IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 10, 2020, accepted May 31, 2020, date of publication June 5, 2020, date of current version June 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3000333

CEEMD: A New Method to Identify Mine Water Inrush Based on the Signal Processing and Laser-Induced Fluorescence

KAI BIAN[1], MENGRAN ZHOU[1,2], FENG HU[1], WENHAO LAI[1], AND MANMAN HUANG[3]

[1]School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China
[2]State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan 232001, China
[3]School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China

Corresponding author: Mengran Zhou (mrzhou8521@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0604503, in part by the Major Science and Technology Program of Anhui Province under Grant 201903a07020013, and in part by the Anhui Province Natural Science Foundation for Youths under Grant 1808085QE157.

ABSTRACT The rapid and accurate identification of water source types in mine water inrush has been achieved by combining laser-induced fluorescence technology (LIF) with artificial intelligence algorithms. However, these algorithms solely rely on data and image processing analysis to identify different kinds of water samples. To address this issue, we analyzed the fluorescence spectrum and the types of mine water inrush sources from the signal point of view. Firstly, a LIF water inrush spectral analysis system was built to collect spectral data and exhibit fluorescence spectra. Different methods of spectral signal decomposition and reconstruction were compared. The complementary ensemble empirical mode decomposition (CEEMD) algorithm with a better signal evaluation index was selected to preprocess raw spectral signals. Then, the multi-class support vector machine of the cuckoo search optimization (CS-MSVM) model was implemented to the reconstructed spectral signals in different stages. The classification accuracy of the reconstructed signals in the fifth stage was 100%. Compared with raw spectra, other signal processing methods, and other different classifiers, the proposed method has the highest classification accuracy. Finally, the reliability of the algorithm was validated by using the LIF spectral signals of different edible oils and the classification accuracy was 100%. The experimental results show that the CEEMD signal processing method combined with LIF spectroscopy is effective for the accurate identification of mine water inrush source, and it also provides a theoretical basis for the spectral analysis technology that can be used for the identification of other substances.

INDEX TERMS Laser-induced fluorescence, signal processing, cuckoo search, mine water inrush.

I. INTRODUCTION

Because of the complexity of hydrogeological conditions and the increase of mining depth over the years, the mine water inrush has become the second most serious problem in production after coal mine gas [1]. Mine water disaster is a serious threat to safety during production and construction in the mining area. Compared to other major accidents and general accidents, the major flood accidents have the characteristics of difficulty in rescue and in post-disaster production recovery and so on. Once it happens, it will cause more serious casualties and economic losses, and will also bring serious adverse effects to the society. The premise to solve the problem of mine water inrush is to distinguish the source of water inrush and effectively perform water prevention and control in advance, so as to prevent and reduce the occurrence of water disasters [2]. Therefore, it is more important to study the previous coal mine flood accidents, to fundamentally prevent the similar coal mine flood accidents in the future. The methods of water source identification of mine water inrush mainly include water chemistry, isotope, water temperature, dynamic observation of groundwater level, etc [3]. Although these methods have reached a certain level of recognition effectiveness, there are some inherited problems such as long analysis time, low recognition accuracy, complex operation process, etc. It is difficult to achieve early online warning and

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

rapid identification of water inrush sources, therefore, they are not useful to prevent and control water disasters.

Laser-induced fluorescence technology (LIF) has the advantage of high sensitivity, strong real-time performance, and high precision [4]. In order to solve the limitations of the existing methods of water inrush source identification, the technology is used in the field of prevention and control of coal mine flood accidents. For example, Wang *et al*. [5] used SG to preprocess the original fluorescence spectral data and used principal component analysis (PCA) to extract feature information. The model of extreme learning machine (ELM) can be used to distinguish the source of water inrush quickly. Hu *et al*. [6] used a successive projection algorithm to select characteristic wavelength points of LIF combined with extreme learning machine to achieve timely recognition of mine water inrush. Yang *et al*. [7] proposed a water source discrimination model of mine water inrush based on the LIF and convolution neural network (CNN). This model can effectively avoid the complicated feature extraction in the recognition process. However, the machine learning algorithms currently in use are all data processing methods. After dimensionality reduction by feature extraction or feature selection, the original form of fluorescence spectra cannot be reconstructed from spectral data. These methods can only use the new characteristic data to establish the water source identification models, and then by applying the accuracy and other evaluation indexes to achieve the identification of mine water inrush, therefore, they can not distinguish and analyze different kinds of water samples directly through the fluorescence spectra. CNN and other deep learning algorithms are not only complex in parameter adjustment, requiring a large number of samples, and time-consuming, but also requiring to draw spectra map before recognition, therefore, can not directly identify the raw spectral data.

Complementary ensemble empirical mode decomposition (CEEMD) is an improved non-stationary and non-linear signal processing method based on EMD. This method can process the non-stationary data stably and decompose and reconstruct the complex signals into new stationary signals [8]. CEEMD has been widely used in medical electrical signal analysis, atmospheric environment detection, crop pest detection, bearing fault diagnosis, and other research fields. For example, Zhang *et al*. [9] combined CEEMD - Lempel-Ziv complexity and extreme learning machine (ELM) to improve the precision and stability of wind power prediction. Zhu *et al*. [10] put forward a hybrid algorithm of CEEMD and particle swarm optimization and gravitational search algorithm (PSOGSA) to realize the prediction and early warning of air quality. Qiao *et al*. [11] realized the classification and detection of wheat pests by using the spectral analysis and feature extraction method based on CEEMD. LU *et al*. [12] proposed an auxiliary CEEMD combined with unsupervised clustering algorithm to diagnose the damage degree of rolling bearing. cuckoo search (CS) is a heuristic intelligent optimization algorithm [13], which can be used to optimize the parameters of a multi-class support vector machine (MSVM),

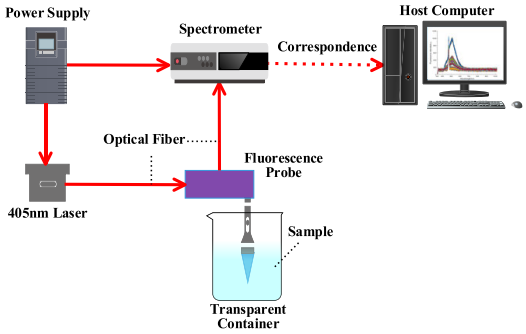


FIGURE 1. Schematic diagram of laser-induced fluorescence spectral analysis system of mine water inrush.

it can effectively avoid being trapped in the local optimal solution, and therefore, improve the classification accuracy. For example, Zhang *et al*. [14] proposed the combination of singular spectrum analysis and CS for short-term electric load forecasting. Arnab *et al*. [15] used a cascaded CS algorithm based on the minimum signal attenuation to realize the efficient transmission of indoor wireless sensor network data. Wang *et al*. [16] used CS-SVM fluorescence spectroscopy to identify polycyclic aromatic hydrocarbons. Yin *et al*. [17] realized the non-destructive identification of rubber by using terahertz spectroscopy technology combined with PCA and CS-SVM.

