying mine acoustic emission signals, as well as the inefficiency and inaccuracy issues prevalent in existing methods, an enhanced CELMD approach is adopted for preprocessing the acoustic emission signals. This method leverages correlation coefficient filtering to extract the primary components, followed by classification and recognition using the Swin Transformer neural network. The results demonstrate that the improved CELMD method effectively extracts the main features of the acoustic emission signals with higher decomposition accuracy and reduced occurrences of mode mixing and end effects. Furthermore, the Swin Transformer neural network exhibits outstanding performance in classifying acoustic emission signals, surpassing both convolutional neural networks and ViT neural networks in terms of accuracy and convergence speed. Moreover, utilizing preprocessed data from the improved CELMD enhances the performance of the Swin Transformer neural network. With an increase in data volume, the accuracy, stability, and convergence speed of the Swin Transformer neural network continuously improve, and using preprocessed data from the enhanced CELMD yields superior training results compared to those obtained without preprocessing.



Citation: Xie, X.; Yang, Y. A Recognition and Classification Method for Underground Acoustic Emission Signals Based on Improved CELMD and Swin Transformer Neural Networks. *Appl. Sci.* **2024**, *14*, 4188. <https://doi.org/10.3390/app14104188>

Academic Editor: Giuseppe Lacidogna

Received: 23 April 2024

Revised: 10 May 2024

Accepted: 14 May 2024

Published: 15 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: acoustic emission signal recognition; improve CELMD; correlation coefficient; Swin Transformer neural network

1. Introduction

The working environment in underground mines is fraught with potential hazards, particularly as mining depth increases. Underground mines face complex ground pressure environments that can easily trigger various ground pressure disasters [1,2]. These disasters not only threaten the lives of workers but also result in equipment damage and production interruptions [3]. However, through real-time monitoring of rock mass acoustic emission signals, combining the mechanical properties of the rock mass with the operational arrangements of the mine, we can analyze the stress state and activity of the rock mass, identify potential ground pressure issues in advance, and take measures to reduce risks and enhance production efficiency [4,5]. Nevertheless, acoustic emission signals often contain significant amounts of noise, which can disrupt the useful information of signals, leading to misjudgments of ground pressure conditions [6,7]. Therefore, accurate identification and classification of acoustic emission monitoring signal data are crucial [8]. Compared to traditional identification and classification methods used previously, researchers have applied many new techniques, such as machine learning, to the recognition and classification of acoustic emission signals, yielding promising results [9]. However, challenges persist, including low success rates in identification and classification, as well as low efficiency [10].

Acoustic emission signals, often induced by changes in the internal state of surrounding rock or underground operations, represent a typical non-stationary signal [11]. Prior to identification, it is typically necessary to preprocess the signals using mathematical

methods to extract key features [12]. Wavelet transform is capable of capturing the transient characteristics of signals but requires appropriate selection of wavelet basis functions and scales [13]. Fourier transform is widely used for frequency domain analysis of acoustic emission signals but is less effective in capturing transient events compared to wavelet transform [14]. Furthermore, Fourier transform is prone to boundary effects when analyzing signals of finite duration [15]. EMD (Empirical Mode Decomposition) decomposes signals into a series of IMFs (Intrinsic Mode Functions), but during the decomposition process, modal overlap may occur due to interference between different modes [16]. Similarly, if the starting or ending parts of the signal are not recorded, errors in decomposition, known as endpoint effects, can occur [17]. LMD (Local Mean Decomposition), proposed by Jonathan S. Smith and others, decomposes signals into multiple product function components with physical significance, alleviating mode mixing and endpoint effects [18]. However, LMD is prone to errors due to noise interference during the decomposition process [19]. ELMD (Enhanced LMD) and CELMD (Complex Enhanced LMD) optimization methods introduce white noise into the original signal to decompose signals into more suitable frequency bands, albeit with varying degrees of residual white noise [20]. CELMDAN (Complex Enhanced LMD with Adaptive Noise) further reduces the impact of noise on decomposition by significantly increasing the number of iterations, but it suffers from the issue of high computational complexity [21].

The method of identification and classification of acoustic emission signals has a significant impact on classification accuracy [22]. Traditional methods primarily rely on manual identification, with researchers utilizing the spectral features of signals and employing methods such as fractal theory, chaos models, and others to assist with signal identification [23]. However, the accuracy of these methods is limited by the experience and expertise of the operator, resulting in generally lower accuracy. In recent years, the development of deep learning has provided new possibilities for the classification of acoustic emission signals. Researchers applied BP neural networks to the classification of acoustic emission signals [24], while other researchers introduced CNN (convolutional neural network), with both achieving favorable recognition rates [25]. Transformer is a neural network model proposed by Google based on self-attention mechanisms, which has achieved great success in natural language processing [26]. Dosovitskiy et al. applied the Transformer neural network to image classification, proposing the ViT (Vision Transformer) neural network, which broke the monopoly of convolutional neural networks in classification tasks [27]. Swin Transformer, based on ViT, optimized attention mechanisms and sampling effects, and learned the hierarchical structure of convolutional neural networks, significantly improving training speed and accuracy [28]. Researchers have applied it to the identification of forest fires, achieving rapid and real-time monitoring of forest fires [29]. Additionally, it has assisted medical professionals in achieving precise classification of tumors, and has also achieved notable results in other fields [30].

Therefore, taking a metal mine in Guangxi Province as an example, this paper applies the Swin Transformer neural network to the classification of underground acoustic emission signals. Firstly, the improved CELMD method is used for the preprocessing of acoustic emission signals, and then the Swin Transformer is employed for the recognition and classification of the processed signals. In this way, the influence of data preprocessing methods and the selection of neural networks on the identification effect of acoustic emission signals is studied, and factors affecting the performance of the Swin Transformer are explored.

2. Basic Principles

2.1. The Empirical Optimal Envelope Method

In the LMD method, the commonly used sliding average method can lead to distortion in constructing envelope lines, prompting researchers to delve into the issue and make some progress. For example, it employs cubic spline interpolation instead of sliding averages and combines extreme point fitting techniques from the EMD method to compute envelope lines, thereby obtaining local mean curves and envelope estimation curves [31].

This method requires only one interpolation calculation to obtain the two required curve functions, avoiding the cumbersome process of sliding averaging and thus improving the computational efficiency of local mean decomposition in signal analysis. However, when applying cubic spline interpolation to analyze strongly non-stationary signals, issues such as “over-enveloping” and “under-enveloping” may arise, affecting the accuracy of signal analysis. To address this problem, scholars have proposed using monotone cubic Hermite interpolation as a replacement for cubic spline interpolation and have achieved some success. However, in practical applications, it has been found that for some vibration signals with strong non-stationary characteristics, using cubic spline interpolation or cubic Hermite interpolation to construct envelope lines may lead to distorted decomposition results [32].

