IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 18, 2019, accepted January 3, 2020, date of publication January 14, 2020, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966645

Stacked Auto-Encoder-Based Transients Recognition in VSC-HVDC

GUOMIN LUO[1], MENGXIAO CHENG[1], HAO SUN[1], MENG LI[1], YINGJIE TAN[1], JINGHAN HE[1], AND HENG ZHANG[2]

[1]School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China
[2]State Grid Beijing Maintenance Company, Beijing 100069, China

Corresponding author: Meng Li (mengl@bjtu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018JBM056, in part by the National Natural Science Foundation of China under Grant 51507008, in part by the China Postdoctoral Science Foundation under Grant 2019M660436, and in part by the State Grid Corporation Technology Project under Grant 5200-201956113A-0-0-00.

ABSTRACT For overhead long-distance high voltage direct current (HVDC) transmission lines, transients are produced due to complicated field conditions and lightning activities. To ensure reliable operation of protection devices, accurate recognition of faults and disturbances is quite critical. The most popular recognition methods include threshold-based ones which require the proper setting of the threshold value, and classifier-based ones that need suitable feature extractions. These methods depend heavily on the experience of engineers or experts and are ineffective in dealing with the variation of system parameters. In this paper, a transient recognition method based on stack auto-encoder (SAE) is proposed to characterize different transients and to avoid human interferences. A symmetrical ±500kv HVDC system is modeled to illustrate the performance of the proposed method. The effect of some factors, such as noises and conductors, are discussed and compared. The simulation results demonstrate that the proposed SAE-based recognition has excellent potential in transient recognition of practical HVDC systems.

INDEX TERMS Stacked auto-encoder, VSC-HVDC, transient recognition.

I. INTRODUCTION

Considering some advantages, such as lower loss, more flexible control, and larger transmission capacity, high voltage direct current (HVDC) becomes a better choice for long-distance power transmission. The traditional Line Commutated Converter (LCC) based HVDC transmission can only provide a point-to-point topology that connects two AC systems with large power sources. However, if the AC system is not strong enough and faults occur, the LCC-HVDC commutation failure may occur consequently. The wide use of controllable power electronics, such as IGBTs (Insulated Gate Bipolar Transistors), accelerates the development of Voltage Source Converters (VSCs), where active power and reactive power can be controlled separately. When compared with LCC-HVDC, the VSC-based ones can easily realize power reversal and avoid commutation failures [1], [2]. Also, VSC-HVDC is feasible to form practical meshed grids. VSC-HVDC transmission is a good choice to deliver large amounts of energy to loads in different regions [3], [4].

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

VSC-HVDC generally adopts capacitors to support the DC bus voltages. Once faults occur on transmission lines, these voltage-supporting capacitors will discharge quickly and result in large fault currents. In order to eliminate the rapidly increasing fault currents and avoid huge damage to the converter, transient protection such as travelling wave protection, which can make a judgment in only a few milliseconds, is commonly used as the primary protection of VSC-HVDCs [5]. Practically, many HVDC systems use overhead transmission lines for their economic construction and high-voltage insulation capability. When compared with cables, overhead transmission lines are economical and convenient for the construction, but more likely to be affected by faults and interferences. Lightning strokes are the most common transients on overhead transmission lines. Even though the lightning strokes may not lead to faults, they can also produce rapid increasing currents, which are similar to grounding faults, and lead to misjudgments of protection devices. So, to ensure the stability and reliability of VSC-HVDC, the protection of the transmission line should be effective and robust [6]. When a large-amplitude transient is detected, it is crucial to identify whether the transient is a disturbance or

a fault and make a response as soon as possible. Therefore, the recognition of transients on HVDC transmission lines is a critical function of protection units.

Traditional transient recognition methods usually contain two parts: feature extraction and classification. A lot of methods are used for feature extraction, such as S transform, wavelet transform, mathematical morphology, and so on. S-transform has high frequency-resolution at low-frequency range, but it cannot produce reliable performance when analyzing electrical transients whose information is contained in high-frequency ranges [7]–[9]. Wavelet transform is also very popular and useful in characterizing time-frequency features [10]–[12]. A number of forms, for example, wavelet entropy [13], wavelet energy [14], and relative entropy of wavelet energy [15], are widely employed. But the wavelet-based features depend on the selection of wavelet basis. Additionally, the time delay is another problem of wavelet-based features in real-time processing [16]. Mathematical morphology-based features, which reveal time-frequency characteristics of transients [17]–[19], are more likely to rely on the experiences of engineers. The existing feature extraction methods are proved to be effective in many circumstances, but most of them need the help of experts to produce reliable performances.

