IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 October 2023, accepted 21 November 2023, date of publication 1 December 2023, date of current version 11 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3338628

RESEARCH ARTICLE

Wavelet Transform-Based Fuzzy Clustering Microseismic First-Arrival Picking Method

TINGTING LIN, JINGWANG CHENG, QIANG CHEN, AND SIYU CUI
College of Mining Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Corresponding author: Jingwang Cheng (Chengjingwang@tyut.edu.cn)

This work was supported by the Fundamental Research Program of Shanxi Province under Grant 202303021211058.

ABSTRACT Microseismic arrival time picking serves as the foundation for microseismic source localization and holds significant importance in the field of microseismic monitoring. Traditional methods, such as Short-Time Average/Long-Time Average (STA/LTA) and clustering methods based on STA/LTA as feature vectors, require manual adjustments of the time window parameters to achieve accurate picking. Furthermore, they are susceptible to inaccuracies in high-background noise environments. And in response to these challenges, this study introduces a fuzzy clustering algorithm based on Continuous Wavelet Transform (CWT-FCM) for microseismic arrival time picking. This method begins by transforming raw data into the wavelet domain and selecting scales with relatively large standard deviations as input for the fuzzy clustering process. Ultimately, it identifies the initial arrivals of microseismic events within the resulting clusters. In this study, our proposed method is applied to microseismic datasets with low signal-to-noise ratios as well as real data, successfully and accurately picking microseismic arrivals. Compared with traditional methods, our approach demonstrates increased robustness and practical value in high-interference scenarios. Notably, it eliminates the need for manual parameter adjustments, thereby enhancing efficiency and precision in automated microseismic signal picking and establishing a foundational dataset for subsequent automatic and high-precision microseismic arrival time localization.

INDEX TERMS Microseismic monitoring, STA/LTA, wavelet transform, fuzzy clustering.

I. INTRODUCTION

Hydraulic fracturing technology [1], [2], [3] plays a significant role in the exploitation and enhanced production of low-permeability oil and shale fields. During the construction process, microseismic monitoring technology was employed to capture the first arrival times of seismic waves and invert them to obtain information about rock fractures. This aids in understanding the spatial distribution and morphology of fractures induced by rock fracturing, ultimately aiming for efficient exploitation. With increasing demand for unconventional oil and gas, microseismic monitoring technology is expected to find more applications. However, in practical exploitation, the presence of high background noise poses challenges for arrival-time picking. Therefore, research on microseismic arrival picking is necessary to improve the accuracy of microseismic monitoring.

Traditional picking methods often utilize Short-Term Average/Long-Term Average (STA/LTA) [4] and Akaike Information Criterion (AIC) [5]. Although STA/LTA and AIC have simple models and short processing times, they are difficult to apply individually. On one hand, they are sensitive to noise and struggle to achieve high-precision picking under low signal-to-noise ratio conditions. However, their parameters must be set manually, resulting in poor adaptability. Scholars have made improvements through two main approaches. First, they have addressed the inherent problems of these methods by incorporating feature functions. For example, the AIC formula has been combined with autoregressive techniques, variance, and kurtosis algorithms, resulting in improved algorithms such as VAR-AIC [6], [7] AR-AIC [8], and Kur-AIC [9]. Lei and Caihua [10] and others introduced a reference threshold in the calculation

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

process of the STA/LTA method to reduce the difficulty of parameter selection. Second, to mitigate the sensitivity of these methods to noise, researchers have applied preprocessing techniques to microseismic data by integrating other algorithms to improve the signal-to-noise ratio. For instance, Quanjie et al. [7] and colleagues employed a wavelet packet denoising model, Ruisheng et al. [6] utilized Hilbert transform to calculate envelope signals and determine the approximate interval of phase arrival, and Eduardo [11] applied the FCM algorithm to select complete microseismic events. Other denoising methods for microseismic signals include wavelet transform [12], empirical mode decomposition [13], filtering [14], masking [15], and top-hat transform [16]. Although the improvements made to the STA/LTA and AIC methods enhanced the robustness of the algorithms against noise and the accuracy of arrival picking, their adaptability remained limited.

With the rapid rise of computer networks, machine learning has become pervasive across various industries, including arrival time picking. Machine learning encompasses both supervised and unsupervised learning. Supervised learning involves labeling the training set data for subsequent data differentiation. Chen et al. [17] used convolutional neural networks (CNN) to classify the first waveform and employed the k-means algorithm for time picking. Owing the problem of vanishing gradients in deep neural networks, J. Zheng et al. [18] trained long short-term memory neural networks and input the original waveform signals into the trained network model to obtain the feature vectors of P-wave arrivals. The aforementioned methods are based on neural networks established by scalar neurons, which are not as flexible as capsule neural networks established by vector neurons. The former method requires a large amount of data for generalization. Chen et al. [19] used capsule neural networks to determine the arrival time of elastic waves.

Because of lack of a reliable event catalog for microseismic data and the complexity of adding labels to a vast number of microseismic events, unsupervised learning is more widely applied in the field of microseismic analysis than supervised learning. Clustering analysis is an unsupervised learning method used to group similar data points into clusters with similar features. Its objective is to discover hidden structures and patterns in the data, partitioning data points into different clusters where data points within the same cluster exhibit high similarity, whereas those from different clusters have low similarity. Current researchers have made improvements to clustering algorithms to achieve accurate picking. For example, Meng [20] classified microseismic data using spectral clustering based on the differences between noise and low-dimensional manifold features of the signal. Ma [21] improved each step of the clustering algorithm, using a locally linear embedding algorithm for Euclidean distance calculation and enhanced particle optimization clustering for cluster centers, further enhancing the performance of clustering analysis. However, the spectral clustering algorithm is sensitive to noise, and Ma's method has long iteration times and high computational complexity. Currently, most researchers achieve accurate classification by modifying the input values of clustering algorithms. For instance, Chen [22] used the mean, power, and STA/LTA as input vectors for c-means and used a threshold to determine the P-wave arrival time. Because STA/LTA is sensitive to noise and its inclusion in the calculations can affect the picking results under low signal-to-noise ratio conditions, Chen [23] further improved the feature vectors by replacing the STA/LTA parameters with spectral centroids. Although this algorithm achieves fast computation without requiring input parameters, spectral centroids are sensitive to noise and require the use of synchronized compressed wavelet technology for denoising, which increases technical difficulty. While the aforementioned methods can accurately measure microseismic arrival times to some extent, they also have limitations in adaptability, as the setting of input parameters relies heavily on manual adjustment. Improper parameter settings can easily lead to inaccurate results in clustering analysis. Therefore, it is necessary to study a stable feature vector.

