IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 23, 2020, accepted May 4, 2020, date of publication May 6, 2020, date of current version May 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992790

Fault Detection on Insulated Overhead Conductors Based on DWT-LSTM and Partial Discharge

NA QU[1,2], ZHONGZHI LI[3], JIANKAI ZUO[3], AND JIATONG CHEN[4]

[1]School of Safety Engineering, Shenyang Aerospace University, Shenyang 110136, China
[2]Liaoning Key Laboratory of Aircraft Safety Airworthiness, Shenyang Aerospace University, Shenyang 110136, China
[3]School of Computer Science, Shenyang Aerospace University, Shenyang 110136, China
[4]School of Aviation Engine, Shenyang Aerospace University, Shenyang 110136, China

Corresponding author: Na Qu (mn_qn@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61901283, and in part by the Research Project on Economic and Social Development of Liaoning Province under Grant 2020lslktqn-056.

ABSTRACT Insulated conductors can improve the stability of power transmission and reduce the construction space compared with traditional bare conductors. Therefore, insulated conductors are used more and more in overhead power transmission. However, a major challenge of using insulated overhead conductors (IOC) is that the ordinary protection devices are not able to detect the phase-to-ground faults and something, such as tree branch, hitting conductor events. This may cause an accident such as power failure or electrical fire and result in serious damage. In this paper, a new approach, which is based on Discrete Wavelet Transform (DWT) and Long Short Term Memory network (LSTM) for detecting of IOC fault according to partial discharge, is presented. Firstly, the original signal is denoised by DWT. Secondly, the denoised signal is decomposed and extracted features on different layers by DWT. Finally, IOC fault is detected by LSTM. This method can improve the detection accuracy of IOC fault which is tested on the ENET public data set and compared with other classification methods.

INDEX TERMS Insulated overhead conductors, partial discharge, discrete wavelet transform, long short term memory network, fault detection.

I. INTRODUCTION

Insulated conductor can improve the stability of power transmission and reduce the construction space compared with traditional bare conductors [1]–[3]. Therefore, insulated conductors are used more and more in overhead power transmission [4], [5]. However, there is a major challenge in using insulated overhead conductor (IOC). When IOC is broken and fall to the ground or something such as a branch hits IOC, it cannot cause overcurrent. Therefore, phase-to-ground and phase-to-phase faults usually cannot be detected by standard protective equipment [6]. This can create a long-term potential reliability threat in some locations where the tree and tree branches frequently hit the conductors or continuously push and bend the conductors. Eventually, the power line may be damaged and broken, which causes power outage or even tree fire [7]. These faults can occur partial discharge (PD) phenomenon [8], [9]. The current value of PD is very small (about 6-10 amps) and easy to be interfered by external background noise, which is the main problem to achieve accurate detection [10].

In recent years, deep learning algorithms, such as deep belief network, deep neural network, convolutional neural network, deep adversarial convolutional neural network, and transfer network, have been gradually applied to fault diagnosis [11]–[15]. In 2016, Li et al proposed a convolutional neural network (CNN) with deep architecture which is established to extrapolate new features automatically to realize ultra-high frequency (UHF) signals recognition in GIS [16]. In 2017, a novel method based on multi-kernel multi-class relevance vector machine (MMRVM) was proposed for partial discharge pattern recognition [17]. In 2018, Wan *et al.* proposed an approach to detecting PD patterns in gas-insulated switchgear (GIS) using long short-term memory (LSTM) and

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

recurrent neural network (RNN) [18], [19]. In the same year, the stacked denoising autoencoder (SDAE) based deep learning method for PD pattern recognition of different insulation defects of high voltage cables was presented in [20]. Based on deep learning, the network of [21] was constructed for pattern recognition straight forwardly. Through on site detection and simulation experiments, image data sets of five partial discharge defects are established and comparative experiments are conducted. Adam *et al.* applied LSTM to identify different types of PD activity in insulated cables, using single PD impulses as input data. The experimental result shows that the recognition accuracy of LSTM is slightly lower than the random forest method, but LSTM method has the advantage of not requiring artificial statistical features [22].

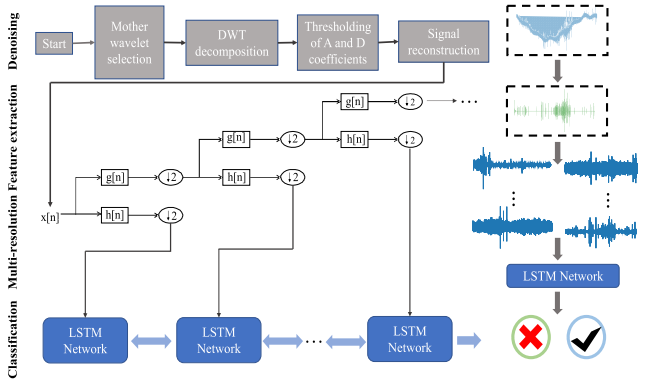


FIGURE 1. The block diagram of complete description.

The spectrum form is usually adopted in the online monitoring system for partial discharge of power equipment. As the discharge time increases and the discharge intensity changes, the PD spectrum will also change accordingly. When the PD spectrum fluctuates with time, its characteristic parameters (such as phase, amplitude, and number of discharges) will also change with time series. The discrete wavelet transform (DWT) can decompose the original signal and extract the features with different resolutions [23]–[25]. LSTM has excellent capability of time series information mining. This paper presents a novel method that combines DWT with LSTM of many-to-one input and output, and increases time series details of PD activity. The method is used to detect PD in the IOC fault. The overall framework of the proposed method is shown as Figure 1. Firstly, the original signal noise is reduced using wavelet method. Secondly, the noise reduction signal is decomposed by DWT and signal features with different resolutions are obtained. Finally, the signal features are input into the many-to-one LSTM model, and the detection result is obtained.

Our contributions are summarized as follows:

(1) We propose an effective learning method for power grid fault diagnosis with noisy signal. This method is implemented by combining DWT and LSTM.
(2) The oversampling technology is proposed to solve the problem of serious imbalance between the number of fault samples and the number of normal samples. The comprehensive evaluation index is designed to evaluate the model performance.
(3) The DWT method is used to solve the problem of the original data with large noise. It can improve the accuracy of LSTM classifier.
(4) It is proved that the accuracy of LSTM detection model can be improved by combining the different levels of signal obtained from DWT decomposition. The best combination level is obtained by designing the contrast experiment.

This paper is organized as follows. Section II introduces the data set. Section III describes the signal noise reduction based on DWT. Section IV describes the signal decomposition and feature extraction based on DWT. Section V describes the IOC fault detection based on LSTM model. Section VI describes the experiment, results and discussion. Section VII concludes the paper.

II. ENET DATASET

Technical University of Ostrava (VSB) devised a special meter to measure the voltage signal of the stray electrical field along IOC, hoping to detect the hazardous PD activities. In 2018, VSB released the ENET data set on Kaggle which is the world's largest data science collaboration platform. The data set contains 8,711 labeled voltage signals from four different locations. Those locations represent the deployment in the real environment (forested and hardly accessible terrain). Each signal is voltage waveform of 50Hz, which contains 800,000 data points and pre-marked as PD (525) or Non-PD (8,186). Due to the large volume of the data set, the Hadoop Distributed File System (HDFS) storage format is used. Examples of the PD signal and the Non-PD signal are shown in Figure 2. It can be seen that the maximum and minimum value of Non-PD signal is about $40mv$ and $-40mv$, and the fluctuation is relatively stable. In contrast, the maximum and minimum value of PD signal is about $60mv$ and $-80mv$, and the signal fluctuation increases significantly.