The CEEMD algorithm adopted in this paper can decompose the raw fluorescence spectral signals into signals of different frequencies, and use the reconstructed spectral signals to establish the CS-MSVM model for water inrush classification. This method provides a theoretical basis for the identification of mine water inrush by LIF spectra.

II. MATERIALS AND METHODS

A. INSTRUMENT AND EQUIPMENT

The schematic diagram of the LIF spectral analysis system of mine water inrush shown in Fig. 1, which is used to complete the collection of spectral data. The system is mainly composed of laser, spectrometer, fluorescence probe, optical fiber, and host computer. The USB2000+ micro-optical fiber spectrometer (Ocean Optics, USA) is selected in the experiment, which contains a linear CCD array with high sensitivity of 2048 pixels (model ILX511, Sony Corporation, Japan). The spectral detection range is 340∼1021nm, the resolution is 0.5nm, the integration time is 1s/1000nm, and the blue-violet laser diode (Beijing Huayuan Tuoda Laser Technology Co., Ltd. China) with the wavelength of 405nm is used. The adjustable range of laser incident power is 100∼130mv, and the experimental setting value is 120mW. The laser is connected with the fluorescence spectrometer through the optical fiber interface of SMA905 connector, and the immersion micro fluorescence probe (model FPB-405-V3, Guangzhou Biaoqi Optoelectronics Technology Development Co., Ltd., China) is used to insert the water samples to be tested to obtain the fluorescence signals.

To avoid the interference of other light sources on the acquisition of fluorescence spectra in the experiment, the experiment must be done in a dark room away from the light. During the experiment, the probe was vertically immersed in the water samples, and experiments were performed repeatedly to ensure that the height of the probe invading the transparent container was always consistent. The hardware conditions of the computer used in the experiment include an Intel Core i7 processor, an NVIDIA RTX 2070 graphics card, a 16G Kingston memory module, etc. The algorithm simulation is run in MATLAB R2019a (MathWorks, USA) environment, using the libsvm-mat-3.1 support vector machine toolkit [18].

B. PREPROCESSING FOR SPECTRAL SIGNALS

In the process of fluorescence spectral data acquisition, there are some problems in the spectrometer, such as spectral amplitude jitter, random error, and light interference. This will cause the received spectral signals to jitter in different degrees in the whole band range, and these spectral signals contain some redundant information. Therefore, the collected raw spectral signals need to be preprocessed. The common preprocessing methods of spectral signals are Savitzky-Golay (SG) smoothing [19], median filter (Median) smoothing [20], robust locally weighted regression (Rlowess) smoothing [21], moving average smoothing [22] and so on.

In order to verify the superiority of the signal processing methods used in this paper, we will make a horizontal comparison and analysis with SG smoothing, Median smoothing, Rlowess smoothing, and moving average smoothing. The identification models of mine water inrush are established for the spectral data after signal processing, and the classification accuracy is taken as the evaluation index of the processing methods. Finally, the most suitable preprocessing method of the spectral signal is selected.

C. MATERIALS AND SAMPLES

In the mine water inrush accident, the goaf water has the feature of fast-moving, strong destructive capacity, large static content, etc., and its harmfulness is higher than other water inrush sources [23]. In this experiment, the goaf water, the limestone water, the sandstone water, and the mixture of the above three were used as the experimental materials. The water inrush samples collected in Huainan mining area on August 21, 2019, were used for the experiment. The goaf water is mixed with limestone water and sandstone water with different proportions to produce seven different kinds of water samples. 50 groups of spectral data are collected from each individually kind of water sample, result in a total of 350 groups of spectral data.

The first is the pure goaf water, the second is the pure limestone water. The third is the pure sandstone water. The fourth is the mixture of goaf water and sandstone water with the volume ratio of 1:1. The fifth is the mixture of goaf water and limestone water with the volume ratio of 1:1. The sixth is the mixture of goaf water and sandstone water with the volume ratio of 1:2. The seventh is the mixture of goaf water and limestone water with the volume ratio of 1:2.

D. CEEMD

Complementary ensemble empirical mode decomposition (CEEMD) is an improved signal decomposition method for ensemble empirical mode decomposition (EEMD) [24] proposed by Yeh *et al*. [25] in 2010. Based on the noise aided analysis, this method can solve problems of residual white noise and too much calculation in EEMD decomposition, and also effectively suppresses the problem of mode mixing [26]. The decomposition process of the CEEMD algorithm is based on empirical mode decomposition (EMD) [27] and EEMD. A pair of auxiliary noises with the same amplitude and opposite sign is added to the signal. This method can effectively eliminate the residual auxiliary noise in the reconstructed signals, reduce the reconstruction error, and improve the performance [28].

The decomposition process of CEEMD is as follows:

1) Add a group of auxiliary noise signals with opposite symbols to the original signals, and the amplitude of each new noise added should be the same.

$$x_i^+(\mathrm{t}) = x(\mathrm{t}) + \mathrm{n}_i^+(\mathrm{t}) \tag{1}$$

$$x_i^-(\mathrm{t}) = x(\mathrm{t}) + \mathrm{n}_i^-(\mathrm{t}) \tag{2}$$

where $x(\mathrm{t})$ is the original signal, $n_i^+(\mathrm{t})$ is the positive noise, and $n_i^-(\mathrm{t})$ is the negative noise;

2) The added noise signals $x_i^+(\mathrm{t})$ and $x_i^-(\mathrm{t})$ are decomposed by EMD to obtain two groups of IMF components $e_i^+(\mathrm{t})$ and $e_i^-(\mathrm{t})$. $e_i^+(\mathrm{t})$ is obtained by the decomposition of $x_i^+(\mathrm{t})$ and $e_i^-(\mathrm{t})$ is obtained by the decomposition of $x_i^-(\mathrm{t})$;

3) Repeat steps 1) and 2)N times to get N IMF components $e_{ni}^+(\mathrm{t})$ and $e_{ni}^-(\mathrm{t})$;

4) Calculate the total integrated average value of components $e_{ni}^+(\mathrm{t})$ and $e_{ni}^-(\mathrm{t})$ respectively.

$$E_i^+(\mathrm{t}) = \frac{1}{N}\sum_{j=1}^{N} e_{ni}^+(\mathrm{t}), \quad j = 1, 2, \ldots, \mathrm{N} \tag{3}$$

$$E_i^-(\mathrm{t}) = \sum_{j=1}^{N} e_{ni}^-(\mathrm{t})/\mathrm{N}, \quad i = 1, 2, \ldots, \mathrm{N} \tag{4}$$

5) Take the average of $E_i^+(\mathrm{t})$ and $E_i^-(\mathrm{t})$ as the final result.

$$\mathrm{IMF} = \frac{1}{2}(E_i^+(\mathrm{t}) + E_i^-(\mathrm{t})) \tag{5}$$

E. CS-MSVM

The cuckoo search (CS) is a heuristic optimization algorithm to simulate the parasitic propagation of cuckoo, which was proposed by Yang [29] of the Engineering Department of Cambridge University in 2009. This algorithm can enhance the global search ability through Levy flight, therefore, it can not only effectively reduce the local search time, but also avoid being trapped in the local optimum in the search process.