Picking operations performed directly in environments characterized by low signal-to-noise ratios present significant challenges. Recently, wavelet transform-based fuzzy clustering methods have been widely applied in the medical domain [24], [25], [26], [27]. By leveraging the unique attributes of microseismic low frequencies, this study also incorporates the wavelet transform as the feature vector for fuzzy clustering. The wavelet transform confers dual advantages. It not only alleviates the impact of noise interference but also directly computes the raw data. Furthermore, the wavelet transform excels in detecting sudden changes, making it particularly suited for capturing the initial arrival times. Building on these principles, this paper introduces an optimized method with the following sequential steps: the raw data undergoes wavelet transform, subsequently enabling fuzzy clustering computations employing time-frequency data to ascertain microseismic initial arrival times. A noteworthy aspect of this method is its independence from the parameter inputs. Application to both simulated and authentic data revealed the robustness of the proposed approach, sustaining a remarkable level of precision even under robust interference conditions.

II. METHOD

A. FUZZY CLUSTERING

The basic concept of clustering analysis is to group similar data points into the same cluster based on the measure of similarity between the data points. Similarity is typically calculated based on the distance or similarity between the features or attributes of the data points, often using Euclidean distance as the similarity metric. In the field of arrival time picking, fuzzy clustering is mainly applied to differentiate between microseismic signals and noise signals and to identify and determine the arrival time in microseismic signals. In this algorithm, the number of clusters is set to 2.

The objective of fuzzy clustering is to minimize the objective function, as shown in Equation (1), which minimizes both the membership degree and the Euclidean distance.

$$J = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{i,j}^{m} \left\| x_i - c_j \right\|^2, \quad 1 \leq m \leq \infty \tag{1}$$

where $u_{i,j}^{m}$ represents the membership degree of the data point in the *j*th cluster, N denotes that there are N data, x_i denotes the *i*th data point, c_j represents the centroid of the *j*th cluster, $\|\bullet\|$ represents the Euclidean distance, and m is the fuzzy parameter, which is usually greater than 1 [11].

The specific steps of fuzzy clustering are as follows:

First, we select the number of clusters (c) and the fuzzy parameter (m), and randomly initialize the membership degree (initial membership matrix U) of each data point to each cluster center.

Next, based on the current membership matrix u, c calculates the cluster center for each cluster and identifies the points around the cluster center for further partitioning. The cluster center vector for the *j*th cluster is given by the following formula (Equation 2):

$$c_j = \frac{\sum_{i=1}^{N} u_{i,j}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{i,j}^{m}} \tag{2}$$

Then, based on the current cluster centers, the membership degree of each data point was calculated for each cluster. The membership degree of the *i*th data point to the *j*th cluster is given by the following equation (Equation 3):

$$u_{i,j} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{3}$$

Finally, update the membership degree u and centroids c, repeat steps 2 and 3, and check whether the cluster centers and objective function values have changed. If there is no change, the iteration is stopped.

When the membership degree of the useful signals changes, it can be determined whether a microseismic event has occurred. Therefore, the arrival time was determined based on this threshold. In the field of arrival time picking, fuzzy clustering is more sensitive than the k-means clustering algorithm because the membership degree in fuzzy clustering can be precise to decimal places, wheras the values obtained from the k-means algorithm are only 0 and 1.

B. WAVLET TRANSFORM

The most important factor influencing the clustering performance is the input feature vector. In this case, wavelet transform was chosen as the input vector for fuzzy clustering.

The wavelet transform was initially proposed by Gabor, who invented the short-time Fourier transform [24]. Although the short-time Fourier transform has strong processing capabilities for frequency-domain stationary signals, it cannot handle non-stationary signals. Continuous wavelet transform, by moving and scaling the mother wavelet, performs local multiplications on the signal to capture abrupt changes and effectively handle non-stationary signals. The formula for the continuous wavelet transform is as follows:

$$WT_x\ (a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} \delta(t) \overline{\psi((t-b)/a)} dt \tag{4}$$

where a represents the scale factor, b represents the translation factor, $\delta(t)$ represents the original signal, and $\psi(t)$ represents the mother wavelet. In this case, the Morlet wavelet was used as the wavelet basis for the time-frequency analysis. The Morlet wavelet expression is as follows:

$$\psi(t) = exp\,(iw_0 t)\, exp\left(-\frac{t^2}{2}\right) \tag{5}$$

The wavelet transform can effectively suppress noise interference. However, the continuous wavelet transform outputs a large amount of scale data to achieve a time-frequency analysis. Using all the time-frequency data as the input feature vector significantly increases the computational complexity and reduces efficiency. Therefore, in this method, the standard deviation algorithm is introduced after continuous wavelet transform to prevent the influence of redundant features caused by the wavelet transform on the clustering results.

The standard deviation reflects the data dispersion or spread. As illustrated in Figure 1a, it represents a single-channel, noise-free microseismic signal. Figure 2a introduces a certain amount of white noise. Figures 1b and 2b depict the corresponding spectral plots in Figures 1a and 2a, respectively. Figure 1b portrays the three-dimensional data, including wavelet coefficients, time, and frequency. To sharpen the focus and enhance clarity, we implemented a screening process based on the standard deviation of the time-frequency data. This screening enabled us to extract

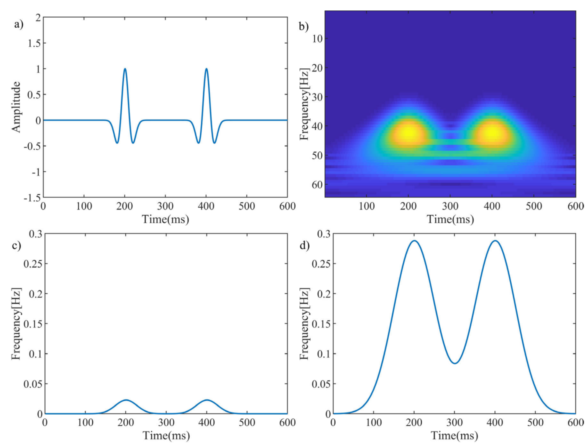

FIGURE 1. (a) Clean single-channel data (b) Spectrogram for a (c) Time-frequency data for std=0.057 (d) Time-frequency data for std=0.103.