III. THE SIGNAL NOISE REDUCTION BASED ON DWT

As an important signal processing method, wavelet transform has the characteristics of multi-layer and multi-resolution analysis. By zooming and panning the wavelet function, signal details in the time and frequency domains can be analyzed [26]–[28]. Wavelet transform mainly includes continuous wavelet transform (CWT), discrete wavelet transform (DWT) and discrete wavelet packet transform (DWPT). CWT requires continuous integration and the calculation is complicated. DWT is the discrete processing of CWT. Combined with Mallat algorithm, computational complexity of DWT can be reduced. The DWT expression is as (1).

$$y(n) = \sum_{k=1}^{N/2^j} a_j(k)\varphi_{j,k}(n) + \sum_{j=1}^{J}\sum_{k=1}^{N/2^j} d_j(k)\psi_{j,k}(n) \qquad (1)$$

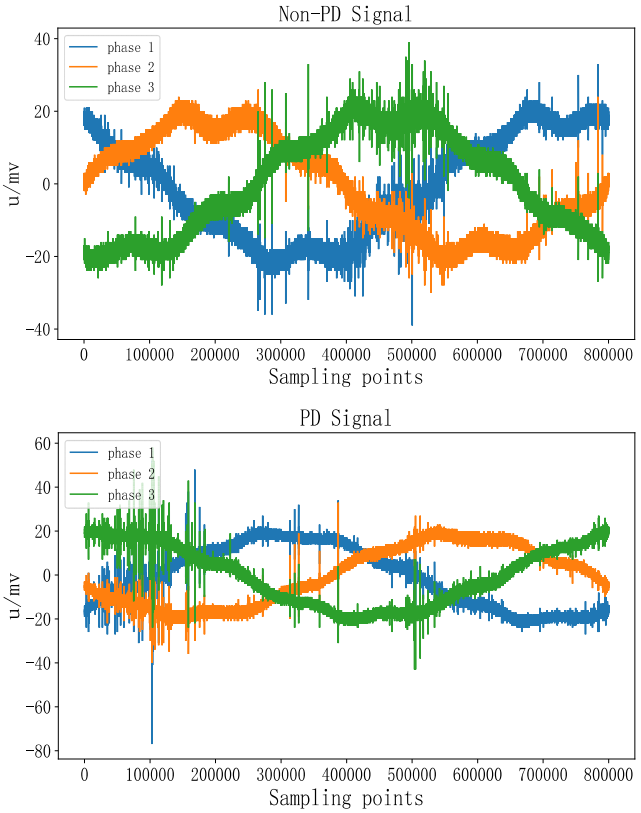


FIGURE 2. Examples of signal records from ENET dataset.

where, $\varphi_{j,k}(n)$ and $\psi_{j,k}(n)$ are the scaling function and wavelet function respectively, J represents the series of wavelet decomposition, and N represents the total number of coefficients of wavelet decomposition, $a_j(k)$ and $d_j(k)$ are the approximate coefficient part and the detailed coefficient part respectively, which can be expressed as (2).

$$\begin{cases} a_j(k) = \langle y(n), \varphi_{j,k}(n) \rangle \\ d_j(k) = \langle y(n), \psi_{j,k}(n) \rangle \end{cases} \tag{2}$$

A. NOISE REDUCTION ALGORITHM BASED ON DWT

High-frequency signals with small amplitude and low-frequency signals with large amplitude can be obtained after signals are decomposed by wavelet. A series of wavelet coefficient prediction values are obtained after mapping through the threshold function. These prediction values are used to reconstruct noise reduction signals. Signal noise reduction process is shown as Figure 3. The right part shows the change of signal waveform during noise reduction. The blue curve represents the original signal, the red curve represents the signal after passing through the high-pass filter, and the green curve represents the signal after passing through DWT noise reduction. The left part is the signal processing flow. The noise reduction algorithm includes wavelet decomposition, threshold quantization processing and wavelet reconstruction. First of all, input the raw voltage signal. Then determine the number of decomposition layers, the scale equation and wavelet threshold. As the mother wavelet of signal decomposition, the noise signal in the original signal is filtered out to obtain useful signals, and these useful signals are reconstructed.

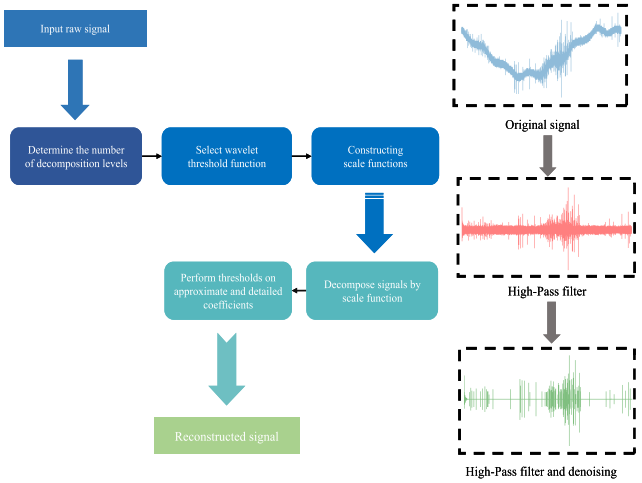


FIGURE 3. The signal noise reduction process.

In order to obtain the optimal frequency resolution and noise reduction, *Daubechies*6 wavelet function is adopted. The wavelet decomposition structure is shown in Figure 4. Y is a discrete sequence of noisy signals, $a_j(k)$ and $d_j(k)$ are the approximate coefficients and detail coefficients on the scale j ($j = 1, 2, \ldots,$ n) respectively.

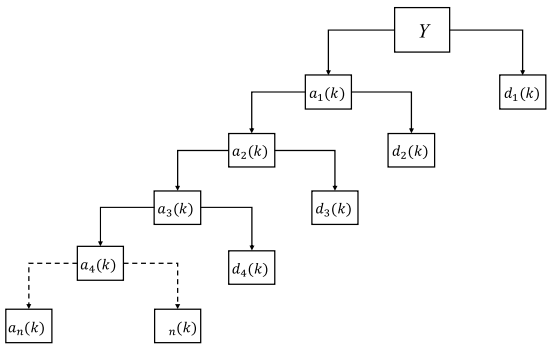


FIGURE 4. Wavelet decomposition structure.

The original signal can be decompose into n detail coefficients d_i and n approximation coefficients a_i. If the threshold of d_i is W, then

$$\bar{d}_{ij} = \begin{cases} 0, & d_{ij} \le W \\ d_{ij} - W, & d_{ij} > W \end{cases} \tag{3}$$

where $_{dij}$ is the wavelet coefficient without threshold processing, and $\bar{d}_{ij}$ is the value after truncation processing.

