Intelligent Microseismic Events Recognition in Fiber-Optic Microseismic Monitoring System Compared With Electronic One

Fei Liu, Meng Wang, Min Zhang, XiaoWei Dong, Guo Zhu, and Xian Zhou

***Abstract*—Fiber-optic seismic sensor has become an emerging effective tool for microseismic monitoring, which is of importance for oil and gas production improvement. When processing the fiber-optic data, the recognition of microseismic events is the first crucial step, which is rarely reported due to the lack of such data. In this paper, based on the features of central data distribution, skewness, kurtosis, energy entropy, etc., the classification accuracy of microseismic (MS) events obtained by fiber-optic MS monitoring system and electronic one is compared using machine learning algorithm (e. g. SVM and KNN) for the first time. The results show that fiber-optic data has higher classification accuracy than electronic data when using the feature of central data distribution owing to the high signal-to-noise ratio of fiber-optic data. However, when choosing the features of energy entropy, zero-crossing rate and the energy proportion of specific frequency band, electronic data has higher classification accuracy than fiber-optic data benefiting from the longer events duration and the lower frequency components of electronic data. When using skewness and kurtosis, the classification performance in fiber-optic data is almost consistent with the electronic one. Moreover, the results indicate that the characteristics of MS signal itself have a greater impact on the discrimination ability of MS events than the applied machine learning classification algorithm.**

***Index Terms*—Fiber-optic sensor, machine learning, microseismic events recognition.**

I. Introduction

MICROSEISMIC (MS) monitoring has a pivotal role in oil and gas industry, in which arrays of geophone are deployed in downhole to monitor the seismic wave generated by the fracturing well during the fracturing process. Recent developments in the field of MS monitoring have led to a renewed interest in MS signal processing methods, which helps to acquire information such as position, occurrence time of source and other information [1]–[3]. However, the obtained MS signal is always overwhelmed in the noise caused by sampling error, device non-ideality, environmental coupling, etc. It is still challenging to pick-up the MS events from the noisy background data.

Recently, machine learning method to recognize MS events from noise has been explored [4]–[6]. Bing Ji *et al.* classified high energy tremors from general MS events and optimized parameters using support vector machine (SVM) and the genetic algorithm respectively. The results showed that the classifier achieves 98% sensitivity, 88% accuracy and 87% specificity using both MS raw wave and energy data, which is better than solely utilizing MS raw wave or energy data [7]. Pingan Peng *et al.* proposed an automatic classification method based on a deep learning approach for MS records classifying in underground mines. The event type was correctly determined by the trained convolutional neural network (CNN) classifier with an accuracy of 98.2% [8]. Guangdong Song *et al.* converted raw data into two-dimensional image format by S-Transform and used CNN to recognize the MS signal. The tested accuracy achieved 96.15% with an image size of 180 $\times$ 140 pixels [9]. Yumei Kang *et al.* used the deep belief network model to identify MS events and blasting signals generated in construction processes, which took the advantage of the strong learning and feature extraction abilities of such models and achieved a classification accuracy of 94.4% [10]. Ruisheng Jia *et al.* presented permutation entropy and SVM to detect low signal-to-noise ratio (SNR) MS events more efficiently [11].

Although extensive research has been carried out in MS recognition methods, research on the subject is found to be restricted to limited comparisons of features and classification algorithm. The study focused on the influence of MS signal acquisition method on classification accuracy is barely reported. On the other hand, to date, the MS events classification investigations are limited to the traditional electronic microseismic monitoring system (EMMS) [12]. With the rapid development and application of optical fiber sensing technology, the emerging fiber-optic MS monitoring system (FMMS) [13] has become a promising alternative approach with its effective anti-electromagnetic interference ability [14], high SNR, etc. Consequently, research on the difference of recognition methods to the MS events collected by FMMS and EMMS is of great interest.