For the classification, thresholding is the basic method, while classifiers are regarded to produce better results. But the performance of classifiers depends on the extracted features and their application scenarios. For example, Support Vector Machine (SVM) is suitable for solving two-type classifications with small samples [20], [21]; the shallow neural network is difficult to represent complicated functions, and training deep neural network is a tough job as the number of training samples in power system is usually limited [22], [23]. For a practical HVDC transmission, only a few transients can be collected to form a small sample. The classifier trained with this kind of sample might not perform well on the transients from other similar systems. Therefore, for transient recognition in practical systems, a classifier with higher accuracy and better generalization is needed.

The performance of above mentioned traditional recognition methods depends heavily on how unique and stable the features are. But fault transients on HVDC transmission lines show obvious nonlinear characteristics and are easily affected by high frequency disturbances. It is not an easy job to find a stable and reliable representation for such nonlinear signals. The deep learning algorithms with layerwise unsupervised learning can provide a better solution for extracting nonlinear features and modeling complex relationships between variables. Also, better generalization performance can be achieved. As one of the most popular unsupervised learning methods, a Stacked Auto-Encoder (SAE) is a neural network consisting of multiple layers of sparse Auto-encoders in which the outputs of each layer are wired to the inputs of the successive layer. It extracts features of signals layer by layer. The SAE-based features are independent of experiences and human interferences, and they are suitable for fault classification in complex nonlinear systems. But the related research are rarely reported.

Based on the feature learning capability of SAE, a SAE-based end-to-end transient recognition method is proposed in this paper. The following contents of this paper are organized as follows: Section 2 introduced the fundamentals of VSC-HVDC systems, and three kinds of transients are modelled and analyzed. In Section 3, the definition of SAE is introduced, and a SAE is designed for transient recognition in HVDC. In Section 4, the architecture and functions of the proposed SAE are discussed and properly selected. The performance of the proposed method is demonstrated and compared in Section 5. The simulation results show the proposed SAE-based method is effective, robust, and has better generalization performance.

II. TRANSIENTS IN HVDC TRANSMISSIONS

A. FUNDAMENTALS OF DC TRANSMISSIONS

In HVDC transmission systems, the converter station is an important equipment that converts energy and connects both DC and AC systems. The topology and grounding methods of the converter determine the transient features of HVDC systems. The most commonly discussed topologies of VSC-HVDC systems include: symmetrical single polar HVDC, unsymmetrical single polar HVDC, and bipolar HVDC [24], [25]. Due to its economical operation and high efficiency, symmetrical single polar HVDC is used in some practical projects, for example, Shanghai Nanhui HVDC that connects wind farms to transmission systems. A typical symmetrical single polar HVDC is presented in Fig.1. Here, two-level VSC convertors are adopted. The VSC convertors are directly grounded at the DC side and grounded with a large resistor at the AC side. Since the neutral point of supporting capacitor is grounded, the potentials of two transmission lines have opposite polarities but same amplitudes. Current limiting reactors L are usually equipped at the terminals of transmission lines to eliminate the rise rate of fault current.

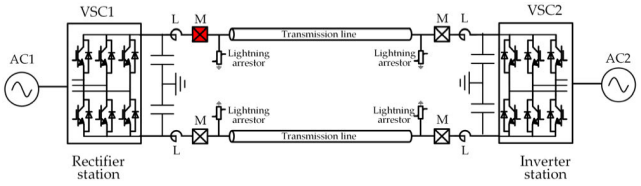

FIGURE 1. Diagram of a point-to-point VSC-HVDC transmission system.

B. TYPES OF TRANSIENTS

For overhead transmission lines, most of transients are generated by lightning strokes, which account for more than 40% of faults or even around 80% in some cases. Additionally, 70% of short-circuit faults are pole-to-ground ones [5].

If a large-amplitude transient is detected by the measurement units M, the protection devices of transmission line will be activated and distinguish disturbances from faults. Due to

the existence of reactors, transients beyond the transmission lines will be eliminated, and only the faults within the transmission line need to be identified. Therefore, three types of transients are discussed in this paper: single-pole to ground fault, lightning disturbance and lightning fault.