Therefore, this paper adopts an empirical optimal envelope method based on Hermite interpolation, which utilizes Hermite interpolation to obtain the optimal envelope curve of the signal, and subsequently obtains local mean functions and envelope estimation functions. The main steps include:

1. First item; given a signal $x(t)$, identify all of its local maxima and minima points. Use Hermite interpolation to interpolate the local maxima points of the signal to obtain the initial upper envelope curve $c_{+0}(t)$. Similarly, obtain the initial lower envelope curve $c_{-0}(t)$.
2. Second item; utilize the upper envelope curves $c_{+0}(t)$ and lower envelope curves $c_{-0}(t)$ to calculate the upper distance function $e_{+0}(t)$ and lower distance function $e_{-0}(t)$ of the signal.

$$\begin{cases} e_{+0}(t) = x(t) - c_{+0}(t) \\ e_{-0}(t) = c_{-0}(t) - x(t) \end{cases} \quad (1)$$

3. Third item; employ Hermite interpolation to interpolate the sequences of local maxima points of the upper envelope curves $e_{+0}(t)$ and lower envelope curves $e_{-0}(t)$ separately, obtaining the first iteration upper envelope curves $c_{+1}(t)$ and lower envelope curves $c_{-1}(t)$.
4. Fourth item; repeat the second and third items until all of the local maxima points of the upper distance function $e_{+n}(t)$ of the n th iteration upper envelope curve $c_{+n}(t)$ equal zero, resulting in the empirical optimal upper envelope curve. Similarly, obtain the empirical optimal lower envelope curve $c_{-n}(t)$.
5. Fifth item; compute the local mean function $f_{11}(t)$ and the optimal envelope estimation curve $g_{11}(t)$.

$$\begin{cases} f_{11}(t) = \frac{c_{+n}(t) + c_{-n}(t)}{2} \\ g_{11}(t) = \frac{c_{+n}(t) - c_{-n}(t)}{2} \end{cases} \quad (2)$$

2.2. Improve CELMD

Compared to the basic CELMD algorithm, the improved CELMD utilizes the empirical optimal envelope method to compute the envelope curve instead of the sliding average method, the empirical optimal envelope method has been applied by researchers to obtain envelope curves in EMD, achieving favorable results. However, it has not yet been applied to LMD. Furthermore, it employs mirror extension and endpoint correction methods to further reduce the endpoint effect, these methods were often used separately in the past. Additionally, it utilizes the correlation coefficient to filter out the main components [33]. The specific steps are as follows:

1. First item; add white noise signals a_1 with an amplitude ratio coefficient α to the original signal $X(t)$, and perform endpoint correction and mirror extension to obtain the processed signal.

$$X_1(t) = X(t) + \alpha a_1 \quad (3)$$

2. Second item; utilize the empirical optimal envelope method to obtain the empirical optimal upper envelope curves $c_{+n}(t)$ and lower envelope curves $c_{-n}(t)$ of the signal, and calculate the local mean function $f_1(t)$ and local envelope function $g_1(t)$.

$$\begin{cases} f_1(t) = \frac{c_{+n}(t) + c_{-n}(t)}{2} \\ g_1(t) = \frac{c_{+n}(t) - c_{-n}(t)}{2} \end{cases} \quad (4)$$

3. Third item; demodulate the signal $X_1(t)$ to obtain a frequency-modulated signal $s_n(t)$, treating it as the original signal, and repeat the second item until the function becomes a pure frequency-modulated signal.

$$s_n(t) = \frac{X_1(t) - f_n(t)}{g_n(t)} \quad (5)$$

4. Fourth item; obtain the first PF (Product Function) component of the signal.

$$PF_1(t) = g_1(t)g_2(t) \cdots g_n(t)s_n(t) \quad (6)$$

5. Fifth item; separate the PF component from the original signal and repeat items 2–4 until all PF components and residual components of the signal are obtained.

$$\begin{cases} X_{11}(t) = X_1(t) - PF_1(t) \\ X_1(t) = \sum_{i=1}^n PF_i(t) + u_1(t) \end{cases} \quad (7)$$

6. Sixth item; repeat items 1–5 for an even number of rounds. In each round, the added white noise has opposite signs between consecutive rounds. Obtain the corresponding PF components and residual components for each round.
7. Seventh item; average the PF components and residual components obtained from multiple rounds of decomposition. Select the main components based on correlation coefficients and add them together to obtain the final signal.

2.3. Swin Transformer

The Swin Transformer neural network architecture is specifically designed for computer vision tasks, particularly image classification and object detection. In recent years, the Swin Transformer neural network has been widely applied in fields such as medicine, transportation, and fire prevention, improving the efficiency of identifying objects such as tumors, fires, and abnormal vehicles. However, its application in the field of signal recognition, especially in acoustic emission signal recognition, is relatively limited. Its overall architecture, as shown in Figure 1, consists of four stages, each containing similar repeating units. Firstly, the Patch Partition divides the input RGB (Red, Green, Blue) image of dimension $H(\text{High}) \times W(\text{Width}) \times 3$ into non-overlapping patches of equal size. Each patch is treated as a token and serves as the effective input sequence length for the Swin Transformer.

The first stage consists of a Linear Embedding and a Swin Transformer Block. It projects the input sequence to an arbitrary dimension and conducts feature learning using the Swin Transformer Block. In the remaining three stages, Patch Merging is used instead of Linear Embedding. Its function is to improve sampling accuracy, i.e., to reduce the range of each token while increasing the number of channels. This allows the Swin Transformer Block to learn more features. After four stages of downsampling and feature learning, the classification results are finally output through global pooling and fully connected layers [34,35].

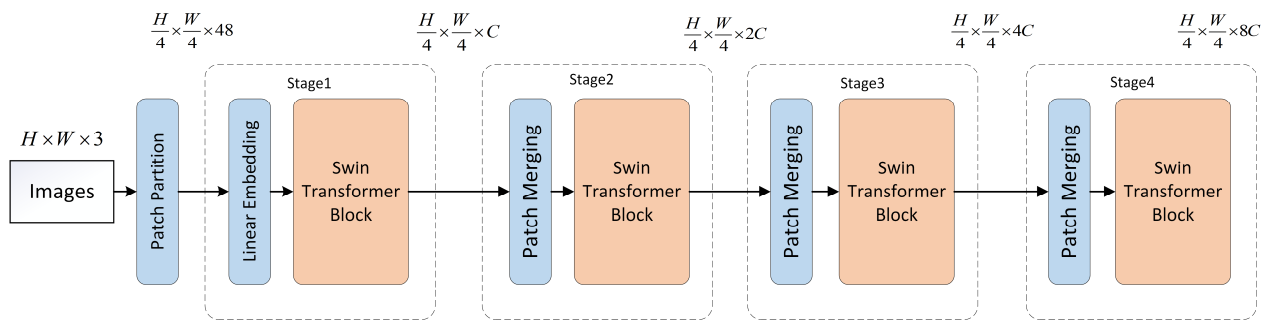


Figure 1. Swin Transformer Network Architecture.

The structure of the Swin Transformer Block is illustrated in Figure 2. It mainly consists of LN (Layer Normalization), W-MSA (Window-based Multi-head Self-Attention), SW-MSA (Shifted Window-based Multi-head Self-Attention), MLP (Multi-Layer Perceptron), and residual connections.

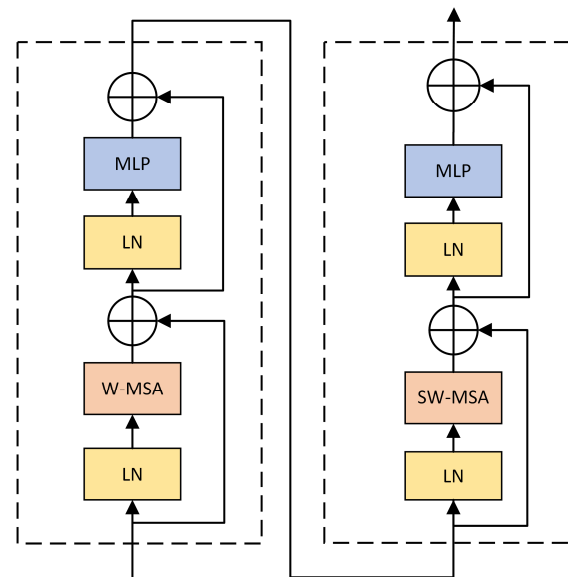


Figure 2. Swin Transformer Block structure.

Generally, blocks within each stage comprise an even number of similar Swin Transformer Blocks. The main difference between Swin Transformer Blocks lies in whether SW-MSA or W-MSA is used. The input features are first normalized by LN, followed by feature learning using W-MSA or SW-MSA. Residuals are then computed, followed by normalization again. Finally, output features are obtained through MLP and residual calculations.

Swin Transformer has several variants according to the differences in the number of Swin Transformer Blocks (B) in each stage, the number of heads (H) in the multi-head self-attention layer, and other parameters. In this paper, we utilized Swin T (Tiny), Swin S (Small), Swin B (Base), and Swin L (Large). The main structures corresponding to each model are summarized in Table 1.

Table 1. Swin Transformer subspecies network architecture.