The noise reduction result can be obtained from the threshold processing result and the *n-th* construction signal. Let $\varphi(x)$ be a scale function of multi-resolution analysis, then

$$\phi(x) = \sum_k P_k \varphi(2x - k) \tag{4}$$

where k is a real number, $(2x - k)$ is a standard orthogonal basis for multi-resolution analysis, and P_k is a scale function.

The threshold W is an important parameter in the noise reduction process [29], [30]. Based on [31] and the average absolute deviation, the threshold W of the decomposition layer n is calculated as (5) and (6).

$$\sigma = 1/0.6745MAD(|d_i|) \tag{5}$$
$$W = \sigma\sqrt{2\log(n)} \tag{6}$$

The hard threshold method is used to deal with the approximate and detailed coefficients, as(7).

$$T_{hard}(x) = \begin{cases} x & if\ |x| > W \\ 0 & otherwise \end{cases} \tag{7}$$

The two-scale equation of Daubechies wavelet is

$$\varphi(x) = \sqrt{2}\sum_{k\in Z} P_k\varphi(2x - k) \tag{8}$$

where Z is the set of integers.

The steps for coefficients construction of the two-scale equation P_k are as follows:

① Select a positive integer ($n \geq 2$) and a polynomial:

$$P(y) = P_n(y) + y^n R(1/2 - y) \geq 0, \quad 0 \leq y \leq 1 \tag{9}$$

② Select $Q(z)$, let $P(\sin^2\frac{\omega}{2}) = |Q(e^{i,w})|^2$;

③ Let $\tilde{H}(z) = (\frac{1+z}{2})^n Q(z)$, then

$$\tilde{H}(z) = \frac{\sqrt{2}}{2}\sum_{k=0}^{L} P_k z^k \tag{10}$$

where $\omega \in [0, 2\pi]$, L is the maximum value of non-zero data in the sequence, z is the noise signal, R, $Q(z)$, $P_n(y)$ is algebraic polynomial of real coefficients, y is a real number between [0,1], and $\tilde{H}(z)$ is high-pass filter with linear phase. As the number of decomposition layers n increases, the smoothness of *Daubechies* wavelet increases.

B. TEST AND SIMULATION

Three-phase signal noise of normal samples and fault samples is reduced, as shown from Figure 5 to Figure 10, respectively. In each set of figures, three figures of the first row are all data points of the signal, three figures of the second row are the first 10000 data points of the signal to observe the noise reduction details. Two figures of the first column are the original signal. Two figures of the second column are the signal through the high-pass filter. Two figures of the third column are the signal after noise reduction. The noise reduction is that keep the approximate and detailed coefficients amplitude when it is greater than the threshold value, and 0 when it is less than the threshold value. It can help to extract and classify the signal features.

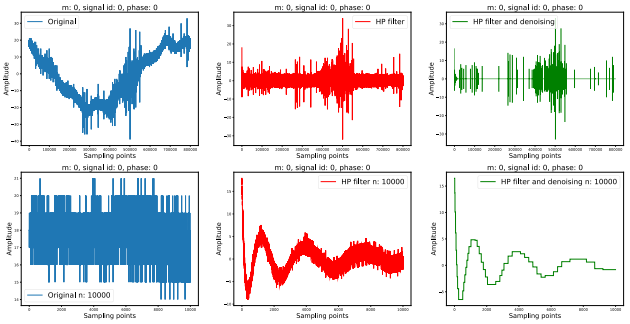


FIGURE 5. 0-phase signal denoise without fault (a) the original signal (the blue curve). (b) the signal after passing the high-pass filter (the red curve). (c) the signal after noise reduction (the green curve).

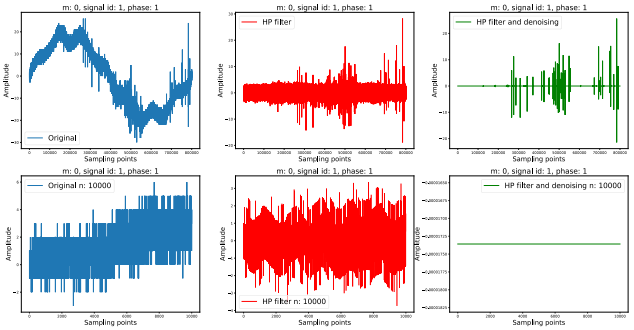


FIGURE 6. 1-phase signal noise reduction without fault (a) the original signal (the blue curve). (b) the signal after passing the high-pass filter (the red curve). (c) the signal after noise reduction (the green curve).

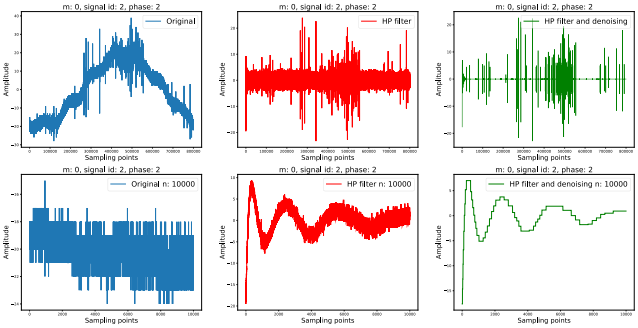


FIGURE 7. 2-phase signal noise reduction without fault (a) the original signal (the blue curve). (b) the signal after passing the high-pass filter (the red curve). (c) the signal after noise reduction (the green curve).

IV. FEATURE EXTRACTION BASED ON DWT

When PD activity occurs, each vibration signal contains unique information about the particular condition of IOC. It is called fault feature frequency signal. In practice, IOC tend to work with non-stationary power transmission, which results in additional aperiodic pulses. In this way, the traditional feature analysis method based on the envelopment is no longer applicable.

Therefore, DWT is chosen to decompose the original signal and obtain signal feature with different resolutions. The noise reduction signal is input and decomposed by DWT. The DWT calculation equation is shown in equation (1).

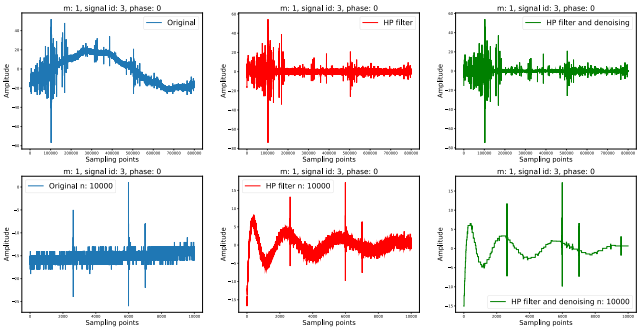


FIGURE 8. 0-phase signal noise reduction with fault (a) the original signal (the blue curve). (b) the signal after passing the high-pass filter (the red curve). (c) the signal after noise reduction (the green curve).

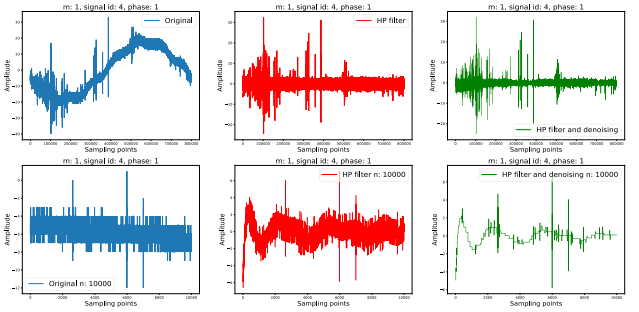


FIGURE 9. 1-phase signal noise reduction with fault (a) the original signal (the blue curve). (b) the signal after passing the high-pass filter (the red curve). (c) the signal after noise reduction (the green curve).