Manuscript received December 7, 2021; revised January 26, 2022; accepted January 28, 2022. Date of publication February 4, 2022; date of current version February 15, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61905004, and in part by the Fundamental Research Funds for the Central Universities under Grant FRF-TP-20-067A1Z. *(Fei Liu and Meng Wang contributed equally to this work.) (Corresponding authors: Xian Zhou.)*

Fei Liu, Meng Wang, Guo Zhu, and Xian Zhou are with the School of Computer and Communication Engineering, University of Science & Technology Beijing, Beijing 100083, China (e-mail: liufei12@hotmail.com; 13103146211@163.com; 2902831435@qq.com; zhouxian219@gmail.com).

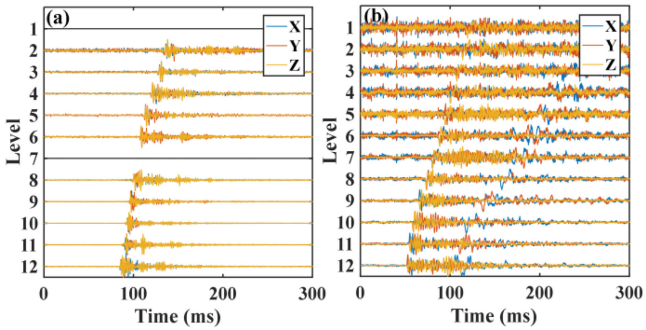
Min Zhang is with the Beijing International Center for Gas Hydrate, Peking University, Beijing 100871, China (e-mail: zhang_min@pku.edu.cn).

XiaoWei Dong is with the Engineering Technology Research Institute of PetroChina, Xinjiang Oilfield Company, Karamay 834000, China (e-mail: dongxw646567@126.com).

Digital Object Identifier 10.1109/JPHOT.2022.3148315

(a) (b)

Fig. 1. Configuration of seismic sensor: (a) FMMS, (b) EMMS.



(a) (b)

Fig. 2. The waveform of typical MS signal: (a) FMMS, (b) EMMS.

The aim of this paper is to explore the difference of MS recognition methods between EMMS and FMMS. First, this paper examines the difference of MS signal produced by two monitoring systems. After that, the MS classification ability using several typical features in the two types of MS monitoring systems is compared via SVM and KNN respectively. It is found that the classification ability between MS events and noise with the same feature is different due to the different MS monitoring systems, namely EMMS or FMMS. Also, it has been indicated that the selected features affect the final classification accuracy more than the classification algorithm used for FMMS and EMMS. It means that the proper feature is of great importance for MS recognition.

II. Signal Analysis Processing

A. The Field Data Description

The MS data are obtained in Xinjiang Oilfield, China, using FMMS and EMMS respectively. The FMMS is based on a self-developed fiber-optic seismic sensor array which contains three-component fiber-optic accelerometers mounted orthogonally to each other respectively, as shown in Fig. 1(a). More detailed configuration can be referred in [13]. The EMMS is based on a commercial geophone array with the geophone model of SGO-15HT, as shown in Fig. 1(b). Both systems have the same sampling frequency of 4 kHz.

The two MS monitoring systems used in this experiment have 12 channels. In each channel the signals detected by the three sensors are denoted by X, Y and Z respectively. In FMMS, the first and seventh channels are reference channels while EMMS has no such configuration. Therefore, there are 30 seismic sensors in FMMS and 36 seismic sensors in EMMS as shown in Fig. 2. Fig. 2 demonstrates the waveform of an identical MS in

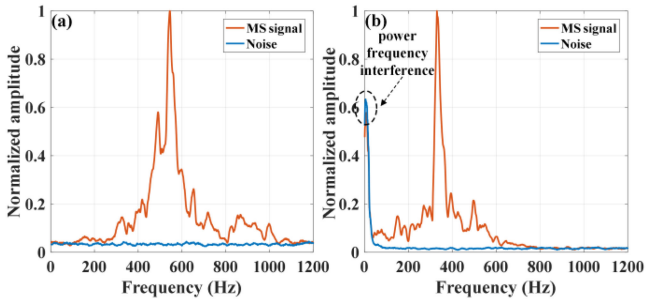


Fig. 3. Spectrum diagram: (a) FMMS, (b) EMMS.

all the channels in both FMMS and EMMS. As shown in Fig. 2, when there is no MS event, the detected signal is dominated by Gaussian noise [15]. When a MS event appears, the seismic sensors in multiple channels can detect MS waveform with the increasing amplitude at first, and then decreasing amplitude with the oscillation in a short time.