Model	Swin T	Swin S	Swin B	Swin L
stage1	2B, 3H	2B, 3H	2B, 4H	2B, 6H
stage2	2B, 6H	2B, 6H	2B, 8H	2B, 12H
stage3	6B, 12H	18B, 12H	18B, 16H	18B, 24H
stage4	2B, 24H	2B, 24H	2B, 32H	2B, 48H

2.4. Model Building

The paper proposes an underground acoustic emission signal recognition and classification method based on improved CELMD and Swin Transformer neural networks. The method mainly consists of three stages: acoustic emission data collection and partitioning, data preprocessing, and neural network training and classification. The flowchart depicting the process is illustrated in Figure 3.

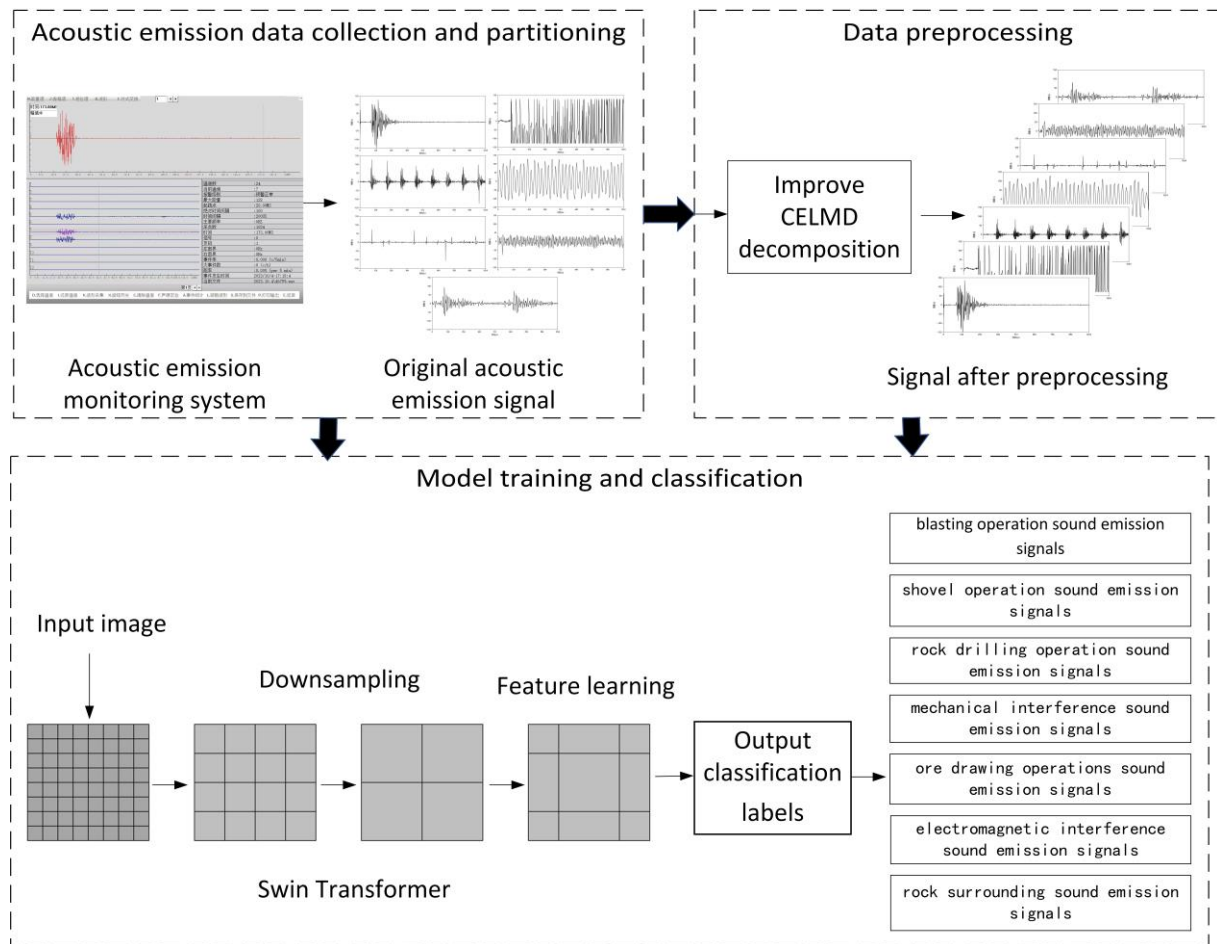


Figure 3. Identification and classification model of mine acoustic emission signals.

3. Application Instance

3.1. Ground Pressure Monitoring System

This study takes the ground pressure monitoring system of Zhongjin Lingnan Panlong Lead-Zinc Mine in Guangxi as the engineering background to verify the effectiveness and feasibility of the proposed method. The network structure of the ground pressure monitoring system, as illustrated in Figure 4, consists of 24 acoustic emission monitoring channels, a sampling interval of 200 microseconds, and a signal generated every 1024 sampling points.

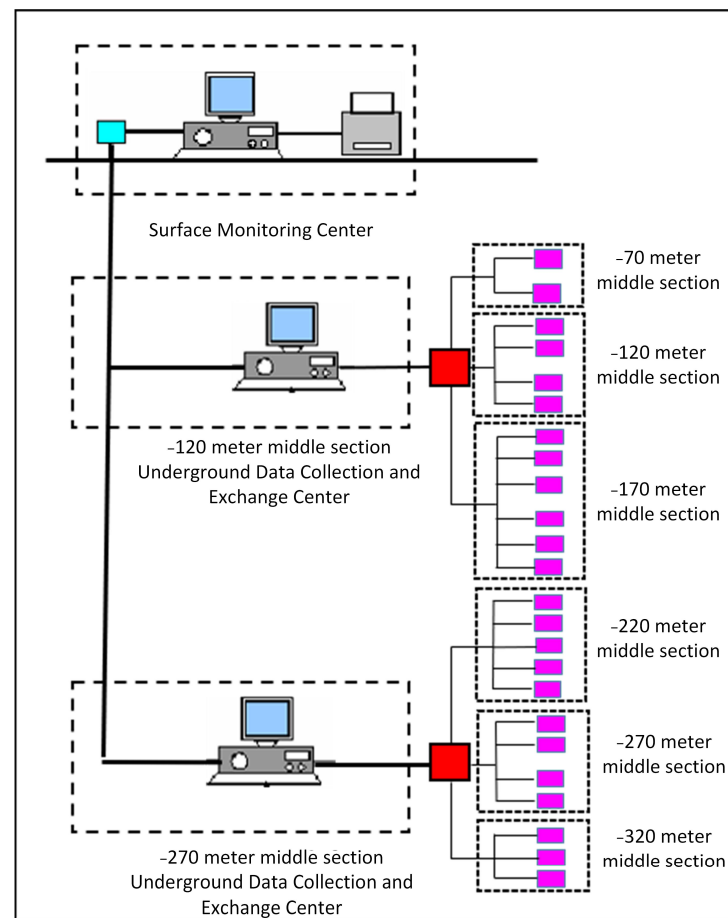


Figure 4. Schematic diagram of ground pressure monitoring system.

3.2. Data Collection

The data used in this study were obtained from the 24-channel acoustic emission monitoring system at the Panlong lead-zinc mine in Guangxi. A significant amount of acoustic emission signals was collected from routine monitoring conducted between June and October 2022. Invalid signals and those with excessive repetition were excluded, resulting in a selection of 35,000 original acoustic emission signals categorized into seven types. Each type comprises approximately 5000 signals and includes signals from BO (blasting operation sound emission signals), SO (shovel operation sound emission signals), RD (rock drilling operation sound emission signals), MI (mechanical interference sound emission signals), OD (ore drawing operations sound emission signals), EI (electromagnetic interference sound emission signals), and RS (rock surrounding sound emission signals).