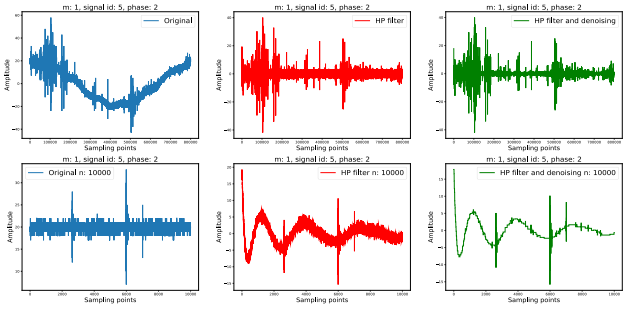


FIGURE 10. 2-phase signal noise reduction with fault (a) the original signal (the blue curve). (b) the signal after passing the high-pass filter (the red curve). (c) the signal after noise reduction (the green curve).

The DB4 wavelet is suitable for transient detection of electrical signal and used as the mother wavelet function. The DB4 function divides the signal into five detailed components (f_1, f_2, f_3, f_4, f_5) and an approximate component (A_1). Five detail components are represented as high frequency components and one approximate component is represented as low frequency components. The signal with fault and the signal without fault are decomposed using DWT, as shown in Figure 11.

V. IOC FAULT DETECTION BASED ON LSTM

LSTM is a sequence prediction model which predicts output through information embedded in a series of time steps.

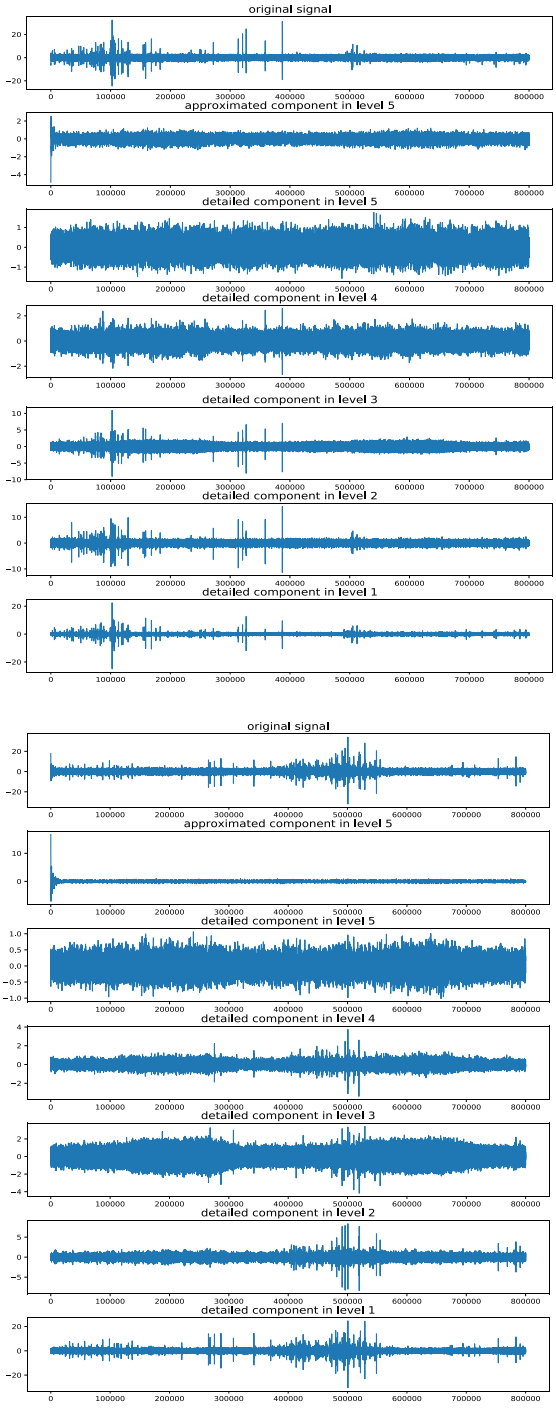


FIGURE 11. DWT decomposition (a) the results of the signal without fault (b) the results of the signal with fault.

In recent years, researchers have applied LSTM to some time series problems such as inventory, weather forecasts, and machine translation [32]–[34]. In these tasks, LSTM is usually superior to traditional machine learning models. Then it is used to detect the IOC fault.

A. RECURRENT NEURAL NETWORK

Since the LSTM neural network is based on enhanced recurrent neural network (RNN), the RNN is reviewed firstly.

The RNN is regarded as a group of feedforward neural network (FNN), where the hidden neurons of the previous time step are connected to the hidden neurons of the next time step. Hidden neurons H_t are obtained by combining the weight of the previous iteration cycle W_h with the weight of the current input information W_x. And so on, this process will continue to the next time iteration cycle. In this way, the RNN can take advantage of sequential information but not regard the signal as a combination of isolated points. The output of the current iteration period is not only based on the current input, but also based on the information of the previous iteration period. The structure of the many-to-one RNN model is shown in Figure 12.

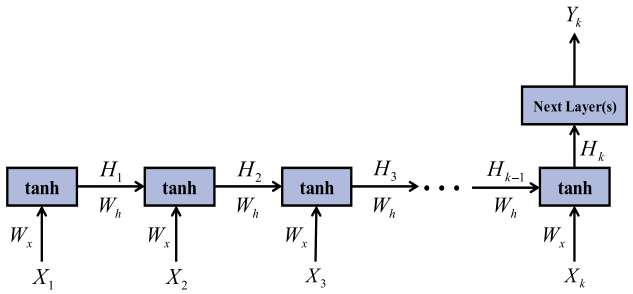


FIGURE 12. The structure of the many-to-one RNN model.

X is the input and H is the vector of hidden layer:

$$H_t = \tanh(W_h H_{t-1} + W_x X_t) \tag{11}$$

According to the chain rule, the network loss gradient is

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial H_k}\frac{\partial H_k}{\partial H_{k-1}}\cdots\frac{\partial H_2}{\partial H_1}\frac{\partial H_1}{\partial W} = \frac{\partial E_k}{\partial H_k}\left(\prod_{t=2}^{k}\frac{\partial H_t}{\partial H_{t-1}}\right)\frac{\partial H_1}{\partial W} \tag{12}$$

and derivative is

$$\begin{aligned}\frac{\partial H_t}{\partial H_{t-1}} &= \tanh'(W_h H_{t-1} + W_x X_t)\cdot\frac{d}{dH_{t-1}}[W_h H_{t-1} + W_x X_t] \\ &= \tanh'(W_h H_{t-1} + W_x X_t)\cdot W_h\end{aligned} \tag{13}$$

Combining (12) and (13),

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial H_k}\left(\prod_{t=2}^{k}\tanh'(W_h H_{t-1} + W_x X_t)\cdot W_h\right)\frac{\partial H_1}{\partial W} \tag{14}$$

According to W_h and $\tanh < 1$, $(\tanh'(W_h H_{t-1} + W_x X_t)\cdot W_h)$ may be less than 1 or greater than 1, which will cause the gradient to disappear or explode [35]–[36]. This will significantly affect the weight update and make it difficult to converge. LSTM uses complicated gate control instead of *tanh* activation function in the gradient flow, so it has better stability and performance.