B. Comparative of MS Signal From FMMS and EMMS

The difference of MS signals from FMMS and EMMS is the basis of classification comparison. For quantitatively investigation, SNR of MS signal, duration of MS event and frequency component, are given in the following taking the MS events from FMMS and EMMS in Fig. 2 as an example.

1) SNR of MS Events: The MS signal and noise before MS signal arrival are recorded as x_1 and x_2 respectively. The SNR is defined as the ratio of the peak value of MS signal to the root mean square value of the noise as following:

$$SNR = \frac{max\,|x_1|}{\sigma\,(x_2)} \tag{1}$$

The SNR of each channel in Fig. 2 is calculated by Eq. (1), and then the SNR of the MS record file is decided by calculating the average SNR of all channels. The SNR of MS signal from FMMS in Fig. 2(a) is 17.62 dB, while that of from EMMS in Fig. 2(b) is 11.37 dB. Thus, the MS signal monitored by FMMS has higher SNR than that of EMMS.

2) Frequency Component: In order to visually observe the difference between the two types of MS signals in the frequency domain, Fig. 3 shows the spectrum of the data in the Z direction of the 12th channel in Fig. 2. Here, more attention is paid to the distribution of spectral components. As shown in Fig. 3, the MS event from FMMS contains more high-frequency components than the MS event from EMMS. Moreover, there is 50 Hz power frequency interference in spectrum diagram from EMMS. Whereas no such frequency interference occurs in FMMS due to its immunity to electro-magnetic interference.

3) Duration of MS Event: First, the signal is divided into frames and multiplied by window function. The data between two adjacent frames has a repetition rate of 1/3 to ensure the continuity of signal. The window function adopts rectangular window function with a length of q:

$$w\ (m) = \begin{cases} 1, & m = 0 - (q-1) \\ 0, & m = others \end{cases} \tag{2}$$

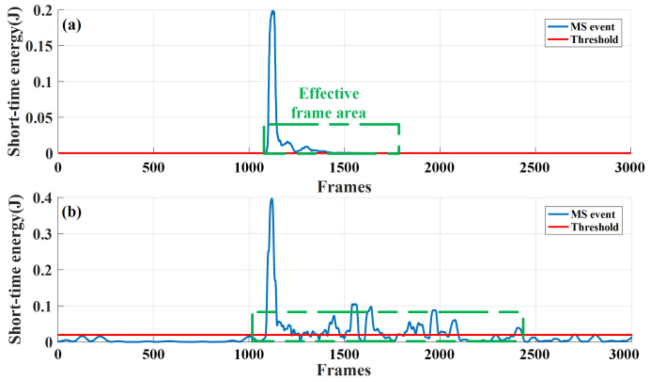


Fig. 4. The duration of MS signal: (a) FMMS, (b) EMMS.

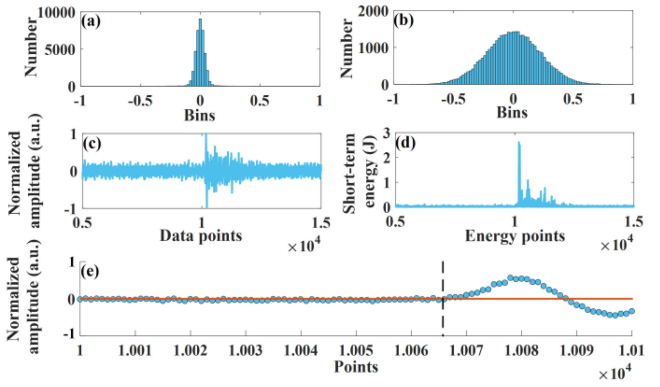


Fig. 5. The comparison of the same feature in MS signal and noise (taking EMMS as an example): (a) Amplitude histogram of MS events, (b) Amplitude histogram of noise, (c) MS event time domain diagram, (d) Short-time energy of MS event in Fig. 5(c), (e) Scatter plot of MS events.