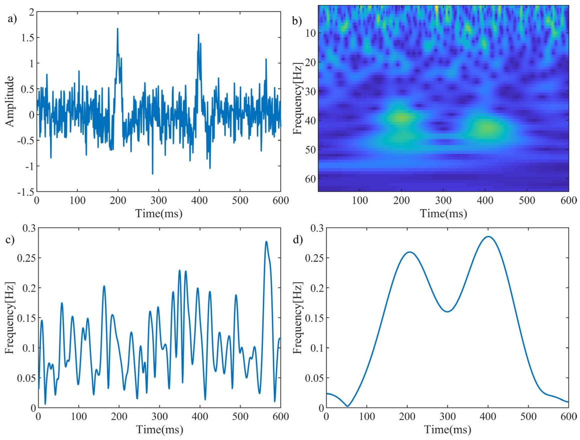

FIGURE 2. **(a) Noised single-channel data (b) Spectrogram for a (c) Time-frequency data for std=0.053 (d) Time-frequency data for std=0.096.**

time-domain data selectively, resulting in the effects depicted in Figures 1c and 1d.

Based on the information presented in the figures, we can draw the following conclusions:1) after wavelet transform, the number of samples significantly increases; 2) with the increase in noise, useful information is covered up; 3) under low signal-to-noise ratio, time-frequency data with larger standard deviation exhibits better noise resistance, while time-frequency data with smaller standard deviation has poor denoising capabilities. In other words, data with larger standard deviations are more stable, which is beneficial for fuzzy clustering.

After conducting extensive testing, we opted to use the eight values with the maximum standard deviations as inputs for the fuzzy clustering process. The formula for the standard deviation is:

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|A_i - \mu|} \tag{6}$$

where μ represents the average value in data A, and the formula is $\mu = \frac{1}{N}\sum_{i=1}^{N} A_i$

III. RESULTS

A. NON-WAVEFORM VECTOR INTERFERENCE

First, to verify the robustness of the proposed model against non-waveform disturbances, we introduce noise into the theoretical data model. To accurately represent the intensity of the noise, we utilize the signal-to-noise ratio (SNR), which is calculated using the following formula:

$$SNR = 20\log\frac{\|s\|}{\|s-n\|} \tag{7}$$

Within the simulated datasets, the generated signal represents an ideal data signal, devoid of any noise, and can be directly employed as the clean signal s for computations. In contrast, n denotes the synthetic data acquired by introducing a specified level of random noise to the signal s. So the formula is applicable to this simulated data set.

The simulated data model without noise is shown in Figure 3, representing a single-source signal with a frequency of 20 Hz. To evaluate the accuracy of the proposed method, we introduced the method proposed in [22] (referred to as the Chen method) and the traditional STA/LTA method as cross-validation techniques.

Figures 3a and 3d correspond to the theoretical picking method and the method proposed in this study, respectively, while Figures 3b and 3e correspond to the Chen method, and Figures 3c and 3f correspond to the STA/LTA method. As both the STA/LTA and Chen method require parameter settings, we used parameter values of 5 for the short-time window length (nsta), 15 for the long-time window length (nlta), and 25 for the analysis window length (q), as shown in Figures 3b - 3c. These specific parameter values were determined by the authors of the Chen method in their original study, where they demonstrated the effectiveness of these settings in microseismic arrival picking. Through extensive experiments, we conducted a comprehensive analysis of the different parameter combinations for the proposed method. After careful evaluation, we found that at a signal-to-noise ratio (SNR) of −2.09 dB, setting nsta, nlta, and q to 87, 146, and 51, respectively, produced the most favorable initial picking results to a large extent. Therefore, we recommend using these parameter values in our study, as shown in Figures 3e to 3f. Figure 3 reveals that in the absence of noise, all methods produce relatively accurate picking results, with different parameters having minimal impact on the results.

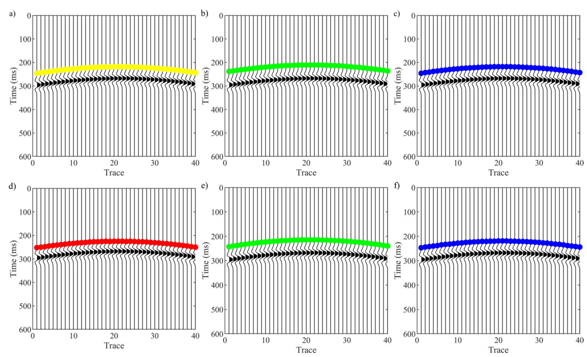

FIGURE 3. **(a) Theoretical pickup data, (b) Chen method, (c) STA/LTA method, (d) method in this paper, Parameter increase of (e) Chen method and (f) STA/LTA method arrival-pickup results for noiseless simulated data.**

Subsequently, using the theoretical initial arrival time as a reference, we introduced Gaussian noise into the simulated data model shown in Figure 3, reducing the signal-to-noise ratio to −2.09 dB and −8.11 dB, respectively. Figures 4 and 5 correspond to SNRs of −2.09 dB and −8.11 dB, respectively. In Figure 4, it is evident that the STA/LTA with smaller parameters is sensitive to noise, leading to significant disturbances during picking. Increasing the parameters results in smoother STA/LTA picking; however, the results shift towards microseismic events, deviating from the realm of the initial arrival times. Similarly, the Chen method using STA/LTA as feature vectors exhibits disturbances with smaller parameters. Several picking points had values close to 0, indicating that q was too short. Increasing the number of parameters significantly reduced the number of erroneous points, as shown in Figure 4e. In contrast, our method demonstrates a relatively consistent picking performance even as the signal-to-noise ratio decreases.

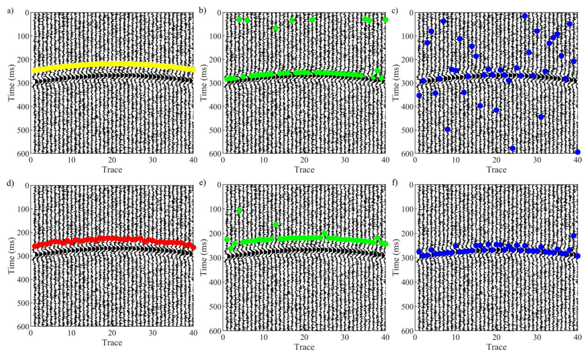

FIGURE 4. (a) Theoretical pickup data, (b) Chen method, (c) STA/LTA method, (d) method in this paper, parameter increase of (e) Chen method and (f) STA/LTA method arrival pickup results at a noise level of SNR = −2.09 dB.