The typical signal waveforms are illustrated in Figure 5; however, some signals were not fully recorded because their length exceeded the sampling length of the monitoring system. Furthermore, each type of signal has its own characteristics and patterns. Among them, the frequency of SO signals is the lowest, and they exhibit strong persistence. This is because shovel operations are typically carried out automatically by machinery and proceed relatively slowly and continuously. This simplicity makes the identification of shovel operation signals relatively straightforward. On the other hand, BO signals are more complex, often containing both high and low frequency components simultaneously. Additionally, their frequency and amplitude may exceed the threshold of the monitoring system, resulting in a slight distortion of the monitored signals. However, the higher amplitude is also beneficial for identification purposes. The patterns of EI signals are the least obvious, exhibiting greater randomness.

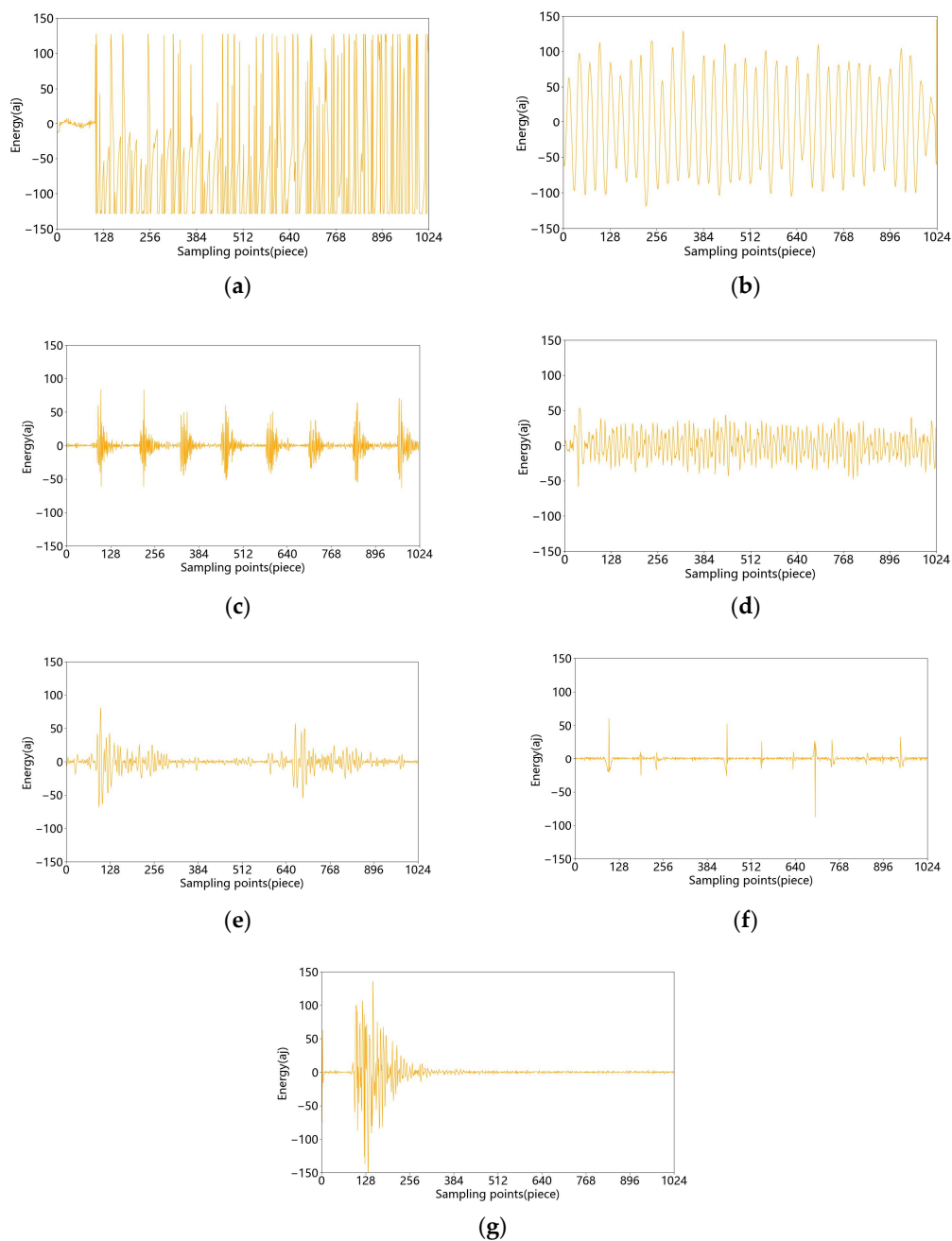


Figure 5. Typical acoustic emission waveform signal: (a) BO; (b) SO; (c) RD; (d) MI; (e) OD; (f) EI; (g) RS.

3.3. Parameter Settings

The three parameters required for the improved CELMD algorithm include the proportion coefficient of adding white noise, the number of times white noise is added, and the correlation coefficient threshold for filtering the main components.

The proportion coefficient α is the ratio of the standard deviation of the white noise signal to the standard deviation of the original signal. A larger proportion coefficient α increases the likelihood of generating false PF components, while a smaller coefficient may affect the effectiveness of reducing mode mixing. According to the method proposed by the researchers, the value of the proportional coefficient is determined to be between 0.05

and 0.8 for various types of signals, as shown in the following formula. In this article, after calculating the signal data used, we used 0.2 [36].

$$\frac{\sigma_h}{4\sigma_0} < \alpha < \frac{3\sigma_h}{4\sigma_0} \quad (8)$$

In the formula, σ_h represents the standard deviation of high-frequency components in the signal under analysis, and σ_0 represents the standard deviation of the signal under analysis.

In theory, the more times white noise is added, the fewer white noise components are included in the final PF components. However, in simulations, it was found that when the number of noise additions is too high, the improvement in decomposition is minimal but the computational time significantly increases. Therefore, the number of times white noise is added is set to 100.

After decomposition, some PF components may not contribute significantly to the overall characteristics. Therefore, it is common to filter the PF components first. Common filtering parameters include correlation coefficients, variance contribution rates, and spectral coefficients. Researchers have demonstrated that these parameters are positively correlated [37]. This paper adopts the correlation coefficient as the filtering parameter. Generally, a correlation coefficient less than 0.5 indicates weak correlation between the PF component and the signal. Therefore, the threshold is set to 0.5.

The learning rate of the Swin Transformer neural network is set to 0.0001, and the epoch is set to 50. In addition, ViT neural network and convolutional neural network are added as control groups, with the same learning rate and epoch as Swin Transformer. The ratio of training set to validation set is set to 8:2. Through experiments, it has been verified that the parameter settings can make the neural network training achieve convergence and meet the requirements. Official pre-training weights are used to accelerate training, and parameters such as success rate, loss rate, time, and confusion matrix are recorded for each training round.

3.4. Data Preprocessing

Taking the collected rock mass acoustic emission signals as an example, the upper and lower envelope curves of the rock mass acoustic emission signals are obtained using the sliding average method, the cubic spline interpolation method, and the optimal empirical envelope method. For the sake of presentation, only the data of the first 400 sampling points in the results are taken. The results are shown in Figures 6–8. The envelope curves constructed by the optimal empirical envelope method exhibit a smooth trend, closely matching the original curve, and almost no over-enveloping or under-enveloping phenomena are observed. On the other hand, both the cubic spline interpolation method and the sliding average method exhibit more pronounced over-enveloping phenomena, with the cubic spline interpolation method also displaying under-enveloping issues. Hence, it can be observed that the optimal empirical envelope method can replace the sliding average method for envelope curve determination.

Taking the collected rock mass acoustic emission signals as an example, both the CELMD and improved CELMD methods were employed for decomposition. The decomposition results are shown in Figures 9 and 10. From the figures, it can be observed that CELMD decomposes a larger number of PF components, but the later components contain more false components, resulting in poor decomposition accuracy. Moreover, noticeable endpoint effects are evident at the left endpoint, and components PF3 to PF5 exhibit significant mode mixing phenomena. In contrast, the decomposition results of the improved CELMD method show almost no false components, and the components of different frequency signals are well decomposed. The influence of white noise is effectively eliminated, and no abnormal amplitude or frequency is observed at the endpoints. There are fewer occurrences of endpoint effects and mode mixing compared to the CELMD method, indicating the significant advantages of the improved CELMD method.