B. LSTM ALGORITHM

Compared with the traditional RNN, LSTM introduces a specially designed unit that can accurately control the hidden state information flow from one time step to another [37]–[39]. The structure of the LSTM is shown in Figure 13.

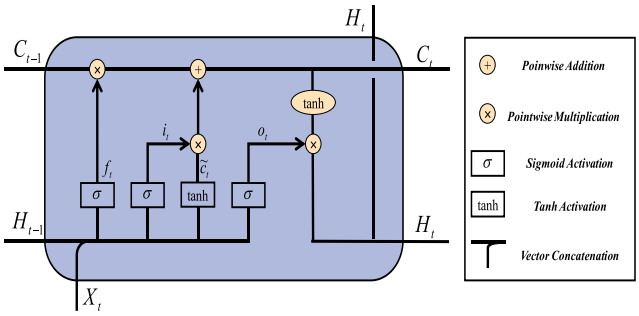


FIGURE 13. The structure of the LSTM.

In Figure 13, X_t and H_t are the input vector and the network hidden state vector at the time iteration period t, respectively. C_t is a vector which is stored in external memory unit. The interaction among the unit state vector, the input vector and the hidden state vector is accomplished through forgetting gate (f_t), input gate (i_t) and output gate (o_t).

The calculation of forgetting gate vector is:

$$f_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f) \tag{15}$$

where, $[H_{t-1}, X_t]$ is the concatenated vector of the previous hidden state vector H_{t-1} and the current input vector X_t, W_f and b_f are the weight and bias of f_t which are determined by network training, σ is the sigmoid activation function. The flow of information in the vector C_t is controlled by dot multiplication of elements. The temporary state vector $\tilde{C}_t$ is calculated by:

$$\tilde{C}_t = \tanh(W_c \cdot [H_{t-1}, X_t] + b_c) \tag{16}$$

where W_c and b_c are the weight and deviation of f_c, *tanh* is the *tanh* activation function.

The calculation of input gate vector is

$$i_t = \sigma(W_i \cdot (H_{t-1}, X_t) + b_i) \tag{17}$$

where W_i and b_i are the weight and deviation of i_t. They are determined by network training.

The state of new C_t in the time step t is updated,

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{18}$$

The current hidden state is determined by the new state and the write gate O_t. Similar to f_t and i_t, O_t can be written as

$$o_t = \sigma(W_o[H_{t-1}, X_t] + b_o) \tag{19}$$

The hidden state of current step H_t is calculated,

$$H_t = o_t * \tanh(C_t) \tag{20}$$

H_t is used to calculate the output of the current time step.

VI. EXPERIMENTS AND DISCUSSIONS

A. FAULT SCENARIOS

Insulated conductors can improve the stability of power transmission and reduce the construction space compared with traditional bare conductors. Therefore, insulated conductors

are used more and more in overhead power transmission. However, the following two faults may occur in IOC, which are difficult to detect:

(1) The standard protection devices used for bare conductor systems are often not able to detect the IOC's phase-to-ground fault. Because of the insulation cover, the phase-to-ground fault will not likely cause an overcurrent when IOC breaks and falls to the ground. This can create risky situation to people in close proximity to the fallen conductors.

(2) Something, such as tree branch, hitting the conductors will not be detected by the upstream protection devices. This can become a potential reliability threat in some locations where the tree branches etc. frequently hit the conductors due to wind or continuously push and bend the conductors. Eventually, the power line will be damaged, causing a power outage or starting a tree fire.

B. EVALUATION INDICATORS

Accuracy rate and error rate are two commonly used evaluation indexes to measure the classification model. However, they are not suitable for analyzing the imbalance dataset of ENET because using accuracy and error rates requires that each type of samples is equally important. In the ENET dataset, it makes more important to classify less samples type correctly than to classify more samples type correctly. So precision rate and recall rate are more suitable for ENET data analysis than accuracy rate and error rate. The less samples type is recorded as positive examples and The more samples type is recorded as negative examples. The labels of true results and forecast results are shown in Table 1.

TABLE 1. The labels of true results and forecast results.

True results	Forecast results	
	Positive	Negative
Positive	*TP*	*FN*
Negative	*FP*	*TN*

TP (true positive) indicates that the prediction result is positive and it is correct. *TN* (True Negative) indicates that the prediction result is negative and it is correct. *FP* (False Positive) indicates that the prediction result is positive and it is false. *FN* (False Negative) indicates that the prediction result is negative and it is false.

The precision rate (P) calculation formula is

$$P = TP/(TP + FP) \tag{21}$$

The recall (R) calculation formula is

$$R = TP/(TP + FN) \tag{22}$$

The F1-Score is the harmonic mean of Precision and Recall.

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{23}$$

Matthews Correlation Coefficient (MCC) is an index used to measure the performance of binary classification. MCC is a correlation coefficient to describe the actual classification and the predicted classification and defined as

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{24}$$

where $(TP + FN)(TN + FP)$ is a constant, which is defined as

$$TP + FN = PosLabelCount \tag{25}$$

$$TN + FP = NegLabelCount \tag{26}$$

$$\begin{aligned} \kappa &= (TP + FN)(TN + FP) \\ &= (PosLabelCount)(NegLabelCount) \end{aligned} \tag{27}$$

Define

$$\begin{aligned} \alpha &= (TP + FP)(TN + FN) \\ &= (Pos\,\mathrm{Pr}\,ediction Count)(Neg\,\mathrm{Pr}\,ediction Count) \end{aligned} \tag{28}$$

$$\begin{aligned} &FP * FN \\ &= (Pos\,\mathrm{Pr}\,ediction Count - TP) \\ &\quad * (Neg\,\mathrm{Pr}\,ediction Count - TN) \end{aligned} \tag{29}$$

$$\begin{aligned} &(TP * TN) - (FP * FN) \\ &= (TP * Neg\,\mathrm{Pr}\,ediction Count) \\ &\quad + (TN * Pos\,\mathrm{Pr}\,ediction Count) - \alpha \end{aligned} \tag{30}$$

Then according to the accuracy definition:

$$TP = Pos\_precision * Pos\,\mathrm{Pr}\,ediction Count \tag{31}$$

$$\begin{aligned} &TP * Neg\,\mathrm{Pr}\,ediction Count \\ &= Pos\_precision * Pos\,\mathrm{Pr}\,ediction Count \\ &\quad * Neg\,\mathrm{Pr}\,ediction Count \end{aligned} \tag{32}$$

Similarly:

$$TN = Neg\_precision * Neg\,\mathrm{Pr}\,ediction Count \tag{33}$$

$$TN * Pos\,\mathrm{Pr}\,ediction Count = Neg\_precision * \alpha \tag{34}$$

Then get:

$$\begin{aligned} &(TP * TN) - (FP * FN) \\ &= \alpha * Pos\_precision + \alpha * Neg\_precision - \alpha \end{aligned} \tag{35}$$

$$\begin{aligned} &(TP * TN) - (FP * FN) \\ &= \alpha * (Pos\_precision + Neg\_precision - 1) \end{aligned} \tag{36}$$

The expression of Matthew correlation coefficient is:

$$MCC = \sqrt{\frac{\alpha}{\kappa}} * (Pos\_precision + Neg\_precision - 1) \tag{37}$$

Defining $PosNegRatio = \sqrt{\frac{\alpha}{\kappa}}$,

$$MCC = \frac{1}{\sqrt{\kappa\alpha}} * (TP * TN - FP * FN) \tag{38}$$

In order to comprehensively evaluate the model performance, the precision rate, recall rate, F1-Score, and MCC are selected as the evaluation indexes.