Second, the short-term energy [16] of each frame is obtained through Eq. (3):

$$E_n = \sum_{(n-1)*\frac{2}{3}q}^{(2n+1)*\frac{q}{3}+1} x_m^2 \tag{3}$$

In Eq. (3), E_n is the short-time energy of frame n and x_m is the amplitude of data point in frame. The frames whose short-energy exceed the threshold are denoted as effective frames. The threshold is determined by the maximum value of short-time energy of noise in a period of time before the MS signal arrival. As the time window size is fixed, the number of effective frames is proportional to the duration of MS event. Therefore, the number of effective frames can measure the duration of MS events. In the experiment, we take the value of q as 30. The thresholds are 1.471×10^{-5} and 0.02 respectively in normalization FMMS and normalization EMMS.

According to the above calculation, Fig. 4 presents the short-time energy of the data in the Z direction of the 12th channel shown in Fig. 2. Fig. 4 shows that the number of effective frames in optical fiber data is less than that of electronic data. Through the calculation, the points number of effective frames in optical fiber data is 970, while that of electronic data is 1732. Thus, the MS signal from FMMS has shorter duration time than that from EMMS, which can be interpreted that the higher frequency of MS events contributes to a larger attenuation when travelling over the reservoir [13].

III. Experiments Results and Discussion

A. Feature Extraction

The time domain feature is described by central data distribution, skewness and kurtosis. Central data distribution is obtained by amplitude histogram. Amplitude histogram characterizes the number of normalized amplitude distribution in each bin between [−1, 1] [17]. Fig. 5(a) and 5(b) shows the amplitude histogram of MS signal and noise respectively. The data points whose amplitudes are distributed in the central bin are called central data. The central data distribution is the ratio of the number of central data to the total data points. Skewness describes the symmetry of variable distribution. Kurtosis reflects the steepness and slowness of variable distribution. The sensor collects background noise in the absence of MS signal. Large amount of field data show that the kurtosis and skewness of background noise sample are about 3 and 0 respectively.

The energy domain is characterized by energy entropy, which represents the complexity and randomness of the signal [18]. The entropy obtained from the short-term energy value of the data in each frame is called energy entropy. Energy entropy is proportional to randomness. The stronger the randomness, the greater the entropy. Moreover, the time domain waveform of MS signal and its short-time energy waveform are shown in Fig. 5(c) and 5(d). Fig. 5(d) reveals that the short-time energy after the signal arrival increases much more sharply than the original waveform (Fig. 5(c)), which is beneficial to the identification of MS events.

In the frequency domain, the zero-crossing rate and the energy proportion of specific frequency band are selected as the extracted features [19]. The zero-crossing rate which can be calculated as Eq. (4) simply indicates the signal frequency. It is noted that N is the number of samples in a frame and the sgn stands for symbolic function. The black dotted line separates the data points before and after the MS signal arrivals in Fig. 5(e). As shown in Fig. 5(e), the frequency of noise crossing the zero is significantly higher than that of MS signal crossing zero. That is because the spectrum of MS signal is mainly concentrated in the low frequency band compared with noise, while there is still a small amount of noise in the high frequency band. Thus, there are some data points of noise with weak energy but high oscillation frequency in time domain as shown in Fig. 5(e).

$$D = \frac{1}{N}\sum_{m=1}^{N} \left|sgn\left[x\left(m\right)\right] - sgn\left[x\left(m-1\right)\right]\right| \tag{4}$$

$$sgn\ [x] = \begin{cases} 1, x \geq 0 \\ -1, x < 0 \end{cases} \tag{5}$$

The energy proportion of specific frequency band describes the distribution characteristic of the spectrum. Define frequency band occupied by effective signal as specific frequency band, such as 420–620Hz in Fig. 3(a) and 300–400Hz in Fig. 3(b).