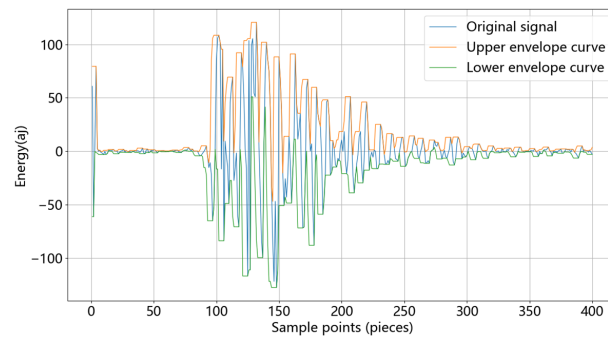


Figure 6. Signal envelope curve constructed by the sliding average method.

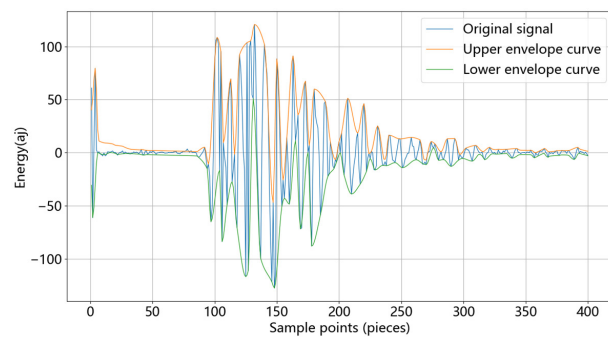


Figure 7. Signal envelope curve constructed by the cubic spline interpolation method.

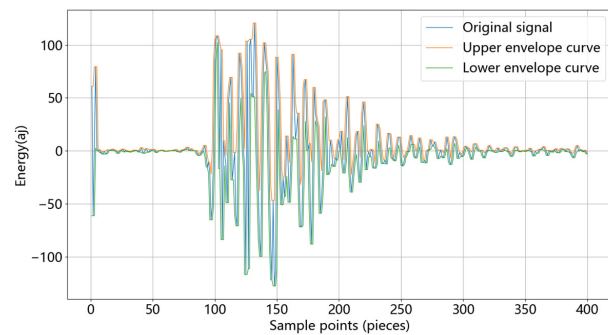


Figure 8. Signal envelope curve constructed by the empirical optimal envelope method.

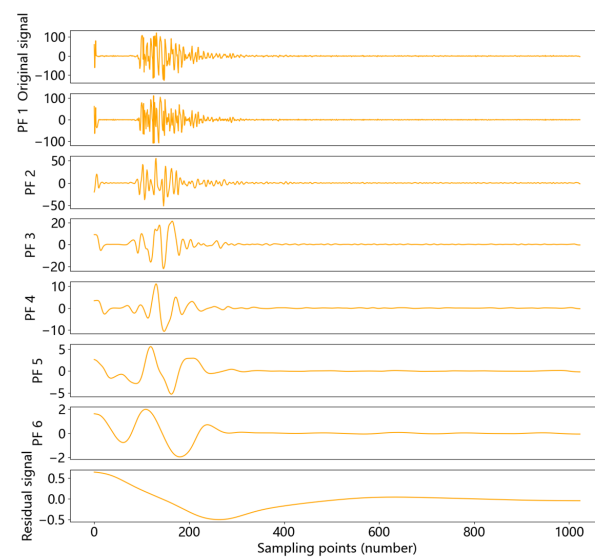


Figure 9. CELMD decomposition results of the RS signal.

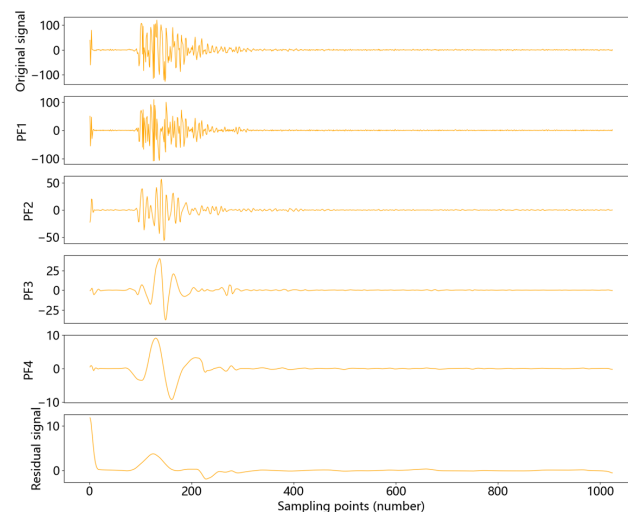


Figure 10. Improved CELMD decomposition results of the RS signal.

4. Classification Results and Discussion

The unprocessed data and the data preprocessed using the improved CELMD method were separately applied to train CNN, ViT networks, and four subtypes of Swin Transformer networks for comparative analysis. The training results of the unprocessed data for each model are presented in Table 2, while the training results of the preprocessed data for each model are shown in Table 3.

Table 2. Training effectiveness of neural networks with unprocessed data.

Model	Accuracy (%)	Model Size (M)	Training Speed (s)	Inference Speed (s)	Top1	Top7
CNN	96.62	9	0.013	0.012	99.04	89.08
ViT	95.92	34	0.005	0.006	98.07	90.05
Swin T	97.89	11	0.007	0.004	99.00	91.28
Swin S	97.93	19	0.049	0.016	99.61	92.50
Swin B	97.78	34	0.019	0.007	98.87	91.48
Swin L	97.45	76	0.083	0.019	99.52	90.67

Table 3. Training effectiveness of neural networks with improved CELMD preprocessing data.

Model	Accuracy (%)	Model Size (M)	Training Speed (s)	Inference Speed (s)	Top1	Top7
CNN	97.57	9	0.010	0.010	99.27	87.63
ViT	97.03	34	0.005	0.006	98.89	89.52
Swin T	98.62	11	0.007	0.004	99.52	95.69
Swin S	98.46	19	0.038	0.010	99.84	95.60
Swin B	98.06	34	0.019	0.007	99.08	95.94
Swin L	98.32	76	0.045	0.011	99.52	93.08

From Tables 2 and 3, it can be observed that the comprehensive accuracy of all six models is above 95%. Among them, the accuracy of all four subtypes of Swin Transformer networks is higher than that of convolutional neural networks and ViT networks. Although the accuracy of the models is already at a high level, preprocessing the data still improves the accuracy of various models. In terms of model size, convolutional neural networks and Swin T networks are relatively lightweight. In particular, the Swin T model not only has high accuracy on a lightweight basis but also exhibits excellent performance in inference

speed. While its training speed is only surpassed by the ViT network, the other subtypes of Swin Transformer networks, compared to the Swin T network, have deeper network structures, resulting in varying degrees of slowdown in training and inference speeds. However, there is no significant improvement in accuracy. This may be partly due to the fact that the accuracy is already at a high level, partly because the acoustic emission waveform data are relatively simple compared to other complex data types. The difficulty in extracting features from simple data is not significant, and compared to the convolutional operations of CNN and the W-MSA mechanism of ViT, the downsampling mechanism and SW-MSA mechanism of Swin Transformer are more adept at handling complex data and uncovering deep-level feature information. Therefore, only limited accuracy improvements have been achieved.

In terms of best classification category performance (Top1), the four subtypes of Swin Transformer exhibit only a slight advantage. However, in terms of worst classification category performance (Top7), Swin Transformer networks outperform convolutional neural networks and ViT networks, especially when trained with data preprocessed using the improved CELMD method, the performance is significantly better than that of other networks.