Continuing to lower the signal-to-noise ratio to −8.11 dB, the picking results are as shown. The STA/LTA method with smaller parameters exhibited more pronounced disturbances, whereas the STA/LTA method with adjusted parameters was relatively stable, with picking points falling on the microseismic events. When microseismic events are weak, the Chen method in Figures 5d and 5e can no longer pick the initial arrivals, and the picking points approach 0, indicating that the parameters require further adjustments. Under low signal-to-noise ratio conditions, our method shows a slight fluctuation in picking as the SNR decreases.

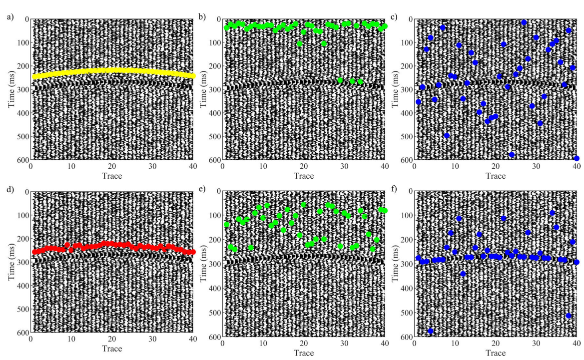

FIGURE 5. (a) Theoretical pickup data, (b) Chen method, (c) STA/LTA method, (d) method in this paper, parameter increase of (e) Chen method and (f) STA/LTA method arrival pickup results at a noise level of SNR = −8.11 dB.

To evaluate the picking results, we calculate the picking error using the formula:

$$\mathrm{E} = \frac{1}{N}\sum_{i}^{N}\left|P(i) - \hat{P}(i)\right| \tag{8}$$

where E represents the picking error measured in the sample, $P(i)$ represents the picked first arrival time of the ith trace, $\hat{P}(i)$ represents the accurate first arrival time of the microseismic event, and N is the total number of traces.

The results are shown in Figure 6. Our method exhibits a slight increase in the error as the noise level increases, which is acceptable. Overall, the picking performance was good. STA/LTA performs well under high signal-to-noise ratio conditions, but as noise increases, the accuracy significantly decreases. Under low SNR conditions, altering the parameters allows the method to identify only the maximum amplitude of the microseismic event, indicating inherent limitations. In contrast, at an SNR of −3.84 dB, manual adjustments were made to the parameters of the Chen method. Therefore, the Chen method exhibits greater noise tolerance than STA/LTA at this SNR. However, at SNRs exceeding −3.84 dB, the Chen method experienced widespread disturbances. Improving the parameters significantly enhances the accuracy of the Chen method, howerver, the parameter data require continuous updates.

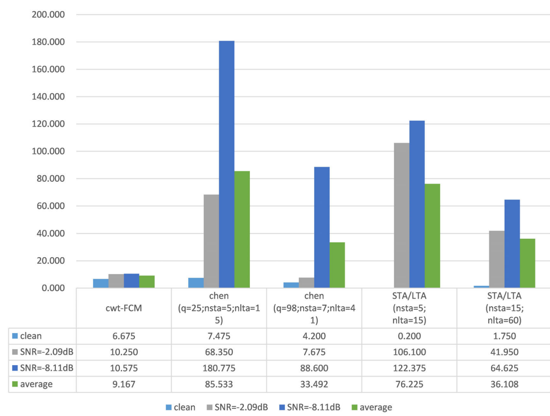

FIGURE 6. Error pick-up diagram.

In conclusion, STA/LTA is highly sensitive to noise, and changes in the parameters have some influence on the picking results, although not significantly. The Chen method is also sensitive to noise, and changing parameters can greatly improve the picking results, however the parameters require continuous updating. In contrast, the method introduced in this study does not require parameter input and exhibits a certain level of robustness to noise, achieving accurate picking even under low signal-to-noise ratio conditions.

B. WAVEFORM VECTOR INTERFERENCE

In addition to non-waveform components, waveform components can also generate energy interference. To verify the robustness of the proposed method under the interference of waveform components, tests were conducted using a theoretical data model.

First, waveform vector interference was added to assess the applicability of the proposed method. Figure 7 shows the simulated microseismic data, consisting of 60 receivers and 2 microseismic sources. The first microseismic event had a dominant frequency of 20 Hz, whereas the second had a dominant frequency of 40 Hz. The picking results are shown in Figure 7. Even in cases with relatively high signal-to-noise ratios, the STA/LTA picking points fell on the second microseismic event. The Chen and proposed methods achieved more accurate picking. During the picking process, we observed that the STA/LTA method tends to select data with higher frequencies, whereas the Chen method, which includes other feature vectors, was not affected by changes in the frequency of the second microseismic event.

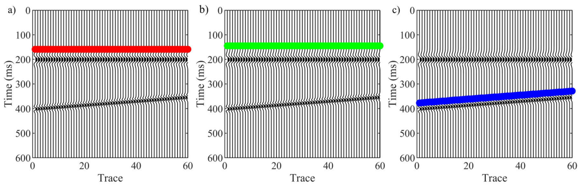

FIGURE 7. (a) Method in this paper (b) Chen method and (c) STA/LTA method pickup results under the interference of high frequency waveform components.

Second, when waves are excited in a well, they undergo reflection and refraction between different media. When the waves reach the bottom of the well or other interfaces, some waves will reflect upward, forming upgoing waves, while others will continue to propagate downward, forming downgoing waves. Considering that upgoing and downgoing waves, as waveform components, can also interfere with the picking of first arrivals, we simulated a situation in which waves were excited in the well using a three-layer velocity model. Figure 8 shows the configuration of 50 receivers represented by vertically arranged black squares and seismic sources represented by red triangles in the third layer. The microseismic event was excited by a 55 Hz Ricker wavelet. The microseismic dataset synthesized from this velocity model is shown in Figure 9. The first appearance of the two downlink waves was located in channels 15 and 28.

The picking results of the three methods are shown in Figure 9. Among these methods, the proposed approach, STA/LTA, and the Chen method show commendable performance in picking.