The multi-class support vector machine (MSVM) is a classification algorithm based on the original two classification support vector machines, which solves the limitation of only two classifications [30]. Combining the advantages of the two algorithms, we use the CS algorithm to optimize the penalty factor C and the RBF kernel function parameter g in MSVM, and find a group of optimal C and g as the optimal parameters to build a CS-MSVM model. Compared with the KNN, Bayesian, decision tree, and random forest, the learning performance of these classification algorithms depend on the manual adjustment of parameters [31]–[34]. CS-MSVM model does not need manual adjustment of parameters. CS optimization algorithm can automatically find the optimal solution of MSVM model parameters, which not only avoids the complex process of manual parameter adjustment, but also reduces the error of subjective parameter adjustment, so that the performance of the classifier is fully improved, and has a good generalization ability [35].

The specific steps are as follows:

1) Determine and set the search range of penalty factor C and kernel function parameter g;

2) Set iterations *time*, the number of nests n, the probability of being discovered $P(\lambda)$, and the number of optimization parameters d;

3) For each nest, the values of penalty factor C and function parameter g are initialized randomly as parameters of MSVM to obtain the accuracy value, and the prediction error rate is taken as the fitness value. Finding the optimal nest is the value of the current optimal C and g.

$$F_N = 1 - \frac{1}{100} fitness \tag{6}$$

where F_N is the fitness at the moment, *fitness* is the accuracy of MSVM under the parameters of C and g;

4) Use Levy flights to update the nest, taking the updated values of C and g as parameters of SVM, repeat step 3) to obtain a new set of nests;

5) Randomly eliminate some nests with probability, and repeat step 3) to get a new set of nests. The new fitness value is compared with the fitness value obtained in step 4), and the optimal nest is obtained;

6) If the function value of the optimization objective meets the end condition, the global optimal nest and fitness value will be output, otherwise, otherwise, step 4) is returned to optimize the parameter values of C and g.

F. EVALUATION INDEX

The effect of a signal processing method is determined by the comparison of some digital evaluation indexes, such as signal to noise ratio (SNR) [36], root means square error (RMSE) [37] and Pearson correlation coefficient (r) [38]. SNR indicates the relationship between signal and noise, RMSE indicates the difference between noise reduction signal and original signal, and r indicates the similarity of signals before and after noise reduction. In general, the larger the SNR value, the truer the signal, and the less distortion. The smaller the RMSE value, the better the signal filtering effect. The larger the r value, the greater the linear correlation between the signals.

The expression of the SNR:

$$SNR = 10\log_{10}[\frac{\sum_{i=1}^{N} w^2(\mathrm{n})}{\sum_{i=1}^{N}(w(\mathrm{n})-w^*(\mathrm{n}))^2}] \tag{7}$$

The expression of the RMSE:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[w^2(\mathrm{n})-w^*(\mathrm{n})^2]} \tag{8}$$

The expression of the r:

$$r = \frac{\sum_{i=1}^{n}(w_i(\mathrm{n}) - \overline{w(n)})(w_i^*(\mathrm{n}) - \overline{w^*(\mathrm{n})})}{\sqrt{\sum_{i=1}^{n}(w_i(\mathrm{n}) - \overline{w(\mathrm{n})})^2}\sqrt{\sum_{i=1}^{n}(w_i^*(\mathrm{n}) - \overline{w^*(\mathrm{n})})^2}} \tag{9}$$

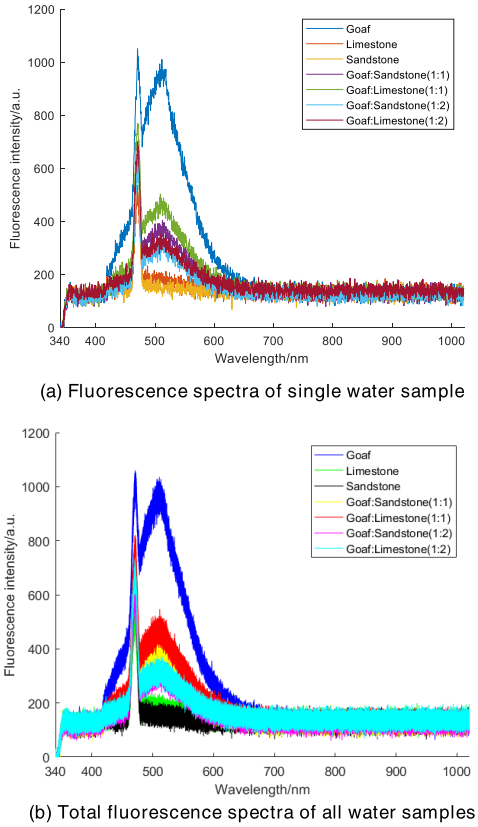
where $w(n)$ is the raw signal, $w_i^*(\mathrm{n})$ is the processed signal, $\overline{w}(n)$ is the average value of the raw signal, and $\overline{w}_i^*(\mathrm{n})$ is the average value of the processed signal.

III. RESULTS AND DISCUSSION

A. RAW SPECTRAL SIGNALS OF WATER INRUSH

The spectral data are collected by the LIF spectral analysis system of mine water inrush, and the collected raw fluorescence spectra are shown in Fig. 2.

The fluorescence spectrum of a single water sample is shown in Fig. 2 (a), the band range is 340 $\sim$ 1021nm, and the difference of water samples is mainly concentrated in the band of 400 $\sim$ 650nm. The trend of other bands is gentle, and the signal overlaps seriously. There are two peaks at 475nm and 514nm for the goaf water, the mixture of goaf water and limestone water, and the mixture of goaf water and sandstone water. The difference between sandstone water and limestone water is mainly reflected in the fluorescence intensity of only one peak at 475nm, and the difference between the goaf water and other water samples is obvious. The spectral signals of sandstone water, limestone water, and three kinds of mixed water are overlapped, therefore, it is difficult to distinguish. It can be clearly seen from Fig. 2(b) that total fluorescence spectra of 350 water samples, except the goaf water, there is little difference among the other six water samples, which contain a large number of redundant signals, so it is difficult to distinguish the types of water samples through observation. Therefore, it is necessary to preprocess the fluorescence spectral signals of mine water inrush with the help of a signal processing algorithm to eliminate the interference of redundant signals and reduce the spectral overlap.



(a) Fluorescence spectra of single water sample

(b) Total fluorescence spectra of all water samples

FIGURE 2. Raw fluorescence spectra.