TABLE 2. Many-to-one LSTM structure.

Layer(type)	Input Shape	Output Shape	Param
Lstm (LSTM)	(160,300)	(128)	219648
Dense-1(Dense)	(128)	(64)	8256
Dense-2(Dense)	(64)	(16)	1040
Dense-3(Dense)	(16)	(1)	17

TABLE 3. Signal feature layers.

Signal label	*Actual signal*
A_1	*Approximated component in level 5*
f_1	*detailed component in level 1*
f_2	*detailed component in level 2*
f_3	*detailed component in level 3*
f_4	*detailed component in level 4*
f_5	*detailed component in level 5*

TABLE 4. Experimental group settings.

Experimental group	*Experimental settings*
1	A_1
2	A_1 *and* f_1
3	A_1, f_1 *and* f_2
4	A_1, f_1, f_2 *and* f_3
5	A_1, f_1, f_2, f_3 *and* f_4
6	A_1, f_1, f_2, f_3, f_4 *and* f_5

C. FUSION LAYER SELECTION BASED ON F1-SCORE

The fusion of signal data from different layers are used for IOC fault detection. The Many-to-one LSTM structure is shown in Table 2. The features of each layer and their corresponding labels are shown in Table 3. There are six sets of comparative tests, as shown in Table 4. The signal data of different decomposition levels are obtained by DWT decomposition, and then they are integrated into the LSTM model in 3D format. For each decomposition level of data, randomly select 1000 groups of 3-phase signal as training data. There are 80000 data points for each phase signal. When the time step is set to 160, 80000/160=5000 data points are taken for each phase signal. In order to reduce the input vector dimension, the average value of 50 data points is taken as a new data point, then each phase signal consists of 5000/50=100 data points.

The F1-score of different fusion layers is shown as Figure 14. It can be seen that the LSTM classifier with four layers of features (A_1, f_2, f_3 and f_4) has obtained the best classification results. Therefore, it is selected as the input to the final classifier. In contrast, two layers of features cannot capture sufficient fine-grained feature changes. Five layers of features may amplify less meaningful feature changes and cause overfitting, then result in reducing classification performance.

D. THE EFFECT OF NOISE REDUCTION AND OVERSAMPLING

In the ENET data set, there are 525 PD signal samples and 8186 non-PD signal samples. The number of two types of

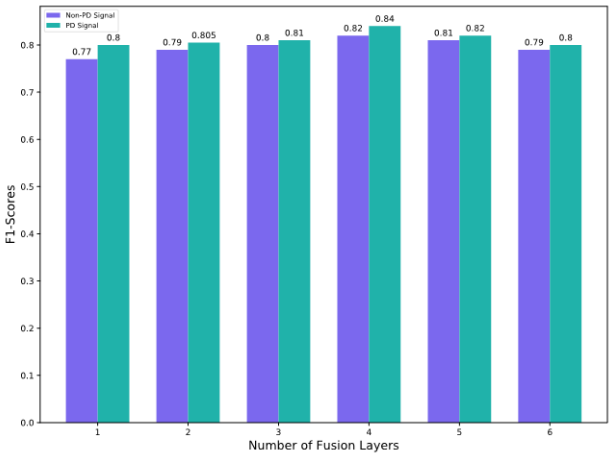


FIGURE 14. The F1-score of different fusion layers.

samples is shown in Figure 15. 0 represents normal samples and 1 represents fault samples. The data amount of normal samples is about 16 times that of fault samples. There is a serious imbalance between normal samples and fault samples.

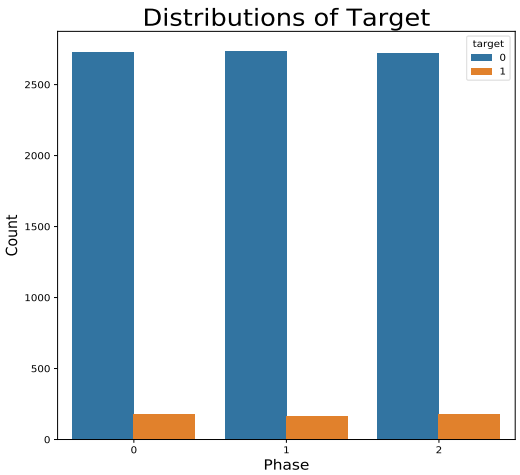


FIGURE 15. The number of normal samples and fault samples.

The problem of serious imbalance in data categories is solved by using Synthetic Minority Oversampling Technique (SMOTE) which is an improved scheme based on random oversampling algorithm. Because random oversampling adopts a simple copying strategy to increase minority samples, it is easy to generate model over simulation. The basic idea of SMOTE algorithm is to analyze a small number of samples and artificially synthesize new samples to add to the data set.

1) For each sample x in the minority class, calculate its distance to all the samples in the minority set using the Euclidean distance as the standard, and obtain its k nearest neighbors.
2) Determine the sampling ratio N according to the sample imbalance ratio. Randomly select several samples

from k nearest neighbors of minority class sample x. Assume that the selected neighbor is x_n.

3) Construct a new sample which is randomly selected neighbor x_n according to the following formula:

$$x_{new} = x + rand(0, 1) * |x - x_n| \tag{39}$$

In order to evaluate the effects of noise reduction and oversampling, the F1 score is used for the test. The test results are shown in Figure 16. It can be seen that both oversampling and noise reduction can improve the classification performance. If oversampling is not used, the classifier will focus on normal signals, thereby reducing the ability to identify fault signals. Wavelet noise reduction can remove the noise information contained in signals, which can reduce the difficulty of LSTM feature extraction and improve the classification accuracy.

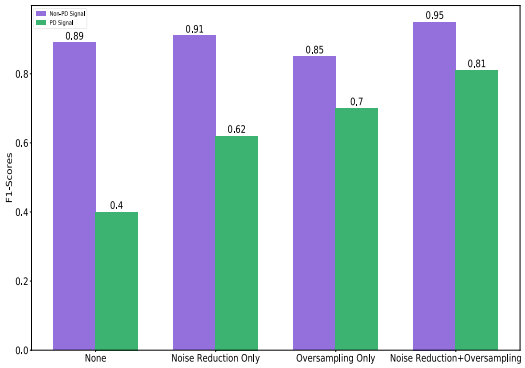


FIGURE 16. The results of noise reduction and oversampling.

