An End-to-end Transient Recognition Method for VSC-HVDC Based on Deep Belief Network

Guomin Luo, Jiaxin Hei, Changyuan Yao, Jinghan He, and Meng Li

***Abstract*—Lightning is one of the most common transient interferences on overhead transmission lines of high-voltage direct current (HVDC) systems. Accurate and effective recognition of faults and disturbances caused by lightning strokes is crucial in transient protections such as traveling wave protection. Traditional recognition methods which adopt feature extraction and classification models rely heavily on the performance of signal processing and practical operation experiences. Misjudgments occur due to the poor generalization performance of recognition models. To improve the recognition rates and reliability of transient protection, this paper proposes a transient recognition method based on the deep belief network. The normalized line-mode components of transient currents on HVDC transmission lines are analyzed by a deep belief network which is properly designed. The feature learning process of the deep belief network can discover the inherent characteristics and improve recognition accuracy. Simulations are carried out to verify the effectiveness of the proposed method. Results demonstrate that the proposed method performs well in various scenarios and shows higher potential in practical applications than traditional machine learning based ones.**

***Index Terms*—Deep belief network, transient recognition, machine learning, voltage source converter based high-voltage direct current (VSC-HVDC).**

I. Introduction

With many advantages such as flexible and controllable operation modes, friendly connection to renewable energy, and convenient construction of multi-terminal or meshed grids, voltage source converter based high-voltage direct current (VSC-HVDC) has been developing rapidly and plays a more and more important role in long-distance transmission [1], [2]. Since the post-fault measurements change quickly, and the faults must be recognized in only a few microseconds, transient protection should be used for VSC-HVDC transmission lines. As a highly erected overhead conductor in wild areas, the transmission line is at high risk of being attacked by lightning strokes. The recognition between fault transients and disturbance transients is critical.

Transient recognition is always an important function of transient protection for both traditional alternative current (AC) and direct current (DC) transmission lines. For AC transmission lines, the sinusoidal sources produce zero-crossing points, where the transient features of grounding faults (GFs) are unobvious; while for DC transmission lines, no zero-crossing points exist [3]. The transient features of GFs on DC lines are clearer than those on AC ones. However, with more unique features, the transient recognition of DC transmission lines is not an easy job since only extremely short data segments can be used. It is hard to balance the speed and the precision of discriminations of faults and disturbances. How to recognize transients quickly and reliably is the primary consideration for VSC-HVDC transient protection [4]-[6].

A lot of recognition methods have been proposed to discriminate faults and disturbances. But no matter which kind of tools are used, the traditional recognition methods must include two crucial technologies: feature extraction and classification. Feature extraction aims to represent the unique characteristics of measurements with lower dimension. Some time-domain features such as principal component analysis can extract different patterns from 3 ms transient data to distinguish shielding failure from back-flashover [4]. Signal processing tools such as Fourier transform and wavelet transform [7]-[10] are generally used to reveal signal differences in the frequency domain or time-frequency domain. Those differences are generally measured by simple processing methods such as amplitude-based ones or change rate-based ones, statistic concepts such as wavelet energy, wavelet entropy, spectrum distribution, and so on [11]-[13]. With those features, classification is performed. Two main types of classification are based on thresholding and machine learning. The thresholding-based classification is mostly composed of a series of logical relationships [14]-[17]. But this linear classification method requires complex selection and calculation of thresholds and logic procedures. Machine learning based classification provides a non-linear solution. Back-propagation (BP) neural network, support vector machine (SVM), and other intelligent algorithms [3], [18], [19] are often used for fitting the classification boundary. Compared with thresholding-based classification, machine learning based classifica-

Manuscript received: March 25, 2020; accepted: October 12, 2020. Date of CrossCheck: October 12, 2020. Date of online publication: November 26, 2020.

This work was supported in part by the National Key R&D Program of China (2018YFB0904600), the National Natural Science Foundation of China (No. 51507008), and the State Grid Corporation technology project (No. 5200-201956113A-0-0-00).

G. Luo (corresponding author), J. Hei, J. He, and M. Li are with the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China (e-mail: guominluo@hotmail.com; 18126094@bjtu.edu.cn; jhhe@bjtu.edu.cn; mengl@bjtu.edu.cn).

C. Yao is with the State Grid Shandong Electric Power Company, Liaocheng, China (email: 16121566@bjtu.edu.cn).

DOI: 10.35833/MPCE.2020.000190

tion can generate better performance. However, the performance of both classifications depends heavily on the representation ability of features. If the feature is unstable in various scenarios, for example, changing with fault parameters, these classification methods might produce misjudgments.

With the development of unsupervised learning techniques, the considerable training burden of machine learning models can be subdivided and done layer by layer. The depth of models can thus be easily increased to complete more complex works. The networks with unsupervised learning techniques, for instance, autoencoder and Boltzmann machine, can reduce the dimension of input data, which provides a kind of representation, or characteristics, of original raw data. This kind of learned features is extracted with data-based methods which are totally different from the traditional human-intervened ones. The learned features do not depend on the selection of characterization tools and the setting of parameters and have great potential in extracting more stable representation of original data [20]. Deep belief network (DBN), which was proposed in 2006, is a probabilistic model that provides a joint probability distribution of data and labels [21]. It is capable of unsupervised feature extraction from complex data, e.g., speech and image recognition [22]. Its powerful characterizing ability can also be explored in the area of electrical measurement analysis.

This paper analyzes the characteristics of three common transients on VSC-HVDC transmission lines, including GF, lightning fault (LF) and lightning disturbance (LD), and proposes a DBN based transient recognition method. The main contributions are as follows:

1) An end-to-end recognition method is proposed for distinguishing fault transients and non-fault lightning transients. The hand-designed components such as feature extraction and classifier selection are not needed. In this paper, the normalized raw data (line-mode current) is used as the inputs of DBN, and the output is the transient type.

2) The transients from different but similar transmission lines are combined to enlarge the datasets and improve the generalization of DBN. Small-size sample set is a tough issue when machine-learning models are trained for power systems. The combination of samples with similar features can help to increase the size of training samples and improve the performance of trained models.

3) The recognition rate of VSC-HVDC transients has been improved to be higher than 96% with 3 ms data length and 10 kHz sampling frequency. Both speed and reliability are required for VSC-HVDC transient recognitions. The existing engineering methods and machine learning based ones cannot fulfill these two requirements at the same time.

The rest of this paper is organized as follows. Section II introduces the fundamentals of HVDC and the types of transients that should be recognized in transient protection. Section III presents the theory of restricted Boltzmann machine (RBM) and its training procedure. The recognition method is given in Section IV, and simulations are performed in Section V. The simulation results show that the proposed method is effective in different scenarios. The comparisons in Section VI and VII prove the excellent generalization performance and stable recognition capability of the proposed method. Finally, conclusions are drawn in Section VIII.