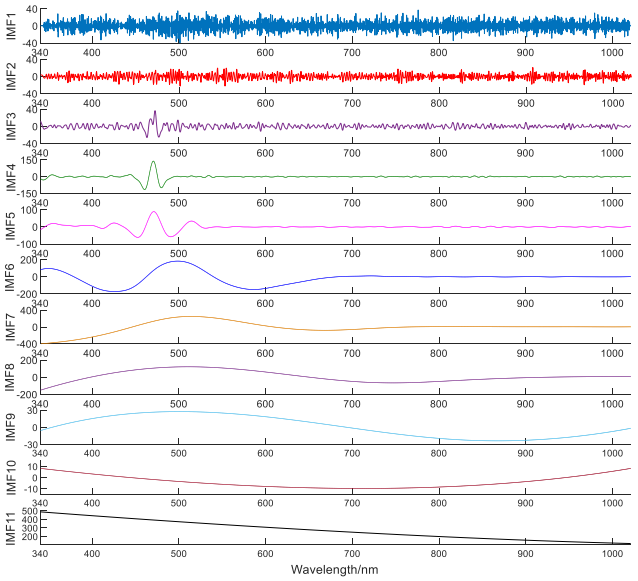


FIGURE 3. Adaptive decomposition of CEEMD.

B. SPECTRAL SIGNALS PREPROCESSING BY CEEMD

By using the principle of the CEEMD algorithm, the fluorescence spectral signals corresponding to each water inrush sample can be adaptively decomposed into a certain number of IMF components. The standard deviation of adding white noise is 0.2 times the original signal standard deviation, and the number of iterations is 100.

The fluorescence spectral signals of the first water inrush samples are taken out for analysis. As can be seen from Fig. 3, the raw fluorescence spectral signals are adaptively decomposed into 11 IMF components. The decomposed IMF components are arranged in the order of frequency from high to low. Each component has its own amplitude and frequency. The random noise of spectra is mainly distributed in the first three IMF components, which have a higher frequency and smaller amplitude and fluctuate violently. Other IMF components belong to low-frequency components. With the increase of the component number, the fluctuation becomes more and more gentle, and the frequency becomes smaller and smaller, which represents the useful signal part. IMF11 is generally called a trend term, which generally shows a slow downward trend. This shows that with the increase of wavelength of fluorescence spectra, the absorption of fluorescence substances in the water inrush to the laser shows a downward trend. Compared with raw fluorescence spectra, only the static category information of the corresponding spectra of different water samples can be obtained, while the CEEMD method can get the implicit information in the spectral data.

Comparing EMD, EEMD, and CEEMD according to the evaluation index of signal processing effect. As shown in Table 1, the three methods decompose the SNR of different IMF with negative values, which indicates that the signal power in these IMF is less than the noise power, and the corresponding RMSE value is also significantly larger, which belongs to the redundant component signal. The positive values of the first three order IMF components are significantly lower than those of the other order components, and all of them are less than 0.1, indicating that there are some noise signals in these IMF components. The number of IMF components of negative SNR of the EMD method is more than that of other two methods, which may lead to mode mixing. It can be seen that the SNR of different IMF decomposed by the raw spectral signals is smaller, and the RMSE is larger. It shows that a single IMF is not enough to represent all the information of the raw spectral signals, and these components have serious distortion problems. It is necessary to reconstruct these signals and select some useful IMF components to eliminate the interference of redundant signals to the subsequent recognition results of mine water inrush.

Firstly, the redundant IMF component signals of negative SNR are eliminated, and then the remaining IMF component signals are summed and reconstructed. Based on the IMF11 component corresponding to the maximum value of SNR and the minimum value of RMSE, different IMF components are gradually superposed according to the reconstruction sequence of increasing SNR value and decreasing RMSE value. The average SNR, the average RMSE, and the average r are used as the evaluation indexes of decomposition and reconstruction of spectral signals.

TABLE 1. Evaluation indexes decomposed by different decomposition methods.

Different modes	EMD		EEMD		CEEMD	
	SNR	RMSE	SNR	RMSE	SNR	RMSE
IMF1	0.0028	340.0403	0.0047	339.9678	0.0076	339.8521
IMF2	0.0042	339.9845	0.0020	340.0719	0.0031	340.0295
IMF3	0.0027	340.0438	0.0017	340.0843	0.0040	339.9923
IMF4	-0.0105	340.5610	-0.0094	340.5177	-0.0320	341.4042
IMF5	-0.0134	340.6772	0.0114	339.7058	0.0235	339.2320
IMF6	-0.0181	340.8586	0.0819	336.9596	0.0434	338.4553
IMF7	0.0404	338.5730	0.6138	316.9439	0.9343	305.4619
IMF8	-0.0642	342.6743	0.5490	319.3159	0.9997	303.1703
IMF9	1.3289	291.8960	0.0966	336.3892	0.2164	331.7785
IMF10	0.7628	311.5528	-0.0606	342.5305	-0.0786	343.2431
IMF11	4.6344	199.5046	2.8970	243.6794	4.2030	209.6634

TABLE 2. Evaluation indexes reconstructed by different decomposition methods.

Method	Average SNR	Average RMSE	Average r
EMD	11.1998	93.7143	0.9210
EEMD	5.7454	196.9972	0.9726
CEEMD	**19.0509**	**51.5819**	0.9503

As can be seen from Table 2. The evaluation indexes of reconstructed spectral signals are obviously better than those of single IMF component signals, which shows that signal reconstruction is a necessary signal processing work. the difference of average r of spectral signals reconstructed by the three decomposition methods is very small, and the difference between the average SNR and the average RMSE is obvious. In conclusion, the filtering effect of the CEEMD method is good. Finally, CEEMD is selected as the processing method of spectral signals of mine water inrush, which makes preparation for the construction of the subsequent classification model.

C. ESTABLISHMENT OF CS-MSVM CLASSIFICATION MODEL

The fluorescence spectral signals of 350 groups of water inrush samples are decomposed by CEEMD, and then all water samples are divided into the training set and the test set according to the ratio of 3:2. Among them, 210 groups of samples are randomly selected as the training set, the remaining 140 groups of samples are selected as the test set, and the data of training set and test set are used as input of the CS-MSVM model. When CS-MSVM is trained in classification, the number of nests is set to 20, and the number of iterations is 50. Radial basis function (RBF) is selected as the kernel function of MSVM, and the fluorescence intensity value is normalized to [0,1] interval for training. It can be

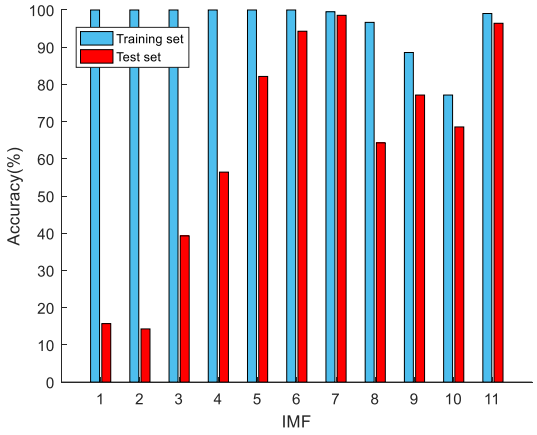


FIGURE 4. Classification accuracy of different modes.

seen from Fig. 4 that all water sample spectral signals are divided into 11 IMF components. Although the accuracy of the first three IMF components in the training set is very high, the accuracy of the test set is very low, which indicates that there is a large noise interference signal in the training data, resulting in a serious overfitting phenomenon. The accuracy of the test set of IMF2 is the lowest. Because of the redundant signals in samples, the accuracy of the training set of IMF10 is the lowest. The accuracy of IMF6, IMF7, and IMFf11 signal components is more than 90%, which shows that the CEEMD processing method can decompose the IMF components that affect the performance of the CS-MSVM classification model.