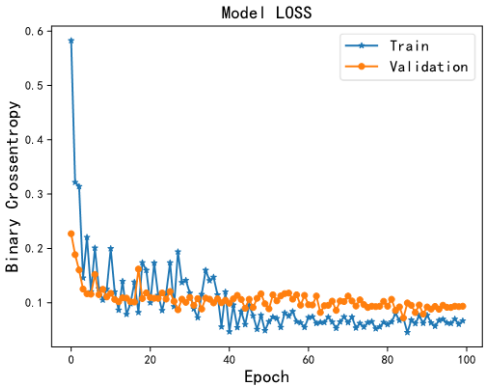


FIGURE 17. The loss function changes during model training.

E. THE COMPARISON OF DIFFERENT ALGORITHMS

The loss value of the proposed model during training is shown in Figure 17. It can be seen that the loss and accuracy tends to converge when the number of iterations is about 80. In the whole process, the loss value of the verification data set drops faster and fluctuates less than those of the training process, and eventually converges to 0.08, which may be caused by less number of verification data set (training data:Validation data = 7: 3). The accuracy of Matthew's correlation coefficient is shown in Figure 18. It can be seen that the accuracy

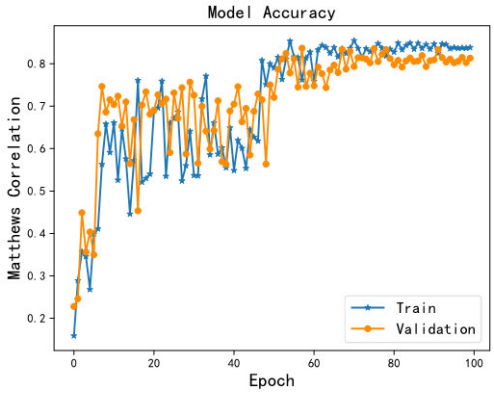


FIGURE 18. The changes in accuracy during model training.

tends to converge when the training iteration reaches about 80 times. The training accuracy is slightly higher than the verification accuracy, and reaches 0.86. The accuracy rates of different evaluation indexes are shown in Table 5.

TABLE 5. The model accuracy.

	Non-PD Signals	*PD Signals*	*Average*
Precision	0.81	0.81	0.81
Recall	0.81	0.81	0.81
F1-score	0.81	0.83	0.82
Matthews Correlation	0.84	0.86	0.85

Fuzzy neural network (FNN) combines the advantages of neural network system and fuzzy system, and it has great advantages in dealing with non-linearity and ambiguity. Support Vector Machine (SVM) is used to solve the problem of data classification and belongs to a kind of supervised learning algorithm [40], [41]. XGBoost is an open source machine learning project and has effectively implemented the GBDT algorithm. The MLR algorithm proposes and implements a non-linear relationship between learning features directly in the original space. The proposed model is compared with five classifiers (FNN, SVM, XGBoost, MLR, and LSTM), and the results are shown in Table 6. It can be seen that the accuracy of the proposed model is better than other models.

TABLE 6. The comparison between different algorithms.

	F1-Non-PD	*F1-PD*
FNN	66%±1.2%	72%±1.1%
SVM	71%±0.5%	72%±1.4%
XGBoost	70%±1.3%	65%±0.9%
MLR	62%±1.4%	53%±1.7%
LSTM	77%±2.3%	80%±2.9%
LSTM with DWT	84%±0.4%	86%±0.9%

VII. CONCLUSION

Aiming at the problem that the phase-to-ground fault and phase-to-phase fault of IOC are difficult to detect, a method based on combination of DWT and LSTM is proposed through detecting partial discharge in this paper. And we get the following conclusions.

(1) DWT can effectively reduce the noise of original voltage signal, so that the signal features are more obvious.
(2) DWT is chosen to decompose the noise reduction signal and obtain signal features of different resolutions. The DB4 wavelet is suitable for transient detection of electrical signal and used as the mother wavelet function.
(3) In some time series problems, LSTM is usually superior to traditional machine learning methods and can enhance the PD recognition performance.
(4) Different algorithms are used to make a comparative test on the ENET data-set. The results show that the DWT-LSTM method provides the best classification results measured by F1-Score. The proposed method based on DWT-LSTM and partial discharge is suitable for IOC fault detection.

In the future, we will use two methods to further improve the accuracy of fault detection. The first method will use bidirectional LSTM to extract the features of fault samples and normal samples. The second method will use small sample learning, migration learning and other means to solve the problem of rare fault samples.