To further comprehensively compare the impact of waveform components and non-waveform components on the picking results, we added noise to the microseismic data set shown in Figure 9, reducing the SNR to −8.27 dB. Considering that the parameters of the Chen and STA/LTA methods can affect the picking results in the presence of noise, we manually optimized the parameters and conducted subsequent tests, resulting in the final results shown

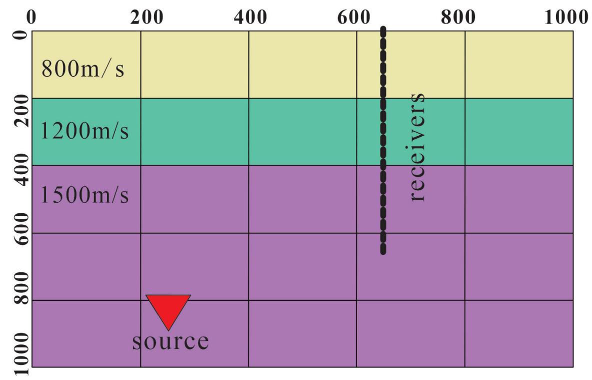

FIGURE 8. Acoustic velocity model.

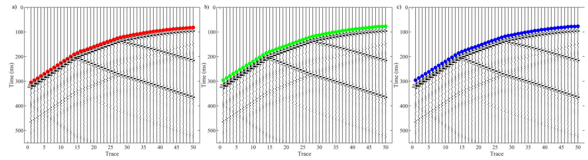

FIGURE 9. (a) Method in this paper (b) Chen method and (c) STA/LTA method pickup results under downlink waveform component interference.

in Figure 10. It can be observed that the proposed method remains robust, the STA/LTA method has half of the picking points falling on the second microseismic event, and the Chen method exhibits significant disorder after the appearance of waveform components, rendering the analysis meaningless.

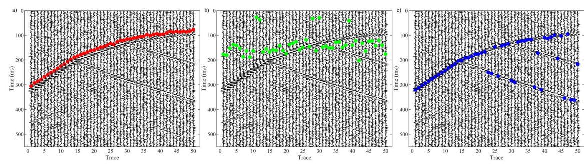

FIGURE 10. (a) Method in this paper (b) Chen method and (c) STA/LTA method pickup results in the presence of both waveform interference and non-waveform interference.

In conclusion, waveform components can interfere with the feature vectors. The STA/LTA method is prone to false picking when the interfering waveform component has a higher frequency. The Chen method is susceptible to interference when the interfering waveform component is adjacent to the microseismic waveform component or when gap waves appear. In both cases, the CWT-FCM method achieves ac-curate picking, demonstrates good generalization, and can accurately perform operations even in the presence of interference from both waveform and non-waveform components.

C. REAL DATA

Finally, we applied the methodology outlined in this study to actual data. Figure 11(a) illustrates the 16-level tricomponent microseismic record obtained during a hydraulic fracturing operation. Detailed information on the detection process is

TABLE 1. Basic well and instrument information.

Monitoring wells	xxx
fracture well	xxx
borehole number	xxx
Acquisition Instrument Model	GEOWAVES
Instrument Sampling Rate	0.5Ms
Record length	3s
Instrument Preamp Gain	40DB
numerical range	10m

presented in Table 1. Notably, there was a significant fluctuation around 2000 ms, which is indicative of a microseismic event occurring at that time. When we zoomed in on the microseismic event around 2000 ms, as demonstrated in Figure 11(b), distinct P and S waves became discernible.

We applied the method proposed in this paper along with STA/LTA to detect microseismic events within the dataset, as shown in Figure 11. The results were magnified

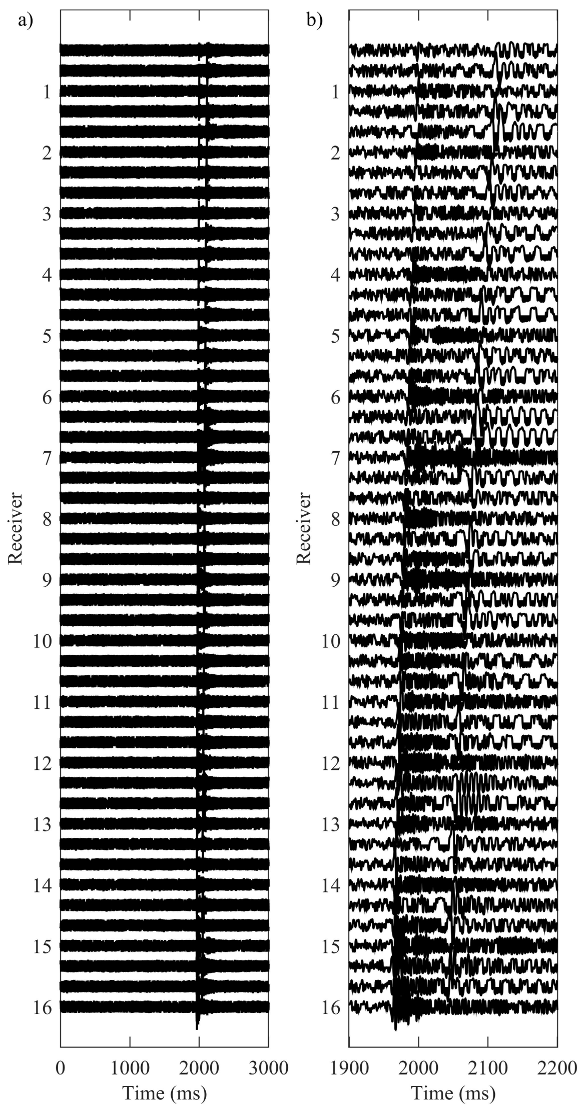

FIGURE 11. (a) Method in this paper and (b) STA/LTA method for the actual data pickup results.