The evaluation indexes SNR and RMSE are used as the basis for the reconstruction order of the spectral signals of all the water inrush samples. Firstly, the IMF10 component of the negative SNR is removed. Finally, IMF11 with the largest SNR and the smallest RMSE is used as the basis for spectral reconstruction. The reconstructed signals are gradually superposed in the order of increasing SNR and decreasing MSE.

The reconstruction process of each stage is shown in Fig. 5, which is divided into 9 stages of (a) ∼ (i), and the last one is raw spectra. The first stage is the reconstruction of IMF11 and IMF7. The (b) refers to add IMF8 on the basis of (a). The (c) refers to add IMF6 on the basis of (b). The (d) refers to add IMF9 on the basis of (c). The (e) refers to add IMF5 on the basis of (d). The (f) refers to add IMF4 on the basis of (e). The (g) refers to add IMF3 on the basis of (f). The (h) refers to add IMF1 on the basis of (g). The (i) refers to add IMF2 on the basis of (h). The spectral signals reconstructed in stages of (a) ∼ (d) have only one peak, and the signal spectral lines are scattered. From the (e) stage, one peak of the spectral signal gradually becomes two peaks, and the spectral lines of the signal become denser and denser. Because the noise signal component is added, the spectral signal reconstructed in stages of (h) ∼ (i) is very similar to the raw signal, and the distortion of the signal processed in stages of (e) ∼ (i) is relatively small, which retains the

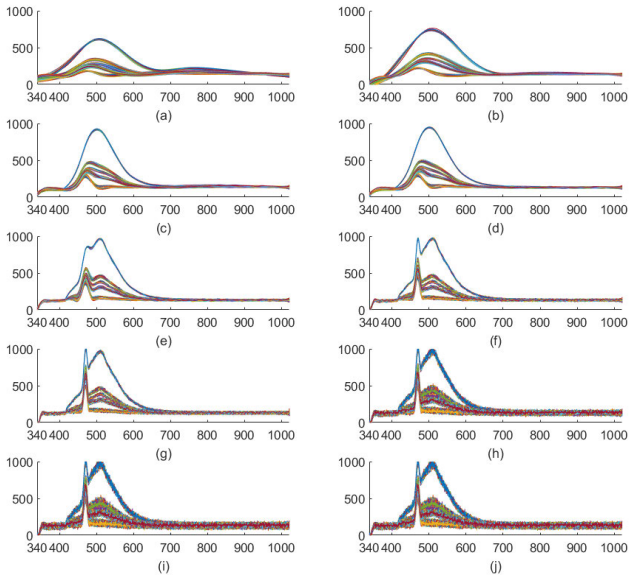


FIGURE 5. Reconstructed spectra of all water samples.

TABLE 3. Classification accuracy results of reconstructed spectra.

Number of reconstruction stages	Best C	Best g	training set/%	test set/%
1	22.2718	0.01	100(210/210)	100(140/140)
2	94.3932	0.01	100(210/210)	99.29(139/140)
3	44.7639	0.01	100(210/210)	99.29(139/140)
4	6.6208	0.01	100(210/210)	99.29(139/140)
5	23.2518	0.01	**100(210/210)**	**100(140/140)**
6	38.1214	0.01	100(210/210)	99.29(139/140)
7	17.5660	0.01	100(210/210)	98.57(138/140)
8	83.6685	0.01	100(210/210)	98.57(138/140)
9	26.4526	0.01	100(210/210)	98.57(138/140)

characteristics of the data itself. Although the spectral lines presented in stages of (a) $\sim$ (g) are more scattered than raw spectra, there are still overlapping spectra which are difficult to distinguish intuitively. Therefore, it is necessary to use the pattern recognition classification algorithm to accurately identify the water inrush samples.

The classification accuracy results of reconstructed spectra in different stages are shown in Table 3, the classification accuracy of the training set and test set of the reconstructed spectral signals is significantly higher than that of the single IMF component. The accuracy of the first stage and the stage of (e) can reach the highest value. Due to the mixing of noise IMF component, the accuracy of the last four stages is slightly reduced. Compared with reconstructed spectra of the water sample in Figure 5, serious distortion occurs in the first stage of spectra. To ensure that spectra can display more information and achieve high classification accuracy, the fifth stage reconstructed spectra are selected as the final signal

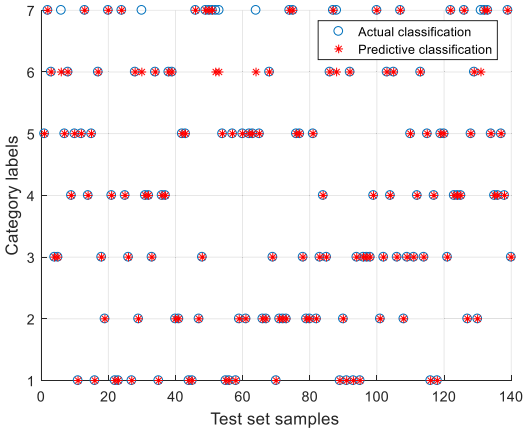


FIGURE 6. Classification results of the test set of raw spectra.

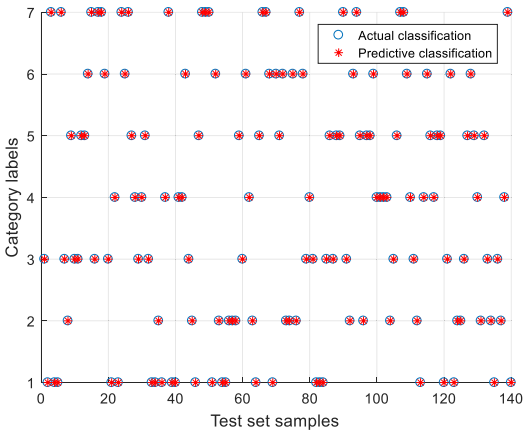


FIGURE 7. Classification results of the test set of spectra processed by CEEMD.

processing result, at which point, the model parameters of the MSVM are $C = 23.2518$, $g = 0.01$.