REFERENCES

[1] G. Hashmi, M. Lehtonen, and M. Nordman, "Modeling and experimental verification of on-line PD detection in MV covered-conductor overhead networks," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 17, no. 1, pp. 167–180, Feb. 2010.
[2] P. Pakonen, "Detection of incipient tree faults on high voltage covered conductor lines," Tampere Univ. Technol., Tampere, Finland, Tech. Rep., Nov. 2007.
[3] S. Dabbak, H. Illias, A. B. Chin, and M. A. Tunio, "Surface discharge characteristics on HDPE, LDPE and PP," *Appl. Mech. Mater.*, vol. 785, pp. 383–387, Aug. 2015.
[4] M. Dong, Z. Sun, and C. Wang, "A pattern recognition method for partial discharge detection on insulated overhead conductors," in *Proc. IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, May 2019, pp. 1–4.
[5] H. K. Agarwal, K. Mukherjee, and P. Barna, "Partially and fully insulated conductor systems for low and medium voltage over head distribution lines," in *Proc. IEEE 1st Int. Conf. Condition Assessment Techn. Electr. Syst. (CATCON)*, Dec. 2013, pp. 100–104.
[6] W. Zhang, Z. Hou, H.-J. Li, C. Liu, and N. Ma, "An improved technique for online PD detection on covered conductor lines," *IEEE Trans. Power Del.*, vol. 29, no. 2, pp. 972–973, Apr. 2014.
[7] J. Zhang, H. Bian, H. Zhao, X. Wang, L. Zhang, and Y. Bai, "Bayesian network-based risk assessment of single-phase grounding accidents of power transmission lines," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 1841, 2020.
[8] S. Mišák and V. Pokorný, "Testing of a covered conductor's fault detectors," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1096–1103, Jun. 2015.
[9] S. Hamacek and S. Misak, "Detector of covered conductor faults," *Adv. Electr. Electron. Eng.*, vol. 10, no. 1, pp. 7–12, 2012.
[10] J. I. N. Jun, "Noise reduction and source recognition of partial discharge signals in gas-insulated substation," Ph.D. dissertation, ScholarBank@NUS Repository, 2006. [Online]. Available: https://scholarbank.nus.edu.sg/handle/10635/15365
[11] Q. Zhang, J. Lin, H. Song, and G. Sheng, "Fault identification based on PD ultrasonic signal using RNN, DNN and CNN," in *Proc. Condition Monitor. Diagnosis (CMD)*, Perth, WA, Australia, Sep. 2018, pp. 1–6.
[12] K. Banno, Y. Nakamura, Y. Fujii, and T. Takano, "Partial discharge source classification for switchgears with transient earth voltage sensor using convolutional neural network," in *Proc. Condition Monitor. Diagnosis (CMD)*, Perth, WA, Australia, Sep. 2018, pp. 1–5.
[13] H. Song, J. Dai, W. Zhang, K. Bi, L. Luo, G. Sheng, and X. Jiang, "Partial discharge pattern recognition based on deep convolutional neural network under complex data sources," *High Voltage Eng.*, vol. 44, no. 11, pp. 3625–3633, 2018.
[14] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019.
[15] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application," *ISA Trans.*, vol. 97, pp. 269–281, Feb. 2020.
[16] G. Li, M. Rong, X. Wang, X. Li, and Y. Li, "Partial discharge patterns recognition with deep convolutional neural networks," in *Proc. Int. Conf. Condition Monit. Diagnosis (CMD)*, Xi'an, China, Sep. 2016, pp. 324–327.
[17] B. Liu and J. Zheng, "Partial discharge pattern recognition in power transformers based on convolutional neural networks," *High Voltage App.*, vol. 53, no. 5, pp. 70–74, 2017.
[18] X. Wan, H. Song, L. Luo, Z. Li, G. Sheng, and X. Jiang, "Pattern recognition of partial discharge image based on one-dimensional convolutional neural network," in *Proc. Condition Monitor. Diagnosis (CMD)*, Perth, WA, Australia, Sep. 2018, pp. 1–4.
[19] X. Wan, H. Song, L. Luo, Z. Li, G. Sheng, and X. Jiang, "Application of convolutional neural networks in pattern recognition of partial discharge image," *Power Syst. Technol.*, vol. 43, pp. 2219–2226, Jun. 2019.
[20] F. Yang, G. Wang, X. Peng, J. Wen, Q. Chen, G. Yang, and C. Li, "Partial discharge pattern recognition of high-voltage cables based on convolutional neural network," *Electr. Power Automat. Equip.*, vol. 38, no. 5, pp. 123–128, 2018.
[21] M.-T. Nguyen, V.-H. Nguyen, S.-J. Yun, and Y.-H. Kim, "Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear," *Energies*, vol. 11, no. 5, p. 1202, 1202.
[22] B. Adam and S. Tenbohlen, "Classification of multiple PD sources by signal features and LSTM networks," in *Proc. IEEE Int. Conf. High Voltage Eng. Appl. (ICHVE)*, Athens, Greece, Sep. 2018, pp. 1–4.
[23] S. C. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelets Transforms*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[24] C. Lin, W. Gao, and M.-F. Guo, "Discrete wavelet transform-based triggering method for single-phase earth fault in power distribution systems," *IEEE Trans. Power Del.*, vol. 34, no. 5, pp. 2058–2068, Oct. 2019.
[25] M. Kordestani and M. Saif, "Data fusion for fault diagnosis in smart grid power systems," in *Proc. IEEE 30th Can. Conf. Electr. Comput. Eng. (CCECE)*, Apr. 2017, pp. 1–6.
[26] P. D. Agoris, S. Meijer, E. Gulski, and J. J. Smit, "Threshold selection for wavelet denoising of partial discharge data," in *Proc. IEEE Int. Symp. Electr. Insul.*, Sep. 2004, pp. 62–65.
[27] G. Suganya, S. Jayalalitha, K. Kannan, and S. Venkatesh, "Survey of de-noising techniques for partial discharge interferences," vol. 12, pp. 414–427, 2017.
[28] D. L. Donoho and I. M. Johnstone, "Threshold selection for wavelet shrinkage of noisy data," in *Proc. 16th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1, Nov. 1994, pp. A24–A25.
[29] R. Sun and C. L. Giles, "Sequence learning: From recognition and prediction to sequential decision making," *IEEE Intell. Syst.*, vol. 16, no. 4, pp. 67–70, Jul./Aug. 2001.
[30] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models," *Expert Syst. Appl.*, vol. 103, pp. 25–37, Aug. 2018.
[31] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018.
[32] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, and S. Jain, "Machine translation using deep learning: An overview," in *Proc. Int. Conf. Comput., Commun. Electron. (Comptelix)*, Jul. 2017, pp. 162–167.
[33] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[34] J. F. Kolen and S. C. Kremer, "Gradient flow in recurrent nets: The difficulty of learning long-term dependencies," in *A Field Guide to Dynamical Recurrent Networks*. Piscataway, NJ, USA: IEEE Press, 2001, pp. 237–243, doi: 10.1109/9780470544037.ch14.
[35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[36] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.
[37] J. Dong, Y. Wu, and G.-H. Yang, "A new sensor fault isolation method for T–S fuzzy systems," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2437–2447, Sep. 2017.

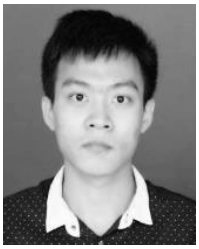
[38] A. H. Etemadi and M. Sanaye-Pasand, "High-impedance fault detection using multi-resolution signal decomposition and adaptive neural fuzzy inference system," *IET Gener., Transmiss. Distrib.*, vol. 2, no. 1, pp. 110–118, 2008.
[39] M. Kordestani, M. Saif, M. E. Orchard, R. Razavi-Far, and K. Khorasani, "Failure prognosis and applications—A survey of recent literature," *IEEE Trans. Rel.*, early access, Sep. 17, 2019, doi: 10.1109/TR.2019.2930195.
[40] J. Zuo, J. Chen, Y. Tan, M. Wang, and L. Wen, "A multi-agent collaborative work planning strategy based on AFSA-PSO algorithm," in *Proc. Int. Conf. Robots Intell. Syst. (ICRIS)*, Haikou, China, Jun. 2019, pp. 254–257.
[41] N. Qu, J. Zuo, J. Chen, and Z. Li, "Series arc fault detection of indoor power distribution system based on LVQ-NN and PSO-SVM," *IEEE Access*, vol. 7, pp. 184020–184028, 2019.

NA QU received the M.S. degree in power electronics and transmission from Northeastern University, Shenyang, China, in 2005, where she is currently pursuing the Ph.D. degree.

She is also an Associate Professor with Shenyang Aerospace University. She has published more than 20 articles in major journals and international conferences. Her research was supported by the National Natural Science Foundation of China. Her current research interests include arc fault, electrical fire detection, and intelligent information processing.

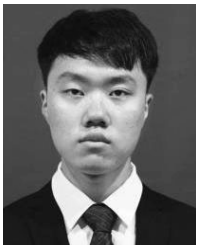
ZHONGZHI LI is currently pursuing the B.S. degree in computer science and technology with Shenyang Aerospace University, Shenyang, China.

He has published more than five articles in journals and international conferences. His research interests include machine learning, deep learning, computer vision, few shot learning, and domain adaptation.



JIANKAI ZUO is currently pursuing the B.S. degree in network engineering with the School of Computer Science, Shenyang Aerospace University, Shenyang, China, and the B.S. degree in business administration with Northeastern University. Due to his excellent academic performance and strong scientific research ability, he was promoted directly to pursue the Ph.D. degree in computer science and technology with Tongji University, Shanghai.

He has published more than ten articles in journals and international conferences. His research interests include machine learning, deep learning, pattern recognition, and computer vision.



JIATONG CHEN is currently pursuing the B.S. degree in energy and power engineering with Shenyang Aerospace University, Shenyang, China.

He has published more than five articles in journals and international conferences. His research interests include intelligent information processing, deep learning, and data mining.