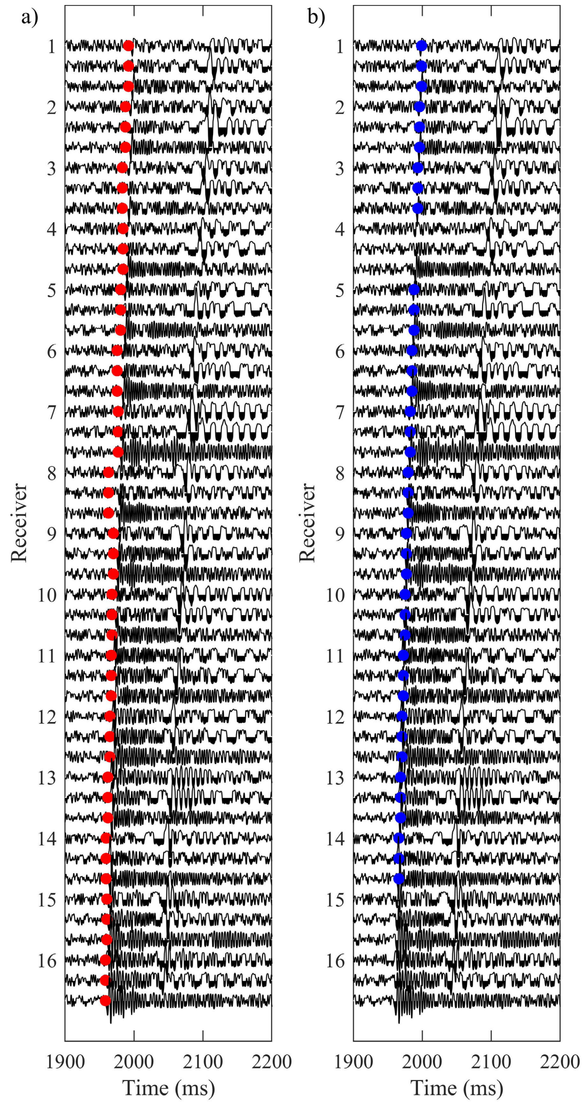

FIGURE 12. (a) Method in this paper and (b) STA/LTA method for the actual data pickup results enlarged.

in Figure 12. It's important to mention that, to best illustrate the performance of the methods employed, we opted not to apply denoising to the STA/LTA method. However, we fine-tuned the parameters for the STA/LTA method. Figure 12 shows that, under conditions of strong interference, the proposed method outperformed STA/LTA. In contrast, the STA method exhibited three segments that did not appear in the 1900 ms to 2200 ms range, indicating a lack of precision in event detection.

To gain deeper insight into these errors, we extracted data from the 15th level using both our method and STA/LTA, as depicted in the figure. These data corresponded to the XYZ axes at the 15th level. It is evident that the microseismic data experienced substantial energy interference. Given our focus on arrival times, we elected to analyze the Z-axis data.

Figure 14 shows the results of the STA/LTA calculations for the microseismic data. It becomes apparent that under

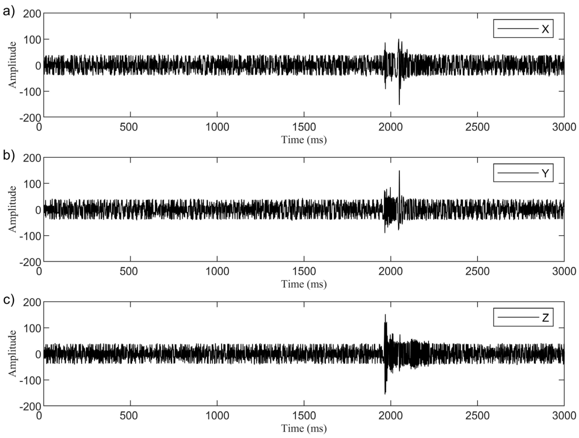

FIGURE 13. 15th channel of microseismic data (a) x-axis data (b) y-axis data (c) z-axis data.

low signal-to-noise ratios, the amplitude of the noise equals or surpasses that of the events. This results in numerous false peaks in the STA/LTA method, rendering it incapable of accurately pinpointing the locations of microseismic events in high-energy environments. As shown by the green line, the STA/LTA values are highest at the false peak around 1250ms, the STA/LTA determines that the microtremor occurs at this point, but this point is well ahead of the true microtremor arrival time (as shown by the red line). In this particular case, the high energy from the non-waveform components contributed to the STA/LTA misjudgment.

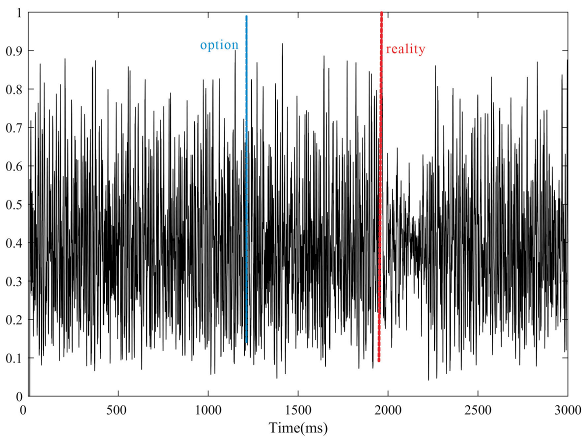

FIGURE 14. The value of STA/LTA.

In Figure 15, we outline the CWT-FCM picking process. Initially, a continuous wavelet transform was applied to the raw data. Figure 15b illustrates that, under low signal-to-noise ratios, strong energy is observed around 2000 ms in the wavelet transform, whereas no discernible energy was apparent around 1200 ms, where STA/LTA erroneously identifies peaks. This highlights the efficacy of the time-frequency analysis in eliminating noise interference. In terms of the noise resistance, the wavelet transform outperformed the STA/LTA algorithm. After filtering the time-frequency data using standard deviation, eight time-frequency data points were obtained, as presented in Figure 15c. It is observed that the standard deviation can effectively distinguish between the microseismic data and noise. As the standard deviation decreased, the amplitude of the microseismic event around 2000 ms also diminished, rendering the demarcation between the microseismic events and noise less distinct. At this juncture, the time-frequency data becomes more stable following interference removal, serving as a feature vector for clustering analysis. This yielded the results shown in Figure 15d. We observed that the signal cluster experienced a jump around 2000 ms, with the post-jump value exceeding 0.5, signifying the occurrence of a microseismic event at that time. This accurately captures the true arrival time of the a microseismic event.

In summary, the presence of noise in real events can lead to false peaks during calculations using the STA/LTA method, resulting in premature data point selection. Conversely, the method introduced in this paper excels in accurately identifying microseismic events even in high-noise environments.

IV. DISCUSSION

In the field of microseismic picking, the STA/LTA method is widely used to determine the first arrival time of events. It compares the short-term average energy (STA) and long-term average energy (LTA) of the signal to distinguish between events and noise. Under high signal-to-noise conditions, seismic events generally exhibit a significant increase in short-term energy, while noise is more prominent in long-term energy. Consequently, the STA/LTA method effectively separates events from noise. However, this method has two limitations: weak resistance to interference and complex parameter settings. These challenges are encountered by several picking algorithms in this field. To address these issues, this paper proposes the use of a fuzzy clustering algorithm based on wavelet transform. Wavelet transform is a widely employed mathematical tool in signal processing that has the unique advantage of capturing signal time-frequency characteristics and facilitating correlation calculations on the original data. Specifically, it enables denoising in the frequency domain and accurate differentiation of waveform components in the time domain, thereby providing precise time information for waveform components and achieving accurate picking results.