To clearly express the difference between the test category and the actual category, the blue "$\circ$" in the figure is the actual category of the input samples, and the red "$*$" is the predicted result of the classification model. If "$\circ$" and "$*$" coincide, the sample is correctly classified. Fig. 6 shows the classification results of CS-MSVM classification model for test set samples after the training of raw spectral data. Among the 140 water inrush samples, the 7th kind water sample is mistakenly identified as the 6th kind water sample, it shows that the two kinds of water samples have similar chemical components and fluorescent substances, which lead to misclassification.

The classification results of the test set treated by CEEMD is shown in Fig. 7, there is no deviation in 7 kinds of water inrush samples, and all of them are classified correctly.

In order to verify the accuracy and effect of CEEMD algorithm combined with CS-MSVM model for water source identification of mine water inrush, the method in this paper will be compared with Savitzky-Golay (SG) smoothing, median filter (Median) smoothing, robot locally weighted

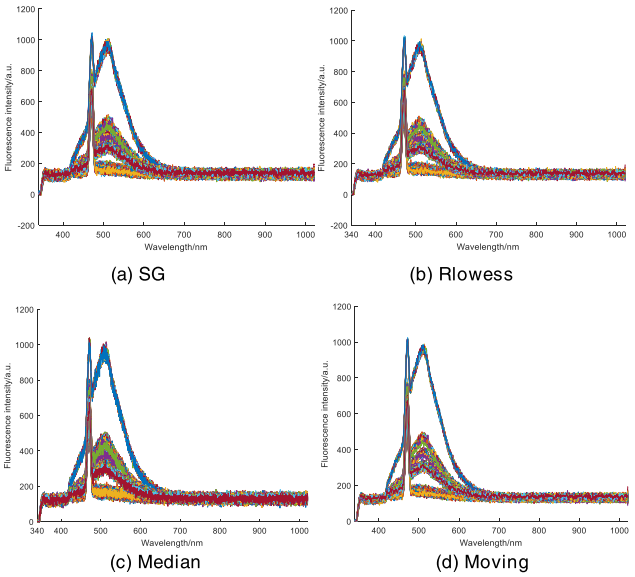


FIGURE 8. Spectra under different signal processing methods.

TABLE 4. Accuracy of different signal processing methods.

Processing method	Best C	Best g	Training set/%	Test set/%
Raw	43.509	0.01	100(210/210)	95(133/140)
CEEMD	38.2416	0.01	**100(210/210)**	**100(140/140)**
SG	87.8629	0.01	100(210/210)	96.43(135/140)
Median	21.6368	0.01	100(210/210)	98.57(138/140)
Rlowess	10.6344	0.01	100(210/210)	97.14(136/140)
Moving	35.5395	0.01	100(210/210)	98.57(138/140)

expression (Rlowess) smoothing, and moving average (Moving) smoothing. The processed spectra are as shown in Fig. 8. Compared with Fig. 5, the CEEMD successfully reduces the overlapping degree of spectral signals of the mixed water samples. However, the other four methods only reduce the burr of the spectral signals and enhance the spectral smoothness, but spectra between the mixed water samples is still stacked seriously and the filtering effect is poor. From the modeling accuracy of different processing methods in Table 4, it can be seen that the various processing methods all can improve the accuracy of the classification model in varying degrees. The classification accuracy of training set and test set after CEEMD processing can reach 100%, and the classification accuracy is also the highest.

We also use random forest (RF), extreme learning machine (ELM), probabilistic neural network (PNN), and decision tree (DT) classifiers to evaluate the preprocessing performance of the signal processing method. The number of trees in RF is set to 400, and each decision tree contains 45 variables. The optimal number of hidden layer nodes in ELM is 143. The optimal parameter value of spread in PNN is

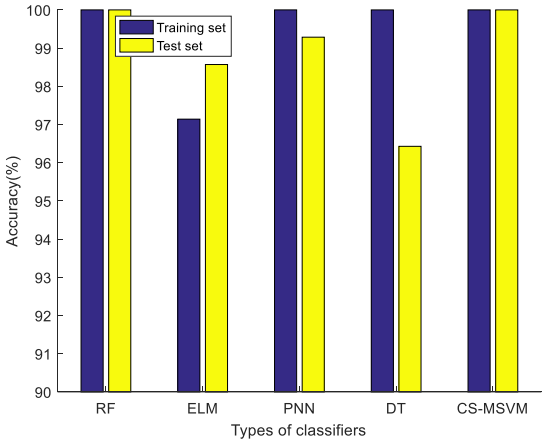


FIGURE 9. Evaluation results of different classifiers.

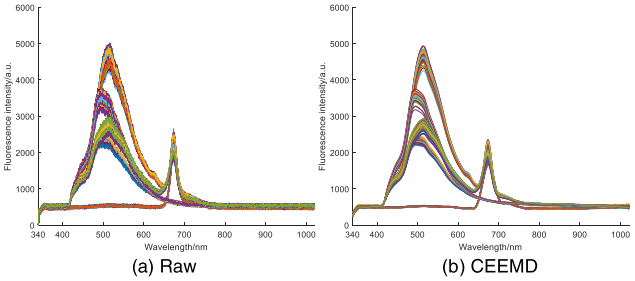


FIGURE 10. Comparison of fluorescence spectra.

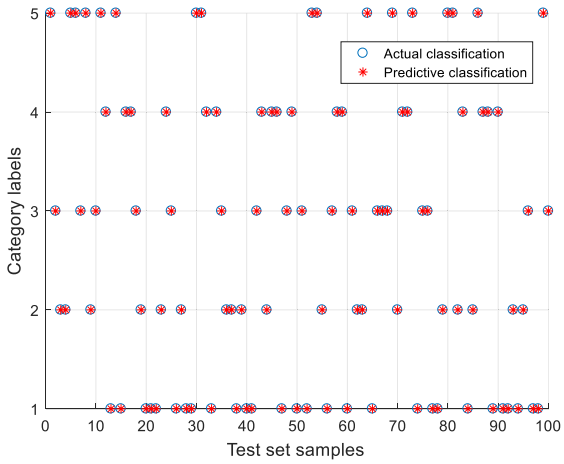


FIGURE 11. Classification results of the test set of spectra processed by CEEMD.

set to 1.5. Fig. 9 provides the classification effect under different classifiers. We can find that in addition to the ELM classifier, the accuracy of other classifiers of training sets can reach 100%, and the accuracy of test sets can reach more than 95%. It shows that the signal preprocessing method proposed in this paper is capable of adapting to different classifiers and has a better performance. Compared with other classifiers, the test set accuracy of the DT classifier is lower, and its value is 96.43%. Only one sample under

the PNN classifier is wrongly identified. The accuracy of the training set and the test set of RF and CS-MSVM both reach 100%, which shows that these two classifiers have an excellent classification learning ability for training samples after preprocessing.