1) GROUND FAULT

Single pole to ground fault means one of the DC transmission lines is grounded with a fault. The fault transients include two phases: capacitor discharging and AC current feeding. In both phases, the voltage of the faulted pole decreases, and the current in the faulted pole increases. As demonstrated by the symmetrical single polar HVDC shown in Fig.1, the single-pole-to-ground fault will result in a voltage increase of the normal pole. As the voltage of the faulted pole decreases, the fault current will reach its peak and decay to a small value within a quite short time.

Short circuit faults can be modelled mathematically by a voltage source with a step waveform, as in (1). The value of superposed voltage source E is the reverse of the voltage at the faulted point just before fault occurrence.

$$V(t) = E\varepsilon(t) \tag{1}$$

2) LIGHTNING DISTURBANCE AND LIGHTNING FAULT

The lightning strokes that cause overvoltage and overcurrent transients on overhead HVDC transmission lines can be classified into two types: indirect strokes which strike on towers, grounding wires, or buildings near the transmission line; and direct strokes which hits the overhead transmission wire directly. In practice, indirect strokes can rarely lead to insulator flashover on transmission lines over 110 kV [26]. So, only the lightning transients caused by direct stokes are analyzed in this paper. When a lightning discharge strikes on HVDC transmission systems, it can be modelled by a single current source before the insulation breaking-down. But when the lightning overvoltage on the HVDC system exceeds the breaking-down threshold, grounding faults will follow, which is modelled by a step voltage source. The mathematic model of the lightning current source is a double-exponential impulse with shorter rise time and longer fall time, as in (2) [27].

$$i\,(t) = AI_L\,(\exp(-\alpha t) - \exp(-\beta t)) \tag{2}$$

where, A is the amplitude correction coefficient, I_L is the peak amplitude of the lightning current, α is the waveform coefficient of the rising time, and β is the waveform coefficient of the falling time. The current sources $i(t)$ with small magnitudes are used to simulate lightning disturbances. Lightning faults are simulated by both current source $i(t)$ and superposed voltage source $V(t)$.

C. WAVEFORM ANALYSIS OF DIFFERENT TRANSIENTS

From the above analysis, the transients on overhead HVDC transmissions are mainly responses of step voltage sources or impulsive current sources with different parameters, for example, magnitude, fault instant, fault location, and so on. The transient responses with the same kind of excitations might have similar waveforms. Thus, the analysis of different waveforms is quite useful for transient recognitions. Normalization is adopted to avoid the effects of amplitude. As in (3), linear normalization is employed in this analysis. Here, $max(x)$ and $min(x)$ are the maximum and minimum values of a transient data vector x, and z is the normalized vector.

$$z = (x - \min(x))/(\max(x) - \min(x)) \tag{3}$$

Fig.2 illustrates both the original waveforms and amplified normalizations of different transients: single-pole-to-ground fault (GF), lightning disturbance (LD), and lightning fault (LF). The effects of various factors, for example, transition resistances and fault locations, are discussed.

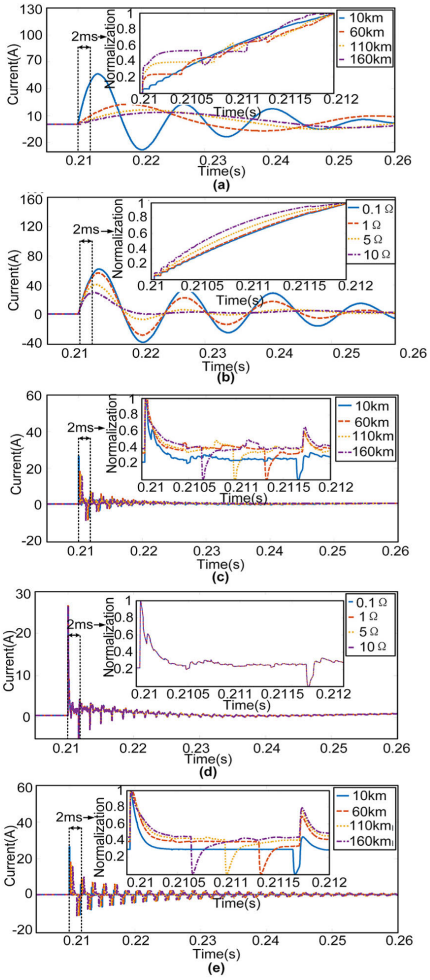

FIGURE 2. Transients before and after normalization, (a) GFs with different locations, (b) GFs with different resistances, (c) LFs with different locations, (d) LFs with different resistances, (e) LDs with different locations.