II. Transients on Overhead HVDC

A. Fundamental of HVDC

The typical structure of a point-to-point VSC-HVDC transmission system is shown in Fig. 1. The system voltage is ±500 kV, the transmission power is 2000 MW, and the total length of the line is 500 km. The middle point of the supporting capacitor is grounded, and the line voltages of both poles are symmetrical. This grounding method could reduce the insulation level of DC lines, and prevent large current flowing through the grounding point under normal conditions [23], [24]. The bus voltage is controlled at the rectifier, and the power is controlled at the inverter. In Fig. 1, L represents the smoothing reactors, M represents the measuring devices for relay protection, and *f* represents the GF.

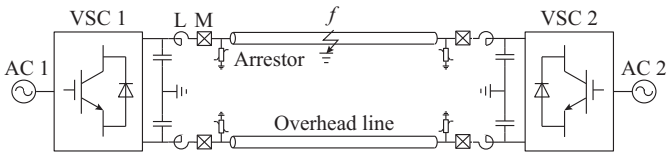

Fig. 1. Structure of VSC-HVDC transmission system.

Once a fault occurs on the transmission line, two transient pulses with opposite polarity travel to both ends of the line. With the smoothing reactors L, the transients will be blocked in the area of transmission lines. Only the transients on the transmission lines should be considered for recognition.

B. GF

GF is the most common fault for overhead HVDC transmission lines. For symmetrical single-polar structure, the post-fault transient process is swift, and can be divided into two stages: capacitor discharging stage and post-fault steady stage [25], [26]. Once a fault occurs, the traveling wave will propagate along the transmission lines, and reach the converter station where supporting capacitors are installed. The supporting capacitor will discharge immediately, and lead to potential drop and current rise of fault pole. However, due to the voltage control of the converter station, the voltage between positive and negative poles keeps unchanged, and the potential of the non-fault pole will increase. This process is a step response, and another steady state will be achieved when the potential of the fault pole reaches its final value. In the post-fault steady state, the potential of the fault pole is almost zero, and the potential of the non-fault pole is close to DC-bus voltage. At the same time, the fault current will decrease due to the potential decrease of the fault pole. A typical fault current waveform of single-pole GF is demonstrated in Fig. 2(a). Its current increment is quite large and reaches its peak in only a few microseconds. It then decays gradually to its steady value.

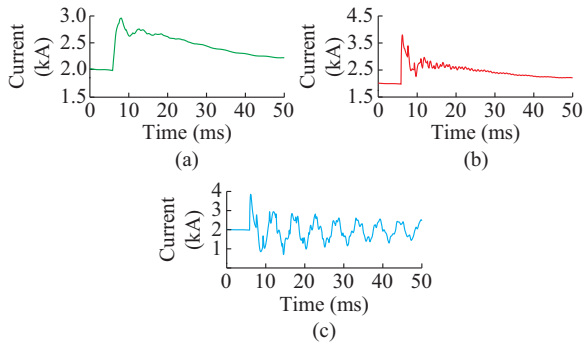

Fig. 2. Transient current waveforms on positive pole. (a) GF. (b) LF. (c) LD.

C. LF

Overhead HVDC transmission lines usually pass through rural, mountainous, or empty wild areas. With a height of tens of meters, the transmission lines are vulnerable to lightning. A lightning stroke is usually regarded as a negative pulse, and also a kind of high-frequency interference source [27], [28]. It can be modeled mathematically by a double exponential current source, as given by (1) [29].

$$i(t)=AI\left(\mathrm{e}^{-\alpha t}-\mathrm{e}^{-\beta t}\right) \tag{1}$$

where I is the lightning current amplitude; A is the correction coefficient of I; and α and β are the rise and fall coefficients, respectively.

When lightning with a number of electrons strikes the wire or the tower top of the HVDC transmission system, insulator flashover might be caused and produce a unipolar GF. Such a phenomenon is called an LF. However, if the lightning amplitude is not large enough, an LD is more likely to be generated. The LD transient pulse will propagate to both ends, and the wave energy will decay to zero due to reflection and refraction [30].

Figure 2(b) and Fig. 2(c) display the typical waveforms of LF and LD in point-to-point VSC-HVDC. The waveforms of LF and GF decay faster than those of LD because most transient energy dissipates quickly to ground through fault point. In the early stage of a lightning flash, the waveforms of LF and LD are similar. However, once the insulation breakdown and GF occur, the waveform of LF becomes more similar to that of GF. These similarities among the three types of signals make accurate recognition difficult, especially for LF and LD.

III. Recognition Model

Deep learning can learn features from raw data and has achieved breakthroughs in pattern recognition in recent years. DBN, one of the popular deep learning networks, consists of multiple RBMs. RBM works well in learning features. A Softmax classification layer is commonly added on the RBMs to form a shallow network, which could perform feature extraction and classification.

A. RBM

An RBM is composed of a visible layer $\boldsymbol{v}$ and a hidden layer $\boldsymbol{h}$. It consists of a weight matrix $\boldsymbol{w}$ that associates with the connection between all units in $\boldsymbol{v}$ and $\boldsymbol{h}$, as well as bias weights $\boldsymbol{a}$ and $\boldsymbol{b}$, which represent the offsets of $\boldsymbol{h}$ and $\boldsymbol{v}$, respectively. Initially, the layer comprises binary-valued random units, i.e., $\boldsymbol{v}\in\{0,1\}^V$, $\boldsymbol{h}\in\{0,1\}^H$, where V is the number of visible layer units and H is the number of hidden layer units. To process real-valued data or measured electrical transient in HVDC transmissions, the visible layer units should be replaced by linear units with independent Gaussian noise, which is called Gaussian-Bernoulli RBM (GB-RBM) [31], [32]. The structure of GB-RBM is shown in Fig. 3.

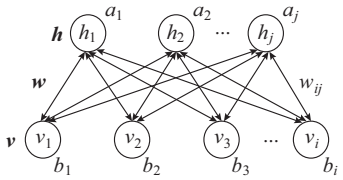

Fig. 3. Structure of GB-RBM.

B. Recognition Model

The goal of GB-RBM is to extract features from original data. After the proper training of the network, the output data (values of hidden units) are a characteristic representation of the input data (values of visible units). The output of the trained GB-RBM can serve as the input of the next GB-RBM. With multiple GB-RBMs which process data sequentially, complex recognition problems can be solved. The network depth depends on the needs of problems and is one of the critical factors that affect the classification accuracy.