TABLE I
RECOGNITION ACCURACY OF MS EVENTS BASED ON SVM

Category	Feature	FMMS	EMMS
Time domain	Central data distribution	93.51%	88.08%
	Skewness	83.07%	81.22%
	Kurtosis	93.94%	92.14%
Energy domain	Energy entropy	92.65%	94.04%
Frequency domain	Zero-crossing rate	89.29%	90.33%
	Energy proportion of specific frequency band	85.07%	85.19%

The energy proportion of specific frequency band is the ratio of signal energy of a specific frequency band to the total signal energy. The data points of noise are randomly distributed in the whole frequency domain, so the energy proportion of specific spectrum energy is less than that of MS signal.

Because the categories of samples have been marked according the information obtained from the experts in advance, the classification task belongs to supervised learning process. Moreover, we focus more on the classifier where features are handcrafted instead of feature extraction based on deep learning. Meanwhile, the target classifiers should have good performance when the training data set is small due to the small number of MS events. Here, SVM and KNN, two commonly used supervised methods which are good at solving the small sample classification, are selected to classify the MS events in this study. Compared to other algorithms which need large amounts of data for automatic feature extraction, the classification algorithms used here perform better in the case of limited data set due to the collection difficulties in the field experiments of MS monitoring.

B. MS Event Recognition Based on SVM

As a binary classifier, SVM is efficient to solve small sample problems and usually applied in MS events recognition. Here SVM is implemented by the LIBSVM software package [20]. SVM is used to pick up MS events in the data obtained by FMMS and EMMS. The data set of each monitoring system contains 480 MS event samples and 480 noise samples. The ratio of the training set to test set is 6:4. 10-fold cross-validation is used to verify the prediction results and the grid search method is used to search for the optimal parameters [21].

Table I shows the MS events recognition results of two sets of data based on SVM. To compare the differences between the two kinds of MS event recognition methods further, the classification results are analyzed in the perspective of the difference between the two types of MS signals. The difference of SNR directly leads to the difference of MS events classification effect of two kinds monitoring systems, especially when using the time domain features. Generally, the accuracy of optical fiber data is higher than that of electronic data. For central data distribution affected by amplitude, the accuracy of optical fiber data, which is 93.51%, is higher than that of electronic data with the accuracy of 88.08%. While for the skewness and kurtosis, the accuracy of optical fiber data is a little higher than that of electronic data. As shown in table, the accuracy of the skewness and kurtosis in optical fiber data are 83.07% and 93.94%, respectively, while those of electronic data are 81.22% and 92.14%.

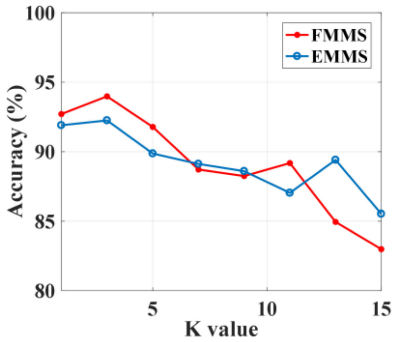


Fig. 6. Relationship between K value and accuracy.

The duration of MS event mainly affects energy entropy and energy proportion of specific frequency band. In the time domain, the longer the effective signal duration, the lower the overall randomness of the sample. Compared with optical fiber data, the longer effective signal duration of electronic data reduces its randomness meanwhile increases the difference with noise. Thus, energy entropy can fully describe the specific characteristics of electronic data and make the better classification accuracy (94.04%) than that of optical fiber data. For energy proportion of specific frequency band, the longer duration contributes to a greater energy of the signal in a specific frequency band, which means a greater discrimination between MS signal and noise. From this point of view, energy proportion of specific frequency band is also applicable to electronic data. The experiment results show that the accuracy of electronic data (85.19%) is slightly higher than that of optical fiber data (85.07%).