D. VERIFICATION OF GENERALIZATION ABILITY

Good generalization ability can ensure the reliability of the training model. If the algorithm used in this paper can achieve good classification effect for different data sets, it can show that the algorithm has strong reliability.

Now, the reliability of the algorithm is verified by the data of LIF spectra of edible oil collected in the laboratory on September 22, 2019. Seven kinds of edible oil are soybean oil, rapeseed oil, peanut oil, corn oil, and sunflower oil purchased on the market. From 250 samples (50 groups of each kind), 150 groups are randomly selected as training sets,and the remaining 100 groups are selected as test sets. The spectral signal is decomposed and reconstructed by CEEMD, and then the edible oil recognition model of CS-MSVM is established. As shown in Fig. 10 (a), raw fluorescence spectra of edible oil are partially overlapped and hard to distinguish. It can be seen from Fig. 10 (b) that spectra of edible oil treated by CEEMD are more dispersed, and different edible oil can be visually distinguished by naked eyes. The actual classification results of the test set are shown in Fig. 11. When the optimization parameter C is 24.4272 and g is 0.01, all samples are classified correctly, which fully reflects the advantages of this method.

IV. CONCLUSION

In this paper, we proposed a new method of mine water inrush identification based on LIF signal processing. The new fluorescence spectra signals are decomposed and reconstructed by the CEEMD algorithm, and the CS-MSVM model is established. This approach is established for training to identify different kinds of mine water inrush. Finally, we compared and analyzed the effect of different processing methods on classification results, and verified the reliability of the algorithm. The experimental process and results show:

1) The CEEMD algorithm based on SNR analysis is superior to the traditional EMD and EEMD algorithm in various evaluation indexes. Because the spectral signals contain all kinds of recessive information, it is difficult to find all the rules through observation. CEEMD can adaptively decompose complex spectral signals into IMF components with practical physical significance. As a result, we can mine hidden signals that cannot be obtained directly from the raw spectral signals, therefore distinguish useful and useless information. By using this method, the noise in raw spectra and the interference of redundant signals on the identification results are both reduced.

2) The normalized data get rid of the influence of too large difference of sample data, which not only speeds up the search for the optimal solution but also improves the classification accuracy. In the training process, the C and g parameters optimized by CS are used in the modeling of water inrush identification of MSVM. This method not only has high recognition accuracy but also avoids the occurrence of overfitting and underfitting.

3) CEEMD combined with CS-MSVM algorithm is applicable to the identification of mine water inrush by LIF technology. The recognition effect of the method used in this paper is better than the classification model established by the signal smoothing method. It can ensure that the underground workers find out the cause of mine water inrush timely and obtain the type of water inrush source safely and accurately. Ultimately, it will be useful to improve water disaster prevention and control.

4) This method is applicable to identify not only mine water inrush, but also edible oil by LIF technology. The experimental results show that the model established in this paper is highly adaptable, reliable, robust, and applicable to other research areas.

All the spectral data in the experiment are collected by the laser-induced fluorescence spectral analysis system in the laboratory. The collection conditions of the spectral data are better than that of the actual water inrush source in the mine, and the interference is relatively less than that in the mine. In this study, 350 groups of sample data (7 kinds of water inrush samples) are tested, but there may be more kinds of water samples with many uncertain factors in the actual industrial field analysis, which brings a lot of challenges to the identification of mine water inrush source. How to overcome these difficulties will become the focus of the next research, and also the key to improving the recognition rate of mine water inrush source. At present, we are all experimenting with static water. We are going to do some research on the identification of dynamic mine water inrush in the future.

In the simulation test, the IMF component is reconstructed mainly based on the signal-to-noise ratio and RMSE, and we finally achieve an ideal recognition effect. In the following research, we may choose more signal evaluation indexes to reconstruct spectral signals and explore its influence on classification accuracy.

CONFLICT OF INTEREST

The authors declared that they have no conflicts of interest regarding the publication of this paper.

REFERENCES

[1] J. Wu, S. Xu, R. Zhou, and Y. Qin, "Scenario analysis of mine water inrush hazard using Bayesian networks," *Saf. Sci.*, vol. 89, pp. 231–239, Nov. 2016.
[2] S. Zhang, W. Guo, and Y. Li, "Experimental simulation of water-inrush disaster from the floor of mine and its mechanism investigation," *Arabian J. Geosci.*, vol. 10, no. 22, p. 503, Nov. 2017.
[3] C. Lin, Y. Wu, W. Lu, A. Chen, and Y. Liu, "Water chemistry and ecotoxicity of an acid mine drainage-affected stream in subtropical China during a major flood event," *J. Hazardous Mater.*, vol. 142, nos. 1–2, pp. 199–207, Apr. 2007.

[4] Q. Yu, X. Wang, and Y. Duan, "Capillary-based three-dimensional immunosensor assembly for high-performance detection of carcinoembryonic antigen using laser-induced fluorescence spectrometry," *Anal. Chem.*, vol. 86, no. 3, pp. 1518–1524, Jan. 2014.

[5] Y. Wang, M. Zhou, P. Yan, F. Hu, W. Lai, Y. Yang, and Y. Zhang, "A rapid identification model of mine water inrush based on extreme learning machine," *J. China Coal Soc.*, vol. 42, no. 9, pp. 2427–2432, Sep. 2017.

[6] F. Hu, M. Zhou, P. Yan, D. Li, W. Lai, S. Zhu, and Y. Wang, "Selection of characteristic wavelengths using SPA for laser induced fluorescence spectroscopy of mine water inrush," *Spectrochimica Acta A, Mol. Biomol. Spectrosc.*, vol. 219, pp. 367–374, Aug. 2019.

[7] Y. Yang, J. Yue, J. Li, and H. Zhang, "Online discrimination model for mine water inrush source based CNN and fluorescence spectrum," *Spectrosc. Spectral Anal.*, vol. 39, no. 8, pp. 2425–2430, Aug. 2019.

[8] J. P. Amezquita-Sanchez, A. Adeli, and H. Adeli, "A new methodology for automated diagnosis of mild cognitive impairment (MCI) using magnetoencephalography (MEG)," *Behavioural Brain Res.*, vol. 305, pp. 174–180, May 2016.

[9] C. Zhang, M. Ding, W. Wang, R. Bi, L. Miao, H. Yu, and L. Liu, "An improved ELM model based on CEEMD-LZC and manifold learning for short-term wind power prediction," *IEEE Access*, vol. 7, pp. 121472–121481, 2019.

[10] S. Zhu, X. Lian, L. Wei, J. Che, X. Shen, L. Yang, X. Qiu, X. Liu, W. Gao, X. Ren, and J. Li, "PM2.5 forecasting using SVR with PSOGSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors," *Atmos. Environ.*, vol. 183, pp. 20–32, Jun. 2018.