Generally, a classification output layer needs to be added behind the last feature layer to realize recognition or classification. This layer, which is essentially a neural network, can perform classification through a transfer function. Logistic function and Softmax function are two common transfer functions used for classification. For multi-classification problems, a Softmax classifier which adopts an exponential function can amplify the differences to produce better performance than logistic classifier. Softmax classifier is thus adopted.

Stacked GB-RBMs (for instance, three layers) with a Softmax classifier forms a DBN. The training process of the typical DBN which includes pre-training and fine-tuning phases is displayed in Fig. 4.

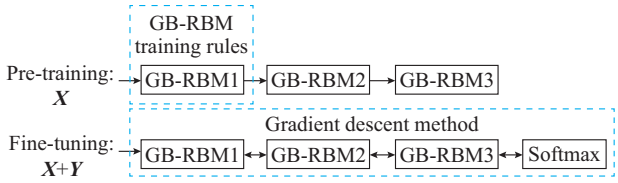

Fig. 4. Training method of DBN.

Pre-training: the training is performed layer by layer according to the training rules of GB-RBM. The sample data $\boldsymbol{X}$ is used as the input of GB-RBM1, and the output of the trained GB-RBM1 is used as the input of GB-RBM2. The training of the following GB-RBMs is sequentially performed in the same way. The process is called unsupervised

pre-training due to the absence of sample label ***Y***.

Fine-tuning: all the pre-trained GB-RBMs are stacked up, and the Softmax layer is then added at the end of the network. Pre-training has specified a rough optimization direction of the entire network when fine-tuning can complete more accurate optimization. The backward propagation algorithm is used to fine-tune the whole network for the purpose of reducing the error between the output and the label. This phase requires label ***Y***, so it is called supervised fine-tuning.

IV. Recognition Method

With a properly designed DBN, the transient measurements from HVDC transmission lines are analyzed to achieve reliable end-to-end recognition. The flow chart of the proposed method is shown in Fig. 5, and the details of main steps are as follows.

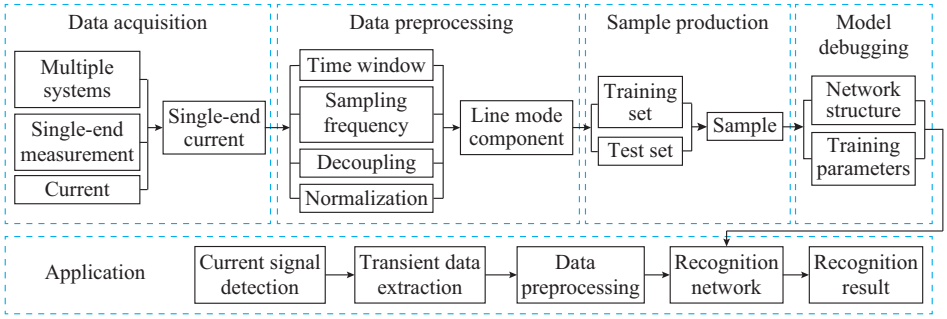

Fig. 5. Flow chart of transient recognition method.

A. Data Acquisition

In the first millisecond of the post-fault process of VSC-HVDC, currents that change quickly and are less likely to be affected by control strategies are chosen to be the analyzed input of the recognition model. To avoid the effect of communications and to reduce the data acquiring time, only the local current measurements are used.

The biggest challenge of using intelligent algorithms in transient recognition is the lack of samples. Transient current samples from one system are limited in practical applications. But for systems with the same voltage level and interconnections, their transient currents appear similar features when they respond to the same kind of excitation, for example, step excitations or double exponential impulses.

Thus, the transient currents from different but similar transmission lines are used in this paper to enlarge the size of samples and enhance the generalization performance of the model.

B. Data Preprocessing

To avoid the influences from measurements such as magnitude and coupling and help to achieve fast convergence of recognition model, the raw data should be processed before being sent to the model. Three aspects need to be considered: data length, decoupling, and normalization.

The data length is determined by the sampling frequency and the length of the time window. The data length should be the same for different systems. For example, if the sampling frequency is 10 kHz and the time window is 3 ms, the data length is 30.

Karenbauer transform, which is shown in (2) and (3), is adopted to decouple the influences between two poles. The original transient currents are decoupled into two components: zero-mode i_0 and line-mode i_1. Compared with i_0, i_1 is more stable with the change of frequency and line surroundings. Hence, the line-mode i_1 is adopted.

$$\begin{bmatrix} i_1 \\ i_0 \end{bmatrix} = \boldsymbol{S}^{-1} \begin{bmatrix} i_+ \\ i_- \end{bmatrix} \tag{2}$$

$$\boldsymbol{S} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{3}$$

where i_+ and i_- are the currents of the positive and negative poles, respectively.

Normalization is generally used in processing the input data of the recognition network to improve the convergence speed and generalization performance. Here, the min-max normalization shown in (4) is selected.

$$i_1^* = \frac{i_1 - \min\{i_1\}}{\max\{i_1\} - \min\{i_1\}} \tag{4}$$

where i_1^* is the normalized i_1.

C. Sample Preparation

Generally, the samples used for the recognition model are divided into the training set and the test set to avoid over-fitting and evaluate the training results.

D. Model Design and Training

The design of a network includes the selection of network size, functions, and parameters such as learning rate, and so on. When considering the network size, the sample dimension and transient types are used as the size of the input layer and output layer, respectively. The depth and the size of hidden layers depend on the samples and applications. In this paper, the hidden layer structure is determined by increasing the number of GB-RBMs gradually, and the structure which achieves maximum recognition rate is selected.

E. Model Application

Once the recognition model is trained, it can be used to recognize unknown abnormal transients with great speed. Although the model training is time-consuming, the implement

of trained model only takes a few milliseconds, which is potential to be applied in practical applications.

V. Simulations and Analysis

A. Simulation Model and Transient Data Processing

As mentioned in Section IV, a ±500 kV point-to-point VSC-HVDC transmission system with four scenarios is modeled on the platform of PSCAD. The structure of the point-to-point transmission system is shown in Fig. 1, and their detailed parameters are listed in Table I. The transmission capacity, transmission distance, and wire are different in the four scenarios, but voltage and control strategies are the same.

TABLE I
Detailed Parameters of Four Scenarios

Scenario	Power (MW)	Distance (km)	Transmission line
1	1500	250	4×LGJ-400
2	2000	500	4×LGJ-500
3	2500	750	4×LGJ-630
4	3000	1000	4×LGJ-720

The three types of transients mentioned in Section II are also simulated. For GFs which are modeled by step voltage functions, only grounding resistances are considered. While for lightning strokes which are modeled by double exponential functions, both time constants and amplitudes are used to simulate different lightning strokes. The location of transients changes arbitrarily in the simulations. A GF is applied 0.1 ms later than lightning strokes when simulating LF transients. Table II lists the setting ranges of parameters of transient signals, where the time constant 1.2(50) represents that the rise time is 1.2 μs and the fall time is 50 μs.