The difference of frequency domain affects the zero-crossing rate. Optical fiber data contains higher frequency components than electronic data, which means higher zero-crossing rate. For electronic data, lower frequency components lead to lower zero-crossing rate. The lower zero-crossing rate increases the difference between MS signal and noise, which is conducive to improve the accuracy of MS signal recognition. The result shows that the accuracy of zero-crossing rate in electronic data is 90.33%, while that of in optical fiber data is 89.29%.

C. MS Event Recognition Based on KNN

Considering the influence of classification algorithm, the same experiment was carried out on these data using KNN again. Fig. 6 shows the relationship between K value and accuracy where kurtosis was taken as an example. Thus, the K value was selected as 3 in subsequent experiments. The MS events recognition results of the two monitoring systems are shown in Table II.

The performance of each feature in the two systems when using KNN is the same as when using SVM, which indicates that the applied classification algorithm has limited influence on the final classification results. To intuitively show the classification effect of the two algorithms on this scenario, the statistical chart of the four cases is shown in Fig. 7. Fig. 7 shows that the classification ability of the same feature using different machine learning algorithms is generally consistent, namely, the

TABLE II
RECOGNITION ACCURACY OF MS EVENTS BASED ON KNN

Category	Feature	FMMS	EMMS
Time domain	Central data distribution	92.45%	90.11%
	Skewness	78.91%	76.99%
	Kurtosis	92.05%	91.87%
Energy domain	Energy entropy	90.20%	93.43%
Frequency domain	Zero-crossing rate	86.27%	89.05%
	Energy proportion of specific frequency band	78.88%	76.30%

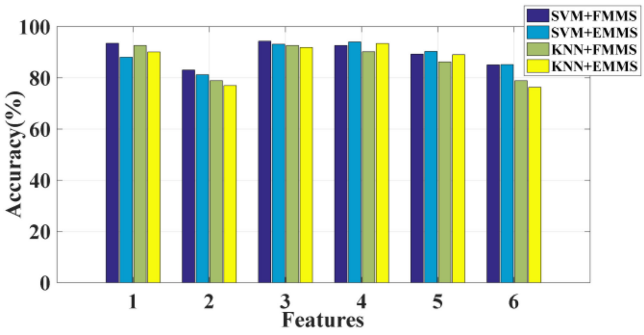


Fig. 7. Classification results (1. Central data distribution, 2. Skewness, 3. Kurtosis, 4. Energy entropy, 5. Zero-crossing rate, 6. Energy proportion of specific frequency band).

classification ability of the features extracted from MS events of the two monitoring systems is less affected by the classification algorithm compared with the characteristics of MS signal itself.

At last we must point out that the work here aims to study the classification ability using the features of the monitored MS data such as SNR, duration time and frequency component. Same conclusion would hold yet as long as the signal was obtained by the same monitoring system in the same or similar field environment. However, if the monitoring system or the field environment was different, leading to the variation of signal features used here, the conclusion of MS event recognition may be different, which needs more investigation in the future.

IV. CONCLUSION

This paper compares the case of MS events recognition between FMMS and EMMS using SVM and KNN, respectively. The experiment results show that different monitoring systems demonstrate different characteristics in time domain and frequency domain about the MS events. When the same feature is applied to the recognition of MS events in the two monitoring systems at the same time, the classification accuracy of the two systems is affected by the MS signal itself more than the classification algorithm. Therefore, appropriate features should be selected according to different systems for MS events recognition. The work set a valuable reference to signal classification for the emerging FMMS.