A. LIMITATIONS OR SHORTCOMINGS OF THIS STUDY

1) This study utilized a single set of actual data, which limits the comprehensiveness of the dataset;
2) Microseismic picking relies heavily on time-frequency data. However, in scenarios with an extremely low signal-to-noise ratio, applying a continuous wavelet transform to the data results in the signal energy being overshadowed by the noise energy. Consequently, subsequent operations lead to a significant decrease in accuracy;

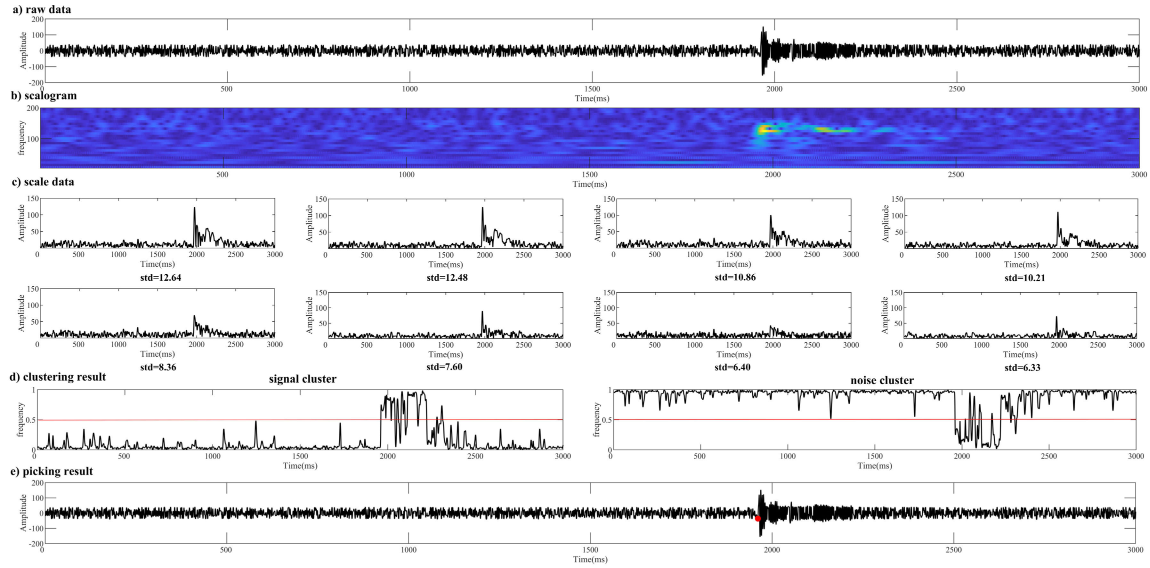

FIGURE 15. **CWT-FCM pickup process (a) raw data (b) scalogram (c) scales of data with the largest standard deviation. (d) clustering results. (e) pickup results.**

3) This method lacks accuracy in selecting microseismic events when they are exceptionally weak. In addition to the exclusion of feature vectors, there are inherent limitations in fuzzy clustering. In reality, microseismic events constitute a relatively small proportion compared to noise data in a given dataset. This biases the centroid calculation in the clustering algorithm towards noise, introducing noise contamination within the signal cluster and significantly reducing the picking accuracy.

B. FUTURE RESEARCH DIRECTIONS

Considering the limitations of this method, we propose a hypothesis to improve the fuzzy clustering. To address the issue of centroids easily skewing towards noise in the presence of strong noise, modifications to the centroid formula can be explored. The standard deviation of the original data or the input feature vectors can be calculated to assess the degree of interference in the original data. By controlling the distance between the two centroids, we can restrict centroid movement based on the strength of the standard deviation.

V. CONCLUSION

This paper proposes a fuzzy clustering-based method for accurate microseismic picking by optimizing time-frequency data. The conclusions drawn from this method are as follows:

1) The raw data undergoes continuous wavelet transform to obtain the time-frequency information of each microseismic trace. Eight data points with the maximum standard deviation are extracted from the low-frequency data as feature vectors for fuzzy c-means clustering. The waveform components and non-waveform components are then processed through clustering. Finally, the first arrival time is extracted from the waveform component data based on a threshold. This method effectively eliminates interference from waveform and non-waveform components and directly computes the original data, thereby efficiently separating the signal from the noise by reducing the redundancy of signal features;
2) STA/LTA, Chen's method, and the method proposed in this paper were applied to simulated and real interference microseismic datasets. The STA/LTA is sensitive to noise, and high-energy waveform interference severely affects its performance. Chen's method relies heavily on manual input parameters, and continuous parameter updates are required to ensure the picking accuracy. In contrast, the method proposed in this study exhibits strong robustness against waveform and non-waveform interference, requires no input parameters, and demonstrates high generalization;
3) When the signal-to-noise ratio is low, the picking accuracy of CWT-FCM changes. There are two reasons for this: useful signals in the scaleogram are masked by irrelevant signals, and the centroids of fuzzy clustering are skewed. These deficiencies point the way for our future work.

ACKNOWLEDGMENT

The authors would like to thank the editors and the reviewers.

REFERENCES

[1] W. Huang, R. Wang, H. Li, and Y. Chen, "Unveiling the signals from extremely noisy microseismic data for high-resolution hydraulic fracturing monitoring," *Sci. Rep.*, vol. 7, no. 1, Sep. 2017, Art. no. 11996.

[2] S. C. Maxwell, J. Rutledge, R. Jones, and M. Fehler, "Petroleum reservoir characterization using downhole microseismic monitoring," *Geophysics*, vol. 75, no. 5, Sep. 2010, Art. no. 75A129.