4.1. Analysis of Classification Accuracy of Swin Transformer Model

Figure 11 presents the confusion matrix results for the four different subtypes of Swin Transformer on the entire dataset, with recognition accuracies exceeding 93% for all categories. The categories with the highest classification accuracy are rock drilling operation acoustic emission signals and mechanical interference acoustic emission signals, with classification accuracies exceeding 98% for all four subtypes, reaching as high as 99.52% and 99.39%, respectively. The categories with poorer classification accuracy are blasting operation acoustic emission signals, ore dumping operation acoustic emission signals, and surrounding rock acoustic emission signals, with the lowest accuracies being 95.69%, 93.08%, and 95.94%, respectively. It is worth noting that among these categories with lower accuracy, the majority of misclassified ore dumping signals are predicted as surrounding rock acoustic emission signals, and the proportion of surrounding rock signals misclassified as electromagnetic interference signals is the highest among the misclassifications of surrounding rock signals. Additionally, the misclassification categories of blasting signals are mainly mechanical interference signals.

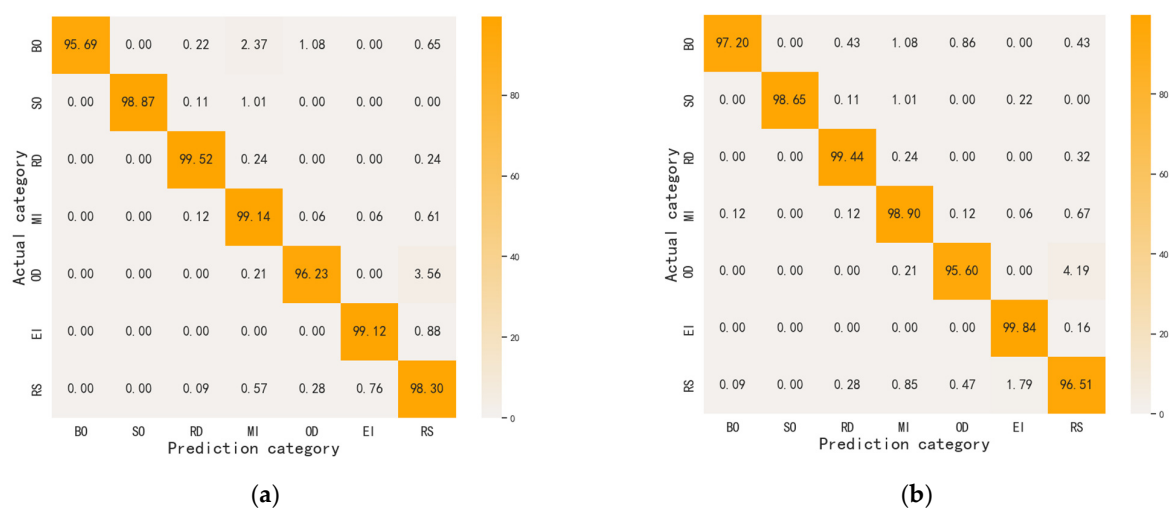


Figure 11. Cont.

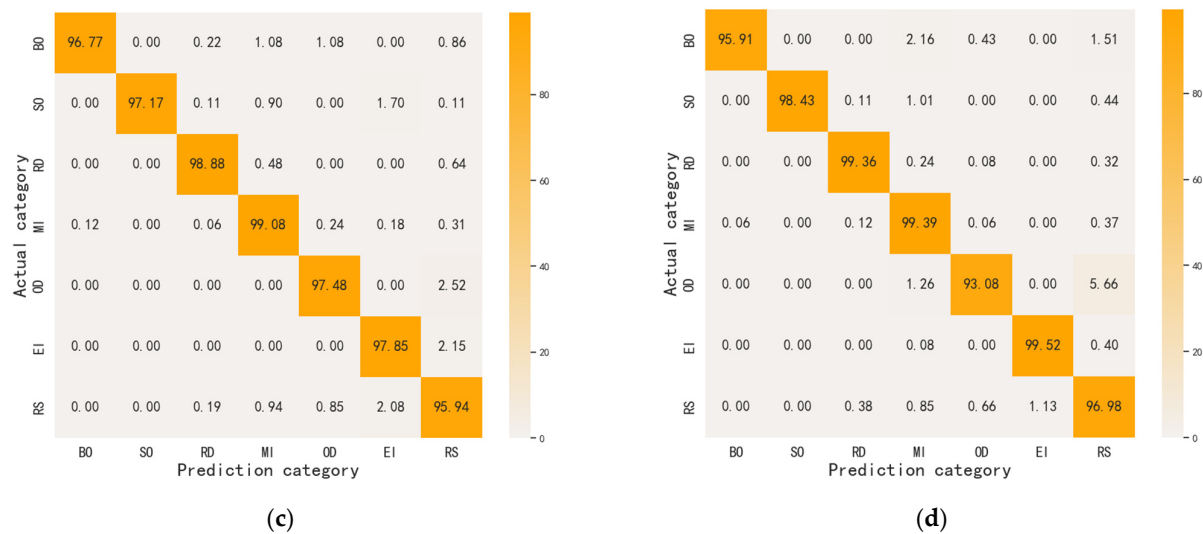


Figure 11. Swin Transformer neural network confusion matrix: (a) Swin T model confusion matrix; (b) Swin S model confusion matrix; (c) Swin B model confusion matrix; (d) Swin L model confusion matrix.

4.2. Comparison of Training Results with Different Data Volumes

Using Python 3.9 scripts, the dataset was randomly split, ensuring a roughly equal distribution of samples for each class. Training and prediction were conducted using dataset sizes of 2500; 5000; 10,000; 20,000; and 35,000. The recognition results of each model without utilizing the improved CELMD preprocessing are presented in Figure 12, while those with improved CELMD preprocessing are depicted in Figure 13.

The results indicate an overall upward trend in model performance with increasing dataset sizes. CNN and ViT network exhibit greater sensitivity to dataset size, showing significant increases in accuracy as the dataset grows. In contrast, Swin Transformer neural networks demonstrate a more gradual improvement trend, yet consistently outperform CNN and ViT network at each dataset size. However, this does not necessarily imply that the learning capacity of these two networks is stronger than that of the Swin Transformer. On the contrary, this phenomenon arises because the Swin Transformer network has a stronger ability to extract primary features than CNN and ViT neural networks. This enables it to learn more feature information from a smaller amount of data, thanks to its unique downsampling mechanism and SW-MSA mechanism, which outperform CNN's convolutional operations and ViT's W-MSA mechanism.

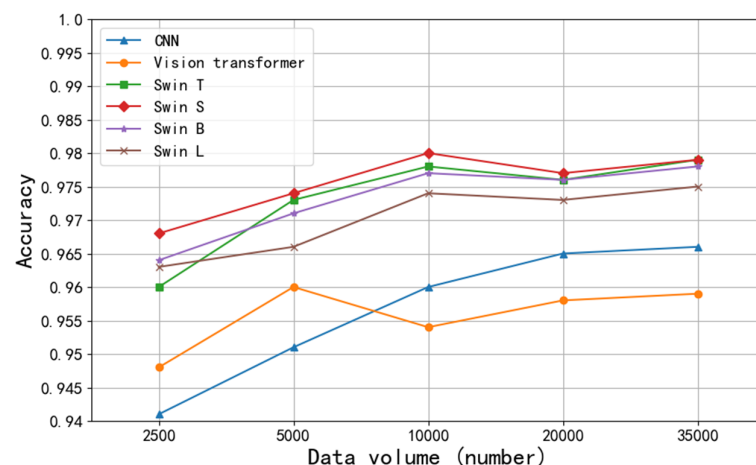


Figure 12. The trend of model accuracy changes with different volumes of unprocessed data.