TABLE II
Parameters of Transient Signals

Type	Location	Grounding resistance (Ω)	Time constant (μs)	Lightning current amplitude (kA)
GF	Random	0.01-100		
LF	Random	0.01-30	1.2(50), 2.6(50), 8(20)	25-80
LD	Random		1.2(50), 2.6(50), 8(20)	10-30

With the simulation setting in Tables I and II, three types of transients are generated. The time window is set to be 3 ms, and the sampling frequency is 10 kHz. Once a transient surge is detected, the measurements 0.2 ms before and 2.8 ms after the detected instant are recorded. Figure 6(a), (c), and (e) shows the line-mode components of three types of transients, while Fig. 6(b), (d), and (f) shows their normalized waveforms. It can be seen from Fig. 6 that the system parameters affect the attenuation speed and the reference value of the waveforms. However, the changing trend remains roughly consistent for the same type of signals.

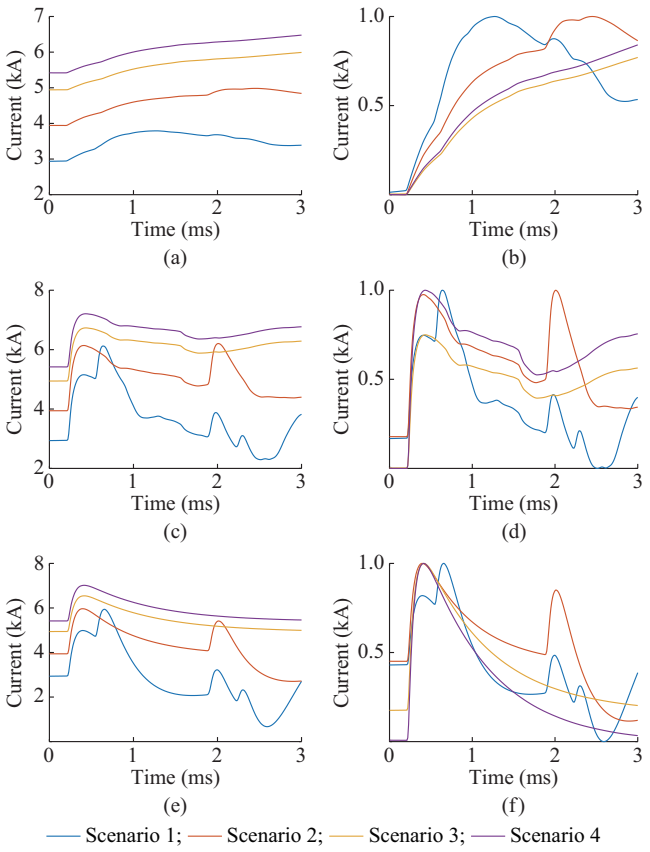

Fig. 6. Waveforms of different transients. (a) Line-mode raw data of GF. (b) Normalized waveforms of GF. (c) Line-mode raw data of LF. (d) Normalized waveforms of LF. (e) Line-mode raw data of LD. (f) Normalized waveform of LD.

B. Sample Preparation

For each simulation model, different sizes of sample sets are built. The longer the transmission line is, the more simulation scenarios will be produced. To analyze the reliability of trained model, the number of test samples is raised. Each sample set is randomly divided into the training set and the test set with equal numbers. Table III lists the detailed sample distributions. For each kind of transients, 200 samples are produced. Thus, 600 samples are used in total.

TABLE III
Sample Distributions

Scenario	No. of samples	No. of training samples	No. of test samples
1	20	10	10
2	40	20	20
3	60	30	30
4	80	40	40
Total	200	100	100

C. Model Design and Training

According to the design and training procedure in Section IV, the recognition model is built and trained. In pre-training and fine-tuning, the minimum-batch size is set to be the number of classes 3, the learning rate ε is set to be 0.01, and the momentum factor mc is set to be 0.5. The number of iter-

ations, which is determined by the convergence curve of the error, is set to be 500 and 300 for pre-training and fine-tuning, respectively. Such training is repeated 10 times to calculate the average misjudgment rate.

The number of neurons cannot be too small or too large, or else the features cannot be adequately represented or the learning process will cost much more time. Figure 7 illustrates the average misjudgment rates with different numbers of neurons. Here, only one hidden layer is used. After 200 iterations or epoches, the average misjudgment rate reaches convergence. For both training sets and test sets, the changing trends of the rate keep the same: the rate decreases with the increase of the number of neurons and rises slightly after reaching its minimum. In this research, the hidden layer with 30 neurons ($h_1 = 30$) always has smaller values than the others: 2.57% for training sets and 3.60% for test sets. Thus, 30 is chosen as the number of neurons in the first hidden layer.

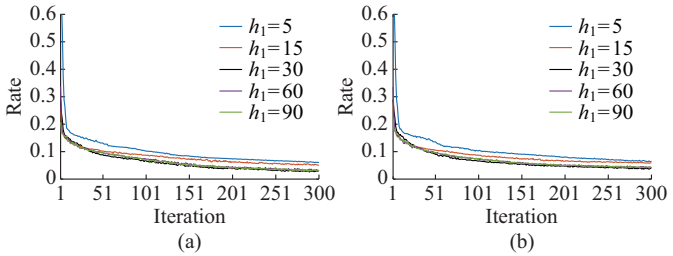

Fig. 7. Average misjudgment rates of network with one hidden layer . (a) Rates with training samples with 300 iterations. (b) Rates with test samples with 300 iterations.

The second hidden layer is then added, and the training process is repeated. The average misjudgment rates of the network with two hidden layers are demonstrated in Fig. 8. It can be seen that the rates do not vary a lot after convergence. The lowest average misjudgment rates are 3.83% and 4.30% for training sets and test sets, respectively, when the second hidden layer contains 30 neurons ($h_2 = 30$). However, they are larger than the rates with only one hidden layer.

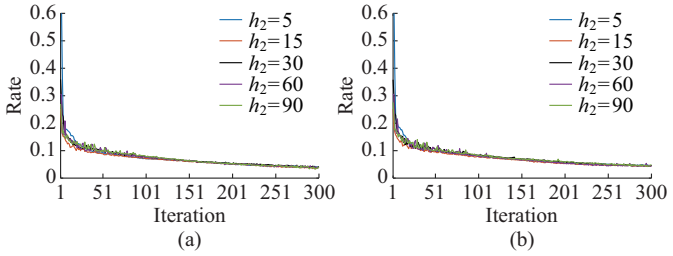

Fig. 8. Average misjudgment rates of network with two hidden layers . (a) Rates with training samples with 300 iterations. (b) Rates with test samples with 300 iterations.