[11] L. Qiao, M. Jia, Y. Liang, and T. Yang, "Spectrum analysis of insect-damaged wheat BPE signal based on CEEMD," *Optik*, vol. 150, pp. 62–70, Dec. 2017.

[12] Y. Lu, R. Xie, and S. Y. Liang, "CEEMD-assisted bearing degradation assessment using tight clustering," *Int. J. Adv. Manuf. Technol.*, vol. 104, nos. 1–4, pp. 1259–1267, Jul. 2019.

[13] X. Li and M. Yin, "Modified cuckoo search algorithm with self adaptive parameter method," *Inf. Sci.*, vol. 298, pp. 80–97, Mar. 2015.

[14] X. Zhang, J. Wang, and K. Zhang, "Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm," *Electr. Power Syst. Res.*, vol. 146, pp. 270–285, May 2017.

[15] A. Ghosh and N. Chakraborty, "Cascaded cuckoo search optimization of router placement in signal attenuation minimization for a wireless sensor network in an indoor environment," *Eng. Optim.*, vol. 51, no. 12, pp. 2127–2146, Feb. 2019.

[16] S.-T. Wang, Y.-Y. Yuan, C.-Y. Zhu, D.-M. Kong, and Y.-T. Wang, "Discrimination of polycyclic aromatic hydrocarbons based on fluorescence spectrometry coupled with CS-SVM," *Measurement*, vol. 139, pp. 475–481, Jun. 2019.

[17] X. Yin, W. Mo, Q. Wang, and B. Qin, "A terahertz spectroscopy non-destructive identification method for rubber based on CS-SVM," *Adv. Condens. Matter Phys.*, vol. 2018, pp. 1–8, Oct. 2018.

[18] 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, Apr. 2011.

[19] B. M. Schettino, C. A. Duque, and P. M. Silveira, "Current-transformer saturation detection using Savitzky-Golay filter," *IEEE Trans. Power Del.*, vol. 31, no. 3, pp. 1400–1401, Jun. 2016.

[20] Z. Zhang, D. Han, J. Dezert, and Y. Yang, "A new adaptive switching median filter for impulse noise reduction with pre-detection based on evidential reasoning," *Signal Process.*, vol. 147, pp. 173–189, Jun. 2018.

[21] K. Motaghi and A. Ghods, "Attenuation of ground-motion spectral amplitudes and its variations across the central alborz mountains," *Bull. Seismolog. Soc. Amer.*, vol. 102, no. 4, pp. 1428–1474, Aug. 2012.

[22] G. Jordan, A. Petrik, B. De Vivo, S. Albanese, A. Demetriades, and M. Sadeghi, "GEMAS: Spatial analysis of the ni distribution on a continental-scale using digital image processing techniques on European agricultural soil data," *J. Geochem. Explor.*, vol. 186, pp. 143–157, Mar. 2018.

[23] J. Chang, B. Su, R. Malekian, and X. Xing, "Detection of water-filled mining goaf using mining transient electromagnetic method," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 2977–2984, May 2020.

[24] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[25] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Adv. Adapt. Data Anal.*, vol. 2, no. 2, pp. 135–156, Apr. 2010.

[26] W. Zhao, Z. Wang, J. Ma, and L. Li, "Fault diagnosis of a hydraulic pump based on the CEEMD-STFT time-frequency entropy method and multiclass SVM classifier," *Shock Vibrat.*, vol. 2016, pp. 1–8, Sep. 2016.

[27] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[28] D. Liu and K. Sun, "Short-term $PM_{2.5}$ forecasting based on CEEMD-RF in five cities of China," *Environ. Sci. Pollut. Res.*, vol. 26, no. 32, pp. 32790–32803, Sep. 2019.

[29] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214.

[30] X. Zhou and D. P. Tuck, "MSVM-RFE: Extensions of SVM-RFE for multiclass gene selection on DNA microarray data," *Bioinformatics*, vol. 23, no. 9, pp. 1106–1114, May 2007.

[31] Z. Deng, X. Zhu, D. Cheng, M. Zong, and S. Zhang, "Efficient kNN classification algorithm for big data," *Neurocomputing*, vol. 195, pp. 143–148, Jun. 2016.

[32] H. Huttunen and J. Tohka, "Model selection for linear classifiers using Bayesian error estimation," *Pattern Recognit.*, vol. 48, no. 11, pp. 3739–3748, Nov. 2015.

[33] Y. Su, J. Shen, H. Qian, H. Ma, J. Ji, H. Ma, L. Ma, W. Zhang, L. Meng, Z. Li, J. Wu, G. Jin, J. Zhang, and C. Shou, "Diagnosis of gastric cancer using decision tree classification of mass spectral data," *Cancer Sci.*, vol. 98, no. 1, pp. 37–43, Jan. 2007.

[34] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2225–2236, Oct. 2010.

[35] M. Komasi, S. Sharghi, and H. R. Safavi, "Wavelet and cuckoo search-support vector machine conjugation for drought forecasting using standardized precipitation index (case study: Urmia lake, Iran)," *J. Hydroinform.*, vol. 20, no. 4, pp. 975–988, Apr. 2018.

[36] R. Lattanzi, G. C. Wiggins, B. Zhang, Q. Duan, R. Brown, and D. K. Sodickson, "Approaching ultimate intrinsic signal-to-noise ratio with loop and dipole antennas," *Magn. Reson. Med.*, vol. 79, no. 3, pp. 1789–1803, Mar. 2018.

[37] J. J. Zhong, S. N. Fang, and C. Y. Linghu, "Research on application of wavelet denoising method based on signal to noise ratio in the bench test," *Appl. Mech. Mater.*, vols. 457–458, pp. 1156–1162, Oct. 2013.

[38] Y. Gu, J. Han, Z. Liang, J. Yan, Z. Li, and X. Li, "Empirical mode decomposition-based motion artifact correction method for functional near-infrared spectroscopy," *J. Biomed. Opt.*, vol. 21, no. 1, Jan. 2016, Art. no. 015002.

KAI BIAN is currently pursuing the Ph.D. degree with the Anhui University of Science and Technology. His research interests include machine learning, photoelectric information processing, spectra analysis, and medical data processing.

MENGRAN ZHOU received the Ph.D. degree in optics from the Anhui Institute of Optics and Fine Mechanics (AIOFM), Chinese Academy of Sciences, in 2006. He is currently a Professor with the Anhui University of Science and Technology. His research interests include photoelectric information processing and coal mine safety monitoring and control.

FENG HU is currently pursuing the Ph.D. degree with the Anhui University of Science and Technology. His research interests include machine learning, photoelectric information processing, and coal mine safety monitoring.

WENHAO LAI is currently pursuing the Ph.D. degree with the Anhui University of Science and Technology. His research interests include pattern recognition, machine learning, and photoelectric information processing.

MANMAN HUANG is currently pursuing the master's degree with the Anhui University of Science and Technology. Her research interests include deep learning, image processing, and machine learning.