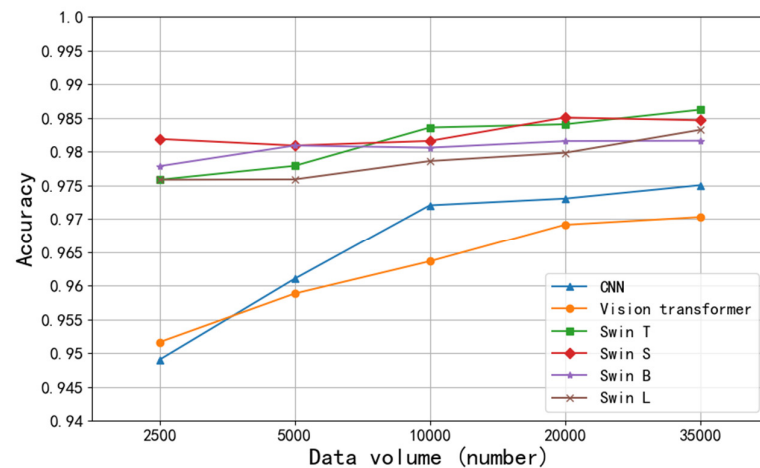


Figure 13. The trend of model accuracy changes with different volumes of preprocessed data.

4.3. Comparison of Convergence Speeds of Different Models

In neural network training, the loss rate represents the disparity between the model's predicted values and the actual values on the training data, serving as a crucial metric for assessing model performance. The change in loss rate reflects the convergence speed and training effectiveness of the model. ViT and Swin T neural networks share similar structures and both compute loss rates using cross-entropy, making them suitable for comparison. The variation of loss rates for these two models under different data preprocessing methods and dataset sizes across training epochs is illustrated in Figures 14 and 15.

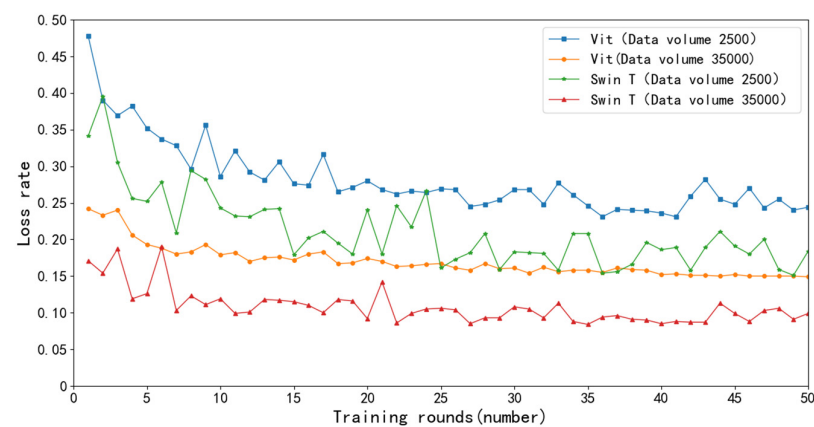


Figure 14. The trend of model loss rate changes with different volumes of unprocessed data.

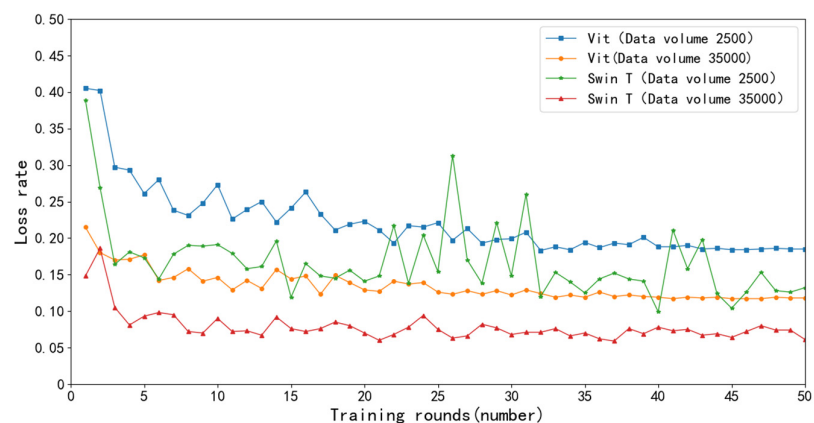


Figure 15. The trend of model loss rate changes with different volumes of preprocessed data.

For the Swin T network, the loss rate generally decreases with increasing training epochs, with lower and more stable loss rates observed under larger dataset sizes. This suggests that increasing the dataset size can accelerate the convergence speed and improve the training effectiveness of the Swin T network. Moreover, training with data preprocessed using the improved CELMD method results in smaller loss rates and faster convergence speed compared to direct training. In comparison to the ViT neural network, regardless of the dataset size or the use of improved CELMD preprocessing, the Swin T network consistently exhibits lower loss rates and faster convergence speed.

5. Conclusions

To address the challenges of low accuracy and difficulty in identifying acoustic emission signals, we propose a multi-classification method for mine acoustic emission signal waveforms based on the improved CELMD method and Swin Transformer neural network. This method aims to achieve intelligent and high-accuracy identification and classification of acoustic emission signals in mines.

The improved CELMD method first utilizes the empirical optimal envelope method instead of the traditional sliding average method to extract the main features of the signal. Furthermore, by incorporating methods such as mirror extension and endpoint correction, it further enhances the decomposition effect and reduces the occurrence of mode mixing and endpoint effects commonly encountered during signal decomposition.

Compared to convolutional neural networks and ViT neural networks, Swin Transformer neural networks demonstrate higher accuracy and convergence speed. Furthermore, utilizing data preprocessed with the improved CELMD method enhances the performance of Swin Transformer neural networks.

With an increase in dataset size, the accuracy, stability, and convergence speed of Swin Transformer neural networks continue to improve. Training with data preprocessed using the improved CELMD method yields superior results compared to training without preprocessing.

Author Contributions: Data curation, X.X.; Formal analysis, X.X.; Methodology, Y.Y.; Resources, X.X. and Y.Y.; Software, Y.Y.; Writing—original draft, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. If any researcher is in need of the data and codes, email: xbxie@csu.edu.cn. The data are not publicly available due to Laboratory's private data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, Y.; Liu, H.; Su, L.; Chen, S.; Zhu, X.; Zhang, P. Developmental Features, Influencing Factors, and Formation Mechanism of Underground Mining-Induced Ground Fissure Disasters in China: A Review. *Int. J. Environ. Res. Public Health* **2023**, *20*, 3511. [\[CrossRef\]](#)
2. Zhao, Y.; Yang, T.; Liu, H.; Wang, S.; Zhang, P.; Jia, P.; Wang, X. A path for evaluating the mechanical response of rock masses based on deep mining-induced microseismic data: A case study. *Tunn. Undergr. Space Technol.* **2021**, *115*, 104025. [\[CrossRef\]](#)
3. Chen, L.; Zhang, D.; Fan, G.; Zhang, S.; Wang, X.; Zhang, W. A New Repeated Mining Method With Preexisting Damage Zones Filled for Ultra-Thick Coal Seam Extraction—Case Study. *Front. Earth Sci.* **2022**, *10*, 835867. [\[CrossRef\]](#)
4. Cai, M.; Lai, X. Monitoring and analysis of nonlinear dynamic damage of transport roadway supported by composite hard rock materials in Linglong gold mine. *Int. J. Miner. Metall. Mater.* **2003**, *10*, 10–15.
5. Di, Y.; Wang, E.; Huang, T. Identification method for microseismic, acoustic emission, and electromagnetic radiation interference signals of rock burst based on deep neural networks. *Int. J. Rock Mech. Min. Sci.* **2023**, *170*, 105541. [\[CrossRef\]](#)
6. Liu, H.; Zhang, X. Predictive analysis of impact hazard level of coal rock mass based on fuzzy inference network. *J. Intell. Fuzzy Syst.* **2020**, *38*, 1509–1518.