To determine the most suitable network depth, the effect of the number of hidden layers on the average recognition rates is demonstrated in Fig. 9. The increase in network depth is unable to improve the average recognition rate. Therefore, the network structure is designed to be 30-30-3, which means that the number of neutrons in the input layer is 30, in the hidden layer is 30, and in the output layer is 3. The input layer size equals to the sample dimension which is actually the length of normalized line-mode currents, and the output size equals to the number of transient types 3. Here, 100, 010 and 001 are used for GFs, LFs, and LDs, respectively.

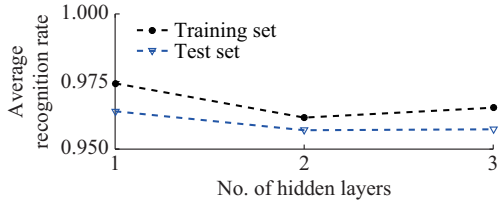

Fig. 9. Effect of number of hidden layers on average recognition rates.

D. Recognition Results

The samples from the four scenarios are recognized with the trained network, and the recognition rates are shown in Table IV. All GFs can be accurately recognized, while some of the lightning transients will be misjudged. The results show that the proposed method is efficient in recognizing transients in VSC-HVDC systems.

TABLE IV
RECOGNITION RESULTS OF TRANSIENTS

Type	Recognition rate (%)	
	Training set	Test set
GF	100.00	100.00
LF	93.80	91.20
LD	98.50	98.00
Overall	97.43	96.40

VI. PERFORMANCE ANALYSIS AND DISCUSSIONS

To demonstrate the effectiveness of the proposed method, its performance is discussed with different inputs, signal-to-noise ratios (SNRs), sample sizes, normalizations, and sampling frequencies.

A. Effect of Inputs

As aforementioned in Section IV, line-mode components of transients are more stable than zero-mode components. The recognition results of transients with line-mode components and zero-mode components are compared and shown in Fig. 10. With a stable representation of transients, the recognition rates of the model with line-mode inputs are higher than thoses with zero-mode inputs. It is suitable to adopt the line-mode components of transients as the input of the recognition model.

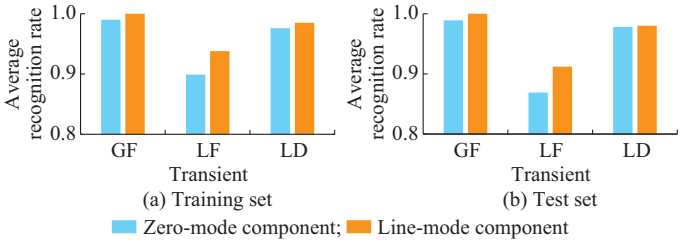

Fig. 10. Comparison of recognition results with different components. (a) Training set. (b) Test set.

B. Effect of SNRs

Gaussian white noises are the primary source that pollutes measured data. Its impact on the recognition results should be considered and discussed. Simulated Gaussian white noises are added to original transient samples and recognized by the proposed DBN model. Figure 11 shows the average recognition rates under different SNRs. With the decrease of SNRs, the recognition rates of faults decrease a little, but the recognition rates of disturbances vary a lot. However, even the SNR is as low as 30 dB, the recognition rate of GF keeps around 98%, while the recognition rates of LF and LD are around 90%. The proposed method can reflects the effect of noise.

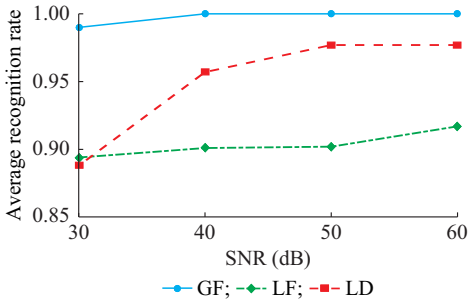

Fig. 11. Effect of SNRs on average recognition rates.

C. Effect of Sample Sizes

The number of samples affects the performance of recognition models. While for electrical power systems, samples of transients on VSC-HVDCs are limited. This paper investigates the recognition rates of the proposed method under different sizes of training sample sets. The results are shown in Fig. 12. The recognition rates can be improved with the increase of the number of training samples. When only 25 training samples from each type of transients, i.e., 75 training samples in total, are adopted, the recognition rates of GF, LD, and LF are close to 100%, 95%, and 80%, respectively. When the training samples of each type of transients reach 100, i.e., 300 training samples in total, the recognition rate of LF is slightly higher than 90%, and the overall recognition rate of three types of transients is close to 95%. In summary, the proposed method can recognize transients effectively with a limited number of training samples, and the recognition performance can be improved with more training samples.

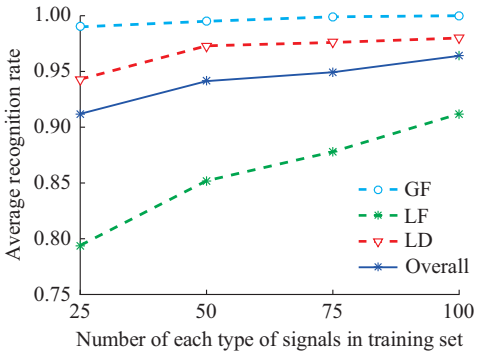

Fig. 12. Effect of number of training samples on average recognition rates.

D. Effect of Normalizations

For neural networks, normalizations of the input and output data can generally speed up learning rates and leads to faster convergence. Different kinds of normalizations are used, for example, min-max normalization, *z*-score normalization, median absolute deviation (MAD) normalization, and sigmoid normalization [33], [34]. Some of these normalizations are robust to outliers, and some of them can change the original distributions of data. The recognition performance of proposed method is tested with different kinds of normalizations, as illustrated in Fig. 13.

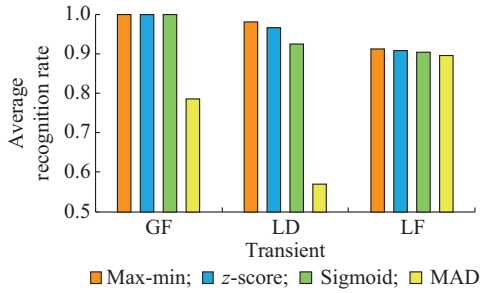

Fig. 13. Effect of normalizations on average recognition rates.