[3] S. A. Shapiro, C. Dinske, and E. Rothert, "Hydraulic-fracturing controlled dynamics of microseismic clouds," *Geophys. Res. Lett.*, vol. 33, no. 14, Jul. 2006, Art. no. L14312.
[4] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
[5] R. Sleeman and T. van Eck, "Robust automatic P-phase picking: An on-line implementation in the analysis of broadband seismogram recordings," *Phys. Earth Planet. Interiors*, vol. 113, nos. 1–4, pp. 265–275, Jun. 1999.
[6] J. Ruisheng, T. Yunliang, S. Hongmei, and H. Yongfa, "Low signal-to-noise ratio microseismic P-wave first-phase automatic pickup method," *J. Coal Sci. Eng. China*, vol. 40, no. 8, pp. 1845–1852, 2015.
[7] Q. Quanjie, J. Fuxing, W. Quande, W. Bo, L. Jinhai, and L. Xiaohui, "Automatic pick-up method of P-wave first arrivals of microseismic signals for hydraulic fracturing of coal seams," *Chin. J. Rock Mech. Eng.*, vol. 37, no. 10, pp. 2319–2333, 2018.
[8] W. L. Guan, W. Hui, and Z. Junjie, "Automatic P-wave S-wave arrival pickup method for single-component microseismic signals," *Sci. Technol. Rev.*, vol. 34, no. 2, pp. 184–189, 2016.
[9] D. Zhao, X. Liu, H. Li, and Y. Zhou, "Kurtosis and AIC methods in the automatic identification of regional seismic events and direct P-wave primaries," *J. Seismolog. Res.*, vol. 35, no. 2, pp. 220–225, 2012.
[10] Q. Lei and L. Caihua, "STA/LTA method for picking up natural seismic wave first arrivals and its improvement analysis," *Prog Geophys.*, vol. 38, no. 4, pp. 1497–1508, 2023.
[11] E. V. Cano, J. Akram, and D. B. Peter, "Automatic seismic phase picking based on unsupervised machine-learning classification and content information analysis," *Geophysics*, vol. 86, no. 4, pp. V299–V315, Jul. 2021.
[12] S. Guanqun, L. Zhenchun, W. Weibo, and G. Lan, "Wavelet decomposition and higher-order statistics based microseismic first-to-find method," *Oil Geophys. Prospecting*, vol. 54, no. 4, pp. 388–395, 2015.
[13] A. G. Hafez, M. Rabie, and T. Kohda, "Seismic noise study for accurate P-wave arrival detection via MODWT," *Comput. Geosci.*, vol. 54, pp. 148–159, Apr. 2013.
[14] S. Gan, S. Wang, Y. Chen, X. Chen, and K. Xiang, "Separation of simultaneous sources using a structural-oriented median filter in the flattened dimension," *Comput. Geosci.*, vol. 86, pp. 46–54, Jan. 2016.
[15] W. Zhu, S. M. Mousavi, and G. C. Beroza, "Seismic signal denoising and decomposition using deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9476–9488, Nov. 2019.
[16] H. Li, R. Wang, S. Cao, Y. Chen, and W. Huang, "A method for low-frequency noise suppression based on mathematical morphology in microseismic monitoring," *Geophysics*, vol. 81, no. 3, pp. V159–V167, May 2016.
[17] Y. Chen, G. Zhang, M. Bai, S. Zu, Z. Guan, and M. Zhang, "Automatic waveform classification and arrival picking based on convolutional neural network," *Earth Space Sci.*, vol. 6, no. 7, pp. 1244–1261, Jul. 2019.
[18] J. Zheng, J. Lu, S. Peng, and T. Jiang, "An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks," *Geophys. J. Int.*, vol. 212, no. 2, pp. 1389–1397, Feb. 2018.
[19] O. M. Saad and Y. Chen, "Earthquake detection and P-wave arrival time picking using capsule neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6234–6243, Jul. 2021.
[20] Y. Meng, Y. Li, C. Zhang, and H. Zhao, "A time picking method based on spectral multimanifold clustering in microseismic data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1273–1277, Aug. 2017.
[21] H. Ma, T. Wang, Y. Li, and Y. Meng, "A time picking method for microseismic data based on LLE and improved PSO clustering algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1677–1681, Nov. 2018.
[22] Y. Chen, "Automatic microseismic event picking via unsupervised machine learning," *Geophys. J. Int.*, vol. 222, no. 1, pp. 1750–1764, Mar. 2020.
[23] Y. Chen, "Fast waveform detection for microseismic imaging using unsupervised machine learning," *Geophys. J. Int.*, vol. 215, no. 2, pp. 1185–1199, Nov. 2018.
[24] D. Gabor, "Theory of communication. Part 1: The analysis of information," *J. Inst. Electr. Eng. III: Radio Commun. Eng.*, vol. 93, no. 26, pp. 429–441, Nov. 1946.
[25] Y. Özbay, R. Ceylan, and B. Karlik, "Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 1004–1010, Jan. 2011.
[26] H. Castillejos, V. Ponomaryov, L. Nino-de-Rivera, and V. Golikov, "Wavelet transform fuzzy algorithms for dermoscopic image segmentation," *Comput. Math. Methods Med.*, vol. 2012, pp. 1–11, Jan. 2012.
[27] N. Chetih, Z. Messali, A. Serir, and N. Ramou, "Robust fuzzy C-means clustering algorithm using non-parametric Bayesian estimation in wavelet transform domain for noisy MR brain image segmentation," *IET Image Process.*, vol. 12, no. 5, pp. 652–660, May 2018.

TINGTING LIN was born in 1999. She received the B.Sc. degree in communication engineering from Zhejiang Sci-Tech University, China, in 2021. She is currently pursuing the master's degree with the School of Geological Engineering, Taiyuan University of Technology. Her research interest includes machine learning in microseismic monitoring.

JINGWANG CHENG received the B.S. degree in geophysics and the Ph.D. degree in earth exploration and information technology from the China University of Geosciences, Wuhan, China, in 2009 and 2014, respectively. He is currently an Associate Professor with the Taiyuan University of Technology, Taiyuan, China. His research interests include seismic data processing, seismic forward modeling, and full waveform inversion.

QIANG CHEN received the B.S. degree in applied geophysics from the China University of Mining and Technology, Xuzhou, in 1993, the M.S. degree in geodetection and information technology from the Taiyuan University of Technology, Taiyuan, in 2008, and the Ph.D. degree in geodetection and information technology from the China University of Geosciences, Beijing, in 2011. From 1993 to 2005, he was an Engineer with the Shanxi Coal Geological Prospecting Institute of Hydrology. Since 2011, he has been an Assistant Professor with the Mining Engineering Department, Taiyuan University of Technology. He is the author of two books, more than 20 articles, and holds four patents. His research interests include seismic and well logging techniques for coal and coal bearing gas.

SIYU CUI received the B.Sc. degree in information management and information systems from the Shanxi Institute of Energy, China, in 2020. He is currently pursuing the master's degree with the School of Geological Engineering, Taiyuan University of Technology. His research interest includes machine learning.