7. Xie, X.; Li, S.; Guo, J. Study on multiple fractal analysis and response characteristics of acoustic emission signals from goaf rock bodies. *Sensors* **2022**, *22*, 2746. [\[CrossRef\]](#)
8. Di, Y.; Wang, E.; Li, Z.; Liu, X.; Huang, T.; Yao, J. Comprehensive early warning method of microseismic, acoustic emission, and electromagnetic radiation signals of rock burst based on deep learning. *Int. J. Rock Mech. Min. Sci.* **2023**, *170*, 105519. [\[CrossRef\]](#)
9. Beijer, A.; Piasecki, T. The use of acoustic emission elastic waves for diagnosing high pressure mud pumps used on drilling rigs. *Energies* **2020**, *13*, 1138. [\[CrossRef\]](#)
10. Wang, T.; Liu, Z.; Liu, L. Investigating a three-dimensional convolution recognition model for acoustic emission signal analysis during uniaxial compression failure of coal. *Geomat. Nat. Hazards Risk* **2024**, *15*, 2322483. [\[CrossRef\]](#)
11. Zhang, Y.; Zhang, Y.; Jia, X.; Li, H.; Tang, S. Feature extraction of mine water inrush precursor. *IEEE Access* **2020**, *8*, 163255–163259. [\[CrossRef\]](#)
12. Tai, J.; Liu, C.; Wu, X.; Yang, J. Bearing fault diagnosis based on wavelet sparse convolutional network and acoustic emission compression signals. *Math. Biosci. Eng.* **2022**, *19*, 8057–8080. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Zhao, H.; Li, Z.; Zhu, S.; Yu, Y. Valve internal leakage rate quantification based on factor analysis and wavelet-BP neural network using acoustic emission. *Appl. Sci.* **2020**, *10*, 5544. [\[CrossRef\]](#)
14. Khoshouei, M.; Bagherpour, R.; Jalalian, M.H. Rock type identification using analysis of the acoustic signal frequency contents propagated while drilling operation. *Geotech. Geol. Eng.* **2022**, *40*, 1237–1250. [\[CrossRef\]](#)
15. Di, Y.; Wang, E.; Li, Z.; Liu, X.; Li, B. Method for EMR and AE interference signal identification in coal rock mining based on recurrent neural networks. *Earth Sci. Inform.* **2021**, *14*, 1521–1536. [\[CrossRef\]](#)
16. Xu, C.; Du, S.; Gong, P.; Li, Z.; Chen, G.; Song, G. An improved method for pipeline leakage localization with a single sensor based on modal acoustic emission and empirical mode decomposition with Hilbert transform. *IEEE Sens. J.* **2020**, *20*, 5480–5491. [\[CrossRef\]](#)
17. Li, B.; Wang, E.; Shang, Z.; Li, Z.; Li, B.; Liu, X.; Wang, H.; Niu, Y.; Wu, Q.; Song, Y. Deep learning approach to coal and gas outburst recognition employing modified AE and EMR signal from empirical mode decomposition and time-frequency analysis. *J. Nat. Gas Sci. Eng.* **2021**, *90*, 103942. [\[CrossRef\]](#)
18. Cheng, Y.; Zou, D. Complementary ensemble local means decomposition method and its application to rolling element bearings fault diagnosis. *Proc. Inst. Mech. Eng. Part O-J. Risk Reliab.* **2019**, *233*, 868–880. [\[CrossRef\]](#)
19. Xiang, L.; Yan, X. A self-adaptive time-frequency analysis method based on local mean decomposition and its application in defect diagnosis. *J. Vib. Control* **2016**, *22*, 1049–1061. [\[CrossRef\]](#)
20. Li, X.; Ma, J.; Wang, X.; Wu, J.; Li, Z. An improved local mean decomposition method based on improved composite interpolation envelope and its application in bearing fault feature extraction. *ISA Trans.* **2020**, *97*, 365–383. [\[CrossRef\]](#)
21. Hao, Y.; Du, Z.; Xing, Z.; Jiang, J.; Yang, K.; Ni, L.; Yan, X. Urban hazardous chemicals pipeline leakage positioning method based on CELMD-MCKD. *Nondestruct. Test. Eval.* **2021**, *36*, 477–493. [\[CrossRef\]](#)
22. Pashmforoush, F.; Fotouhi, M.; Ahmadi, M. Acoustic emission-based damage classification of glass/polyester composites using harmony search k-means algorithm. *J. Reinf. Plast. Compos.* **2012**, *31*, 671–680. [\[CrossRef\]](#)
23. Zhang, M.; Zhang, Q.; Li, J.; Xu, J.; Zheng, J. Classification of acoustic emission signals in wood damage and fracture process based on empirical mode decomposition, discrete wavelet transform methods, and selected features. *J. Wood Sci.* **2021**, *67*, 59. [\[CrossRef\]](#)
24. Ding, Z.W.; Li, X.F.; Huang, X.; Wang, M.; Tang, Q.; Jia, J. Feature extraction, recognition, and classification of acoustic emission waveform signal of coal rock sample under uniaxial compression. *Int. J. Rock Mech. Min. Sci.* **2022**, *160*, 105262. [\[CrossRef\]](#)
25. Pham, M.T.; Kim, J.M.; Kim, C.H. Rolling bearing fault diagnosis based on improved GAN and 2-D representation of acoustic emission signals. *IEEE Access* **2022**, *10*, 78056–78069. [\[CrossRef\]](#)
26. Han, G.; Kim, Y.-M.; Kim, H.; Oh, T.-M.; Song, K.-I.; Kim, A.; Kim, Y.; Cho, Y.; Kwon, T.-H. Auto-detection of acoustic emission signals from cracking of concrete structures using convolutional neural networks: Upscaling from specimen. *Expert Syst. Appl.* **2021**, *186*, 115863. [\[CrossRef\]](#)
27. Wu, Y.; Meng, Y.; Shao, C. End-to-end online quality prediction for ultrasonic metal welding using sensor fusion and deep learning. *J. Manuf. Process.* **2022**, *83*, 685–694. [\[CrossRef\]](#)
28. Xu, Y.; Wang, X.; Zhang, H. SE-Swin: An improved Swin-Transformer network of self-ensemble feature extraction framework for image retrieval. *IET Image Process.* **2024**, *18*, 13–21. [\[CrossRef\]](#)
29. Zheng, F.; Lin, S.; Zhou, W.; Huang, H. A lightweight dual-branch Swin transformer for remote sensing scene classification. *Remote Sens.* **2023**, *15*, 2865. [\[CrossRef\]](#)
30. Sun, R.; Pang, Y.; Li, W. Efficient lung cancer image classification and segmentation algorithm based on an improved Swin transformer. *Electronics* **2023**, *12*, 1024. [\[CrossRef\]](#)
31. Chen, X.; Xiang, Y.; Feng, Y.T. Selection of interpolation methods used to mitigate spectral misregistration of imaging spectrometers. *Spectrosc. Spectr. Anal.* **2011**, *31*, 1147–1150.
32. Zhou, T.; Han, D.; Lai, M.J. Energy minimization method for scattered data Hermite interpolation. *Appl. Numer. Math.* **2008**, *58*, 646–659. [\[CrossRef\]](#)
33. Jia, L.; Zhang, Q.; Zheng, X.; Yao, P.; He, X.; Wei, X. The empirical optimal envelope and its application to local mean decomposition. *Digit. Signal Prog.* **2019**, *87*, 166–177. [\[CrossRef\]](#)

34. Hu, Y.; Jiang, G. Weft-knitted fabric defect classification based on a Swin transformer deformable convolutional network. *Text. Res. J.* **2023**, *93*, 2409–2420. [[CrossRef](#)]
35. Xiao, X.; Guo, W.; Chen, R.; Hui, Y.; Wang, J.; Zhao, H. A Swin Transformer-based encoding booster integrated in U-shaped network for building extraction. *Remote Sens.* **2022**, *14*, 2611. [[CrossRef](#)]
36. Yang, H.-T.; Bai, B.; Lin, H. Seismic magnitude calculation based on rate- and state-dependent friction law. *J. Cent. South Univ.* **2023**, *30*, 2671–2685. [[CrossRef](#)]
37. Li, S.; Lin, H.; Lin, Q.-B.; Wang, Y.-X.; Zhao, Y.-L.; Hu, H.-H. Mechanical behavior and failure characteristics of double-layer composite rock-like specimens with two coplanar joints under uniaxial loading. *Trans. Nonferrous Met. Soc. China* **2023**, *33*, 2815–2831. [[CrossRef](#)]

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