Although the min-max normalization is not robust to outliers, the recognition rates of min-max normalization are better than the other normalizations. Both sigmoid and MAD normalizations are robust. They transform the input values to follow the Gaussian distribution, and the normalized input values are clustered around 0. It makes the features of different transients unclear. Min-max and *z*-score normalizations focus more on the waveforms. In particular, min-max normalization ensures that all input values fall in the interval [0, 1]. The maximum and minimum are 1 and 0 for each input value, respectively. The waveform features of different transients are clearly revealed.

E. Effect of Sampling Frequencies

More information can be included for a certain time window when sampling frequency is increased. Here, the effect of sampling frequency is discussed, and the recognition results are shown in Fig. 14. To avoid the influence from DBN models, only the size of input layer is changed. Thus, the DBN structure is 30-30-3 for 10 kHz scenario, 300-30-3 for 100 kHz scenario, and 3000-30-3 for 1 MHz scenarios. All GB-RBM parameters keep the same. Three hundred samples are used for training and test, respectively.

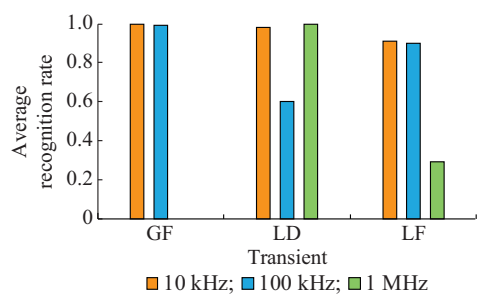

Fig. 14. Effect of sampling frequency on average recognition rates.

As illustrated by Fig. 14, the recognition rate decreases when sampling frequency increases. With the increase of input size, the number of hidden layer and the sample size of each layer are not changed accordingly. It is hard for DBN models to reach convergence. The overall training errors of full batch (all training samples) can be about 1 when 1 MHz sampling data are used. The recognition rate of GF is 0 and the overall rate is only 43%. Thus, when sampling frequency increases, both the width and depth of DBN model must be adjusted to generate better recognition results.

F. Generalization Performance of Different Systems

To further analyze the generalization performance, all the training samples in scenario 1 are removed from the original training set and used as the test set. Thus, the new training set does not include any sample in scenario 1, and the generalization performance of the proposed model can be tested. The same experiment is also implemented in scenarios 2, 3, and 4, respectively. Figures 15 presents the recognition results of each scenario. When the method is applied to the modified scenarios, it has a recognition rate of 100% for GF and a recognition rate of more than 95% for LD. However, the system difference has a more significant impact on the recognition rate of LF, which is about 90% for scenarios 2 and 3, and about 80% for scenarios 1 and 4.

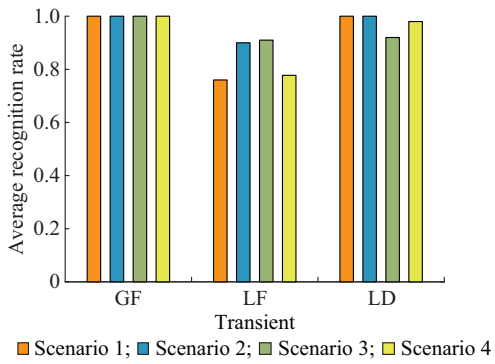

Fig. 15. Recognition rates in different scenarios.

VII. Comparisons with Traditional Methods

To illustrate the effectiveness of the proposed method, its performance is compared with existing methods.

A. Existing Machine Learning Based Recognition Methods

1) Features with Classifiers

Feature extraction with the trained classifier is one of the most common methods in traditional pattern recognition. When characterizing the features of electrical transients, some feature extraction methods are often used, for example, wavelet energy and principle component analysis (PCA). The recognition performance of the two methods are discussed in this paper.

Wavelet energy reveals the spectrum of signals with multi-resolution analysis. The definition of wavelet energy is shown in (5)-(7) [35], [36].

$$E_i = \left\| \boldsymbol{c}_i \right\| = \mathrm{sqrt}\left(\sum_{j=1}^{n} c_{ij}^2 \right) \tag{5}$$

$$\boldsymbol{E}_o = [E_1 \quad E_2 \quad \ldots \quad E_N] \tag{6}$$

$$\boldsymbol{E} = \boldsymbol{E}_o / \left\| \boldsymbol{E}_o \right\| \tag{7}$$

where E_i is the wavelet energy of the i^{th} layer; c_{ij} ($j = 1, 2, \ldots, n$) is the j^{th} element in the wavelet coefficient vector of the i^{th} layer $\boldsymbol{c}_i$; and $\boldsymbol{E}$ is the normalized $\boldsymbol{E}_o$. The dimension of the feature vector is selected as 4 in this paper, and Daubechies "db4" wavelet is applied.

PCA is a dimensionality-reduction method. It transforms a large variable set into a smaller one that still contains most of the information in the large set. Its definition is shown in (8) and (9) [4].

$$\boldsymbol{S} = \frac{1}{n-1} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \tag{8}$$

$$\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\lambda}\boldsymbol{V}^{\mathrm{T}} = [v_1 \quad v_2 \quad \ldots \quad v_p] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_p \end{bmatrix} \tag{9}$$

where $\boldsymbol{S}$ is the variance-covariance matrix, which is used to calculate eigenvalues $\boldsymbol{\lambda}$ and eigenvectors $\boldsymbol{V}$; $\boldsymbol{X}$ is the original raw data; and $\boldsymbol{\Sigma}$ is the largest principal component. The other principal components can be calculated likewise. The largest k components are used as the features of k-dimension vector. In this research, the raw data is condensed to a 6-dimension vector.

For classifiers, BP neural network and SVM are widely used ones. Therefore, with a different combination of features and classifiers, three feature-based recognition methods are discussed: method 1 is the combination of wavelet energy and BP neural network; method 2 is the combination of wavelet energy and SVM; and method 3 is the combination of PCA and BP neural network. The structure of BP neural network is 4-28-3 for method 1 and 6-4-3 for method 3. The activation function is the hyperbolic tangent function, and the training procedure uses the gradient descent algorithm with momentum BP and dynamic adaptive learning rate. The number of iterations is 500, and the learning rate is 0.01. For method 2, classification is carried out by using SVM with a typical "one-to-one" construction. That is, an SVM needs to distinguish two types of samples, and three SVMs in total are needed for three-type recognitions. All SVMs use the radial basis function (RBF) as the kernel function.