REFERENCES

[1] D. S. Collins *et al.*, "Seismic event location and source mechanism accounting for complex block geology and voids," in *Proc. Amer. Rock Mechanics Assoc.*, Jun. 2014, pp. 1–7.
[2] P. Peng, L. Wang, and P. Wang, "Targeted location of microseismic events based on a 3D heterogeneous velocity model in underground mining," *PLoS One*, vol. 14, no. 2, 2019, Art. no. e0212881.
[3] J. Sabbione, M. Sacchi, and D. Velis, "Radon transform-based microseismic event detection and signal-to-noise ratio enhancement," *J. Appl. Geophys.*, vol. 113, pp. 51–63, 2015.
[4] Y. Pu, D. Apel, and R. Hall, "Using machine learning approach for microseismic events recognition in underground excavations: Comparison of ten frequently-used models," *Eng. Geol.*, vol. 268, 2020, Art. no. 105519.
[5] M. Malfante *et al.*, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 20–30, Mar. 2018.
[6] M. Curilem *et al.*, "Classification of volcanic seismic events: An expert knowledge analysis for feature selection," in *Proc. 8th Int. Conf. Pattern Recognit. Syst.*, 2017, pp. 1–6.
[7] B. Ji *et al.*, "Investigate contribution of multi-microseismic data to rockburst risk prediction using support vector machine with genetic algorithm," *IEEE Access*, vol. 8, pp. 58817–58828, 2020.
[8] P. Peng *et al.*, "Automatic classification of microseismic records in underground mining: A deep learning approach," *IEEE Access*, vol. 8, pp. 17863–17876, 2020.
[9] G. Song, J. Cheng, and K. Grattan, "Recognition of microseismic and blasting signals in mines based on convolutional neural network and stockwell transform," *IEEE Access*, vol. 8, pp. 45523–45530, 2020.
[10] Y. Kang *et al.*, "Classification of microseismic events and blasts using deep belief network," in *Proc. Chin. Control Decis. Conf.*, 2020, pp. 5556–5561.
[11] R. Jia *et al.*, "Automatic event detection in low SNR microseismic signals based on multi-scale permutation entropy and a support vector machine," *J. Seismol.*, vol. 21, pp. 735–748, 2017.
[12] J. A. Henfling *et al.*, "Development of a HT seismic monitoring tool for downhole," in *Proc. IEEE Conf. Innov. Technol. for an Efficient Reliable Electricity Supply*, Sep. 2010, pp. 373–378.
[13] F. Liu *et al.*, "Downhole microseismic monitoring using time-division multiplexed fiber-optic accelerometer array," *IEEE Access*, vol. 8, pp. 120104–120113, 2020.
[14] M. Azpurua, M. Pous, and F. Silva, "Decomposition of electromagnetic interferences in the time-domain," *IEEE Trans. Electromagn. Compat.*, vol. 58, no. 2, pp. 385–392, Apr. 2016.
[15] R. Bachu *et al.*, "Detection of gaussian-like vibration in fiber optical distributed sensing by 4th-order cumulants," in *Proc. 9th Int. Symp. Next Gener. Electron.*, 2021, pp. 1–4.
[16] M. Jalil, F. Butt, and A. Malik, "Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals," in *Proc. Int. Conf. Technological Adv. Elect., Electron. Comput. Eng.*, 2013, pp. 208–212.
[17] S. Yelmanov and Y. Romanyshyn, "A new approach to image enhancement based on modified histogram equalization," in *Proc. IEEE 14th Int. Conf. Comput. Sci. Inf. Technol.*, Sep. 2019, pp. 5–9.
[18] Y. Yu and C. Junsheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *J. Sound Vib.*, vol. 294, pp. 269–277, 2006.
[19] R. G. Bachu *et al.*, "Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal," in *Proc. Conf. Amer. Soc. Eng. Educ. Zone*, pp. 1–7, 2008.
[20] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," ACM *Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2017. [Online]. Available: https://www.csie.ntu.edu.tw/∼cjlin/libsvm/
[21] T. Xiao *et al.*, "Based on grid-search and PSO parameter optimization for support vector machine," in *Proc. 11th World Congr. Intell. Control Automat.*, Jun. 2014, pp. 1529–1533.