2) Raw Data with Classifiers

When the input number is not massive enough, the supervised learning algorithms can also process raw data without feature extraction. However, such a procedure will cost more time to get converged. Therefore, the normalized line-mode components of transients are also used as the input of traditional classifiers: BP neural network (method 4) and SVM (method 5). For method 4, the parameters of the BP neural network are the same as those in method 1, but the network structure is chosen as 30-15-3. For method 5, the SVMs are constructed the same as for method 2.

3) Comparison Results and Analysis

The same transient samples are used for the five methods mentioned above, as well as the proposed method. The recognition results are shown in Table V.

TABLE V
COMPARISON OF RECOGNITION RATES WITH DIFFERENT METHODS

Method	Recognition rate (%)			
	GF	LF	LD	Overall
1	99.45	64.78	67.87	77.37
2	99.20	69.00	68.10	78.77
3	100.00	80.00	69.70	83.23
4	98.98	83.30	94.18	92.15
5	99.10	84.40	97.80	93.77
Proposed method	100.00	91.20	98.00	96.40

It can be seen from Table V that no matter which method is used, GF can be effectively recognized. All recognition rates of GF of the six methods are higher than 98%. Both method 3 and the proposed method have the highest recognition rates of 100%.

The recognition rates of LD and LF vary a lot. In particular, the recognition rates of feature-based methods, i.e., methods 1, 2 and 3, are about 70%-80%. This phenomenon might result from the low sampling frequency (10 kHz) and a short data window (3 ms). As only 30 data values are adopted for recognition, the time-domain or the time-frequency domain features of different kinds of transients cannot be fully revealed. Among the feature-based methods, method 3 has better performance. Its overall recognition rate is higher than 80%. For feature-based methods, the overall recognition rates are lower than 90%, while non-feature-based ones have higher overall recognition rates. Two independent processes, i.e., feature extraction and classification, are needed for feature-based methods. Their recognition performance depends on both the quality of features and the design of classifiers. As the dimension of raw data is only 30, it is also possible to apply end-to-end recognition with traditional machine-learning models, such as BP neural network and SVM. The end-to-end solution can avoid the errors induced by hand-designed processes such as feature extraction. Without the dimension reduction of feature extraction, the end-to-end recognition methods can generate higher recognition rates. Compared with traditional classifier-based methods, the proposed method can perform better. DBN learns features from raw data with unsupervised learning algorithms, thus more robust features of raw data can be learnt. The unsupervised learning algorithms included in the proposed methods provide more effective characterizing ability than traditional supervised ones.

B. Existing Engineering Method

Since LD is an interference which will disappear after a short duration, the electrical measurements will get back to their values in normal operations. In practical engineering, the integration of voltage, as demonstrated in (10), is used to discriminate LD from the faults on a current-controlled HVDC transmission line [3]. Generally, the integration requires a delay of 0.5 ms to avoid the transient voltage spike caused by lightning.

$$\left| \sum_{k=N_S+N_{shift}}^{N} \Delta u(k) \right| > V_{th} \tag{10}$$

where N_S is the starting instant of lightning stroke at measuring unit; N_{shift} is the time delay which is 0.5 ms; N is the time length of data, which is usually tens of milliseconds; $\Delta u(k)$ is the voltage gradient; and V_{th} is a threshold. Table VI illustrates the recognition results of practical integration-based recognition method and the proposed method.

TABLE VI
RECOGNITION RATES OF DIFFERENT TYPES OF TRANSIENTS

Type	Recognition rate (%)	
	Practical method	Proposed method
GF	30.20	100.00
LF	100.00	91.20
LD	39.50	98.00
Overall	56.56	96.40

Since only 3 ms data are used, the voltage of LD cannot reach the post-transient stability. The recognition performance of the practical integration-based method is ineffective. Due to the large-amplitude currents of lightning strokes, the voltage gradient of LF is much larger than those of GF and LD, and better recognition rate (100%) can be obtained.

VIII. CONCLUSION

This paper proposes a DBN-based transient recognition method for VSC-HVDC transmission systems. This method provides reliable end-to-end recognition without complex data processing. Fundamentals and characteristics of three typical types of transients on transmission lines are analyzed for designing an effective recognition model. The simulation results illustrate the effectiveness of the proposed method with a low sampling frequency within an extremely short-time window. The following conclusions can be achieved.

1) The proposed method can generate excellent recognition results within a short time with low sampling frequency. The overall recognition rate is higher than 96%. With more training samples, the recognition rates can be further improved.

2) Compared with existing methods, the proposed method shows better performance. The machine learning based methods discussed in this paper has better performance than the practical integration-based method. The overall recognition rate of the proposed method is about 40% higher than practical methods.

3) The proposed method has good recognition capability and generalization performance, and it has high potential in practical applications.

4) In future work, the recognition will be focused on discriminating faults and disturbances. The misjudgment rates of LFs can be reduced because those recognized to be LFs will not be counted.

REFERENCES

[1] J. Beerten, S. Cole, and R. Belmans, "Modeling of multi-terminal VSC HVDC systems with distributed DC voltage control," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 34-42, Jan. 2014.

[2] L. Zhang, L. Harnefors, and H. Nee, "Interconnection of two very weak AC systems by VSC-HVDC links using power-synchronization control," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 344-355, May 2011.
[3] F. Kong, Z. Hao, S. Zhang *et al*., "Development of a novel protection device for bipolar HVDC transmission lines," *IEEE Transactions on Power Delivery*, vol. 29, no. 5, pp. 2270-2278, Mar. 2014.
[4] J. Morales, E. A. Orduña, and C. Rehtanz, "Identification of lightning stroke due to shielding failure and back flashover for ultra-high-speed transmission-line protection," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 2008-2017, Jul. 2014.
[5] J. A. Morales, E. Orduña, and C. Rehtanz, "Classification of lightning stroke on transmission line using multi-resolution analysis and machine learning," *International Journal of Electrical Power & Energy Systems*, vol. 58, pp. 19-31, Jun. 2014.
[6] J. A. Morales, E. Orduña, C. Rehtanz *et al*., "Comparison between principal component analysis and wavelet transform filtering methods for lightning stroke classification on transmission lines," *Electric Power Systems Research*, vol. 118, pp. 37-46, Jan. 2015.
[7] K. Satpathi, Y. M. Yeap, A. Ukil *et al*., "Short-time Fourier transform based transient analysis of VSC interfaced point-to-point DC system," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4080-4091, Oct. 2018.
[8] K. D. Kerf, K. Srivastava, M. Reza *et al*., "Wavelet-based protection strategy for DC faults in multi-terminal VSC HVDC systems," *IET Generation, Transmission & Distribution*, vol. 5, no. 4, pp. 496-503, Apr. 2011.
[9] R. Bertho, V. A. Lacerda, R. M. Monaro *et al*., "Selective nonunit protection technique for multiterminal VSC-HVDC grids," *IEEE Transactions on Power Delivery*, vol. 33, no. 5, pp. 2106-2114, Sept. 2018.
[10] F. Deng, X. Li, and X. Zeng, "Single-ended travelling wave protection algorithm based on full waveform in the time and frequency domains," *IET Generation, Transmission & Distribution*, vol. 12, no. 15, pp. 3680-3691, Aug. 2018.
[11] G. Luo, Q. Lin, L. Zhou *et al*., "Recognition of traveling surges in HVDC with wavelet entropy," *Entropy*, vol. 19, no. 5, May 2017.
[12] G. Luo, C. Yao, Y. Liu *et al*., "Entropy SVM-based recognition of transient surges in HVDC transmissions," *Entropy*, vol. 20, no. 6, pp. 1-13, May 2018.
[13] H. Xiao, Y. Li, R. Liu *et al*., "Single-end time-domain transient electrical signals based protection principle and its efficient setting calculation method for LCC-HVDC lines," *IET Generation, Transmission & Distribution*, vol. 11, no. 5, pp. 1233-1242, Mar. 2017.
[14] Z. He, L. Fu, S. Lin *et al*., "Fault detection and classification in EHV transmission line based on wavelet singular entropy," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2156-2163, Oct. 2010.
[15] Q. Huai, K. Liu, L. Qin *et al*., "Backup-protection scheme for multiterminal HVDC system based on wavelet-packet-energy entropy," *IEEE Access*, vol. 7, pp. 49790-49803, Apr. 2019.
[16] Z. Liu, Z. Han, Y. Zhang *et al*., "Multiwavelet packet entropy and its application in transmission line fault recognition and classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2043-2052, Feb. 2014.
[17] W. Fan and Y. Liao, "Wide area measurements based fault detection and location method for transmission lines," *Protection and Control of Modern Power Systems*, vol. 4, no. 1, pp. 53-64, Mar. 2019.
[18] S. R. Samantaray, P. K. Dash, and S. K. Upadhyay, "Adaptive Kalman filter and neural network based high impedance fault detection in power distribution networks," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 4, pp. 167-172, May 2009.
[19] L. Guo, Y. Lei, S. Xing *et al*., "Deep convolutional transfer learning network: a new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316-7325, Oct. 2019.
[20] T. Wu and W. U. Bajwa, "Learning the nonlinear geometry of high-dimensional data: models and algorithms," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6229-6244, Dec. 2015.
[21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, p. 504, Jul. 2006.
[22] Y. Bengio, *Learning Deep Architectures for AI*. New York: Now Foundations and Trends, 2009.
[23] B. Chang, O. Cwikowski, M. Barnes *et al*., "Point-to-point two-level converter system faults analysis," in *Proceedings of 7th IET International Conference on Power Electronics, Machines and Drives (PEMD 2014)*, Manchester, UK, Apr. 2014, pp. 1-6.
[24] N. Flourentzou, V. G. Agelidis, and G. D. Demetriades, "VSC-based HVDC power transmission systems: an overview," *IEEE Transactions on Power Electronics*, vol. 24, no. 3, pp. 592-602, Feb. 2009.
[25] L. Jiang, Q. Chen, L. Wang *et al*., "Novel protection method for VSC-HVDC transmission lines," *The Journal of Engineering*, vol. 2019, no. 16, pp. 2142-2146, Apr. 2019.
[26] A. Moawwad, M. S. E. Moursi, W. Xiao *et al*., "Novel configuration and transient management control strategy for VSC-HVDC," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2478-2488, Jun. 2014.
[27] J. Takami and S. Okabe, "Observational results of lightning current on transmission towers," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 547-556, Dec. 2007.
[28] S. Okabe, T. Tsuboi, and J. Takami, "Analysis of aspects of lightning strokes to large-sized transmission lines," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 1, pp. 182-191, Jan. 2011.
[29] J. Takami and S. Okabe, "Characteristics of direct lightning strokes to phase conductors of UHV transmission lines," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 537-546, Dec. 2007.
[30] J. A. Martinez and F. Castro-Aranda, "Lightning performance analysis of overhead transmission lines using the EMTP," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 2200-2210, Jun. 2005.
[31] X.-H. He, D. Wang, Y.-F. Li *et al*., "A novel bearing fault diagnosis method based on gaussian restricted Boltzmann machine," *Mathematical Problems in Engineering*, vol. 2016, no. 3, pp. 1-8, Dec. 2016.
[32] V. T. Tran, F. Al-Thobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113-4122, Jul. 2014.
[33] A. Ross and K. Nandakumar, "Fusion, score-level," in *Encyclopedia of Biometrics*, 2nd ed. New York: Springer, 2015, pp. 610-614.
[34] T. Jayalakshmi and A. Santhakumaran, "Statistical normalization and back propagation for classification," *International Journal of Computer Theory and Engineering*, vol. 3, no. 1, pp. 89-93, Feb. 2011.
[35] G. Luo, D. Zhang, K. J. Tseng *et al*., "Impulsive noise reduction for transient earth voltage-based partial discharge using wavelet-entropy," *IET Science, Measurement & Technology*, vol. 10, no. 1, pp. 69-76, Jan. 2016.
[36] B. Li, J. He, Y. Li *et al*., "A review of the protection for the multi-terminal VSC-HVDC grid," *Protection and Control of Modern Power Systems*, vol. 4, no. 3, pp. 239-249, Nov. 2019.

Guomin Luo received the B.Sc. and M.Sc. degrees from Southwest Jiaotong University, Chengdu, China, in 2005 and 2008, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, Singapore, in 2013. She is currently an Associate Professor with Beijing Jiaotong University, Beijing, China. Her research areas include online monitoring, signal processing, and protection of power systems.

Jiaxin Hei is currently a research student in the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China. His research interests include DC system fault location and artificial intelligence.

Changyuan Yao is an engineer in State Grid Shandong Electric Power Company, Jinan, China. His research interest is maintenance of transmission system.

Jinghan He received the B.Sc. and M.Sc. degrees in automation from Tianjing University, Tianjing, China, in 1987 and 1994, respectively, and the Ph.D. degree from Beijing Jiaotong University, Beijing, China. She is currently a Professor with Beijing Jiaotong University, Beijing, China. Her main research interests include online monitoring, protection and control of power systems, power quality, new energy and smart grid, and electrical rail transportation.

Meng Li received the B.S. and Ph.D. degrees in electrical engineering from North China Electric Power University, Beijing, China, in 2003 and 2018, respectively. He is currently a Lecturer at the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China. His research interest is DC grid protection.