As demonstrated by Fig.2, normalization can eliminate the effect of grounding resistance. For example, the normalized waveforms in Fig.2(b) and Fig.2(d) are much more

similar than original ones. But the location-related variations still exist after normalization. The impulsive oscillations in Fig.2(c) repeat with different time delays when lightning strikes different locations. For all three kinds of normalized waveforms, GFs which is the response of voltage excitations are quite different from the others. LDs are the responses of current excitations, and LFs are the responses of both voltage and current excitations. LGs and LDs look quite similar as they are both current excitations before insulation breakdown.

III. FUNDAMENTALS OF STACKED AUTO-ENCODER

An auto-encoder (AE) is a type of artificial neural network used to encode data efficiently in an unsupervised way. It contains two parts: encoder, which compresses the input data with large dimension into a short code, and decoder, which reconstructs the code to be a signal that closely matches the original data. Therefore, with proper design, auto-encoders with unsupervised learning algorithms can be stacked and used in feature extraction and to form a transient-recognizing network.

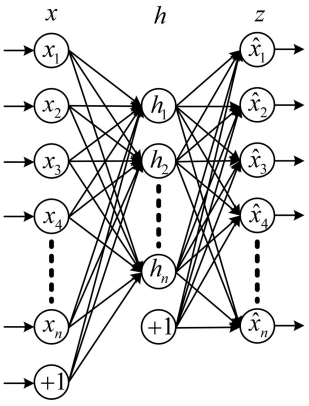

FIGURE 3. A typical structure of an AE.

A. INTRODUCTION OF AUTO-ENCODER

Architecturally, the simplest form of an auto-encoder is a feedforward, non-recurrent neural network which is quite similar to another single-layer perceptron. Fig.3 shows the structure of the simplest auto-encoder, where only one hidden layer is included. This encoder, which includes the input layer and the hidden layer, maps input x to space h. This mapping progress is as in (4). Here, W is a weight matrix, b is a bias vector, and S is an activation function such as a sigmoid function or a rectified linear unit. The vector $h = [h^{(1)}\ h^{(2)} \ldots h^{(m)}]$ is usually referred to as a code of the input x, m suggests the hidden layer dimension.

$$h = f(x) = S(Wx + b) \tag{4}$$

After that, the decoder, which contains the hidden layer and the output layer, maps h to the reconstruction z of the same shape as x. This decoding progress is as in (5). Here, $z = [z^{(1)}\ z^{(2)} \ldots z^{(n)}]$ is the reconstructed vector of code h, and n is the dimension of the input layer. W' and b' are the weight matrix and bias vector of the decoder. If the code space has lower dimensionality than the input space, the vector h can be regarded as a compressed representation, or in other words, the feature of the input x.

$$z = g(y) = S(W'y + b') \tag{5}$$

The auto-encoder is usually trained with a backpropagation algorithm to minimize its cost function, as in (6) and (7). By adjusting network parameters θ, weight matrix (W and W') and bias vectors (b and b'), the reconstruction error $L(x,y)$ can be minimized. A stochastic gradient descent algorithm, as in (8) and (9), is often used to update weights and bias for each iteration.

$$L(x, y) = -\sum_{i=1}^{n} [x_i \ln y_i + (1 - x_i) \ln (1 - y_i)] \tag{6}$$

$$J_{AE}(\theta) = \frac{1}{N} \sum_{i=1} L\left\{x^{(i)}, g\left[f\left(x^{(i)}\right)\right]\right\} \tag{7}$$

$$W_{ij} = W_{ij} - \varepsilon \frac{\partial}{\partial W_{ij}} J \tag{8}$$

$$b = b - \varepsilon \frac{\partial}{\partial b} J \tag{9}$$

B. INTRODUCTION OF STACKED AUTO-ENCODER

A stacked auto-encoder (SAE) is a neural network consisting of multiple layers of sparse auto-encoders in which the outputs of each layer are wired to the inputs of the successive layer. The formation process of a two-layer SAE is illustrated in Fig.4. The primary code $y1$ of the first AE (AE1) serves as the input of the second AE (AE2). The secondary code $y2$ of AE2 can be used as the features of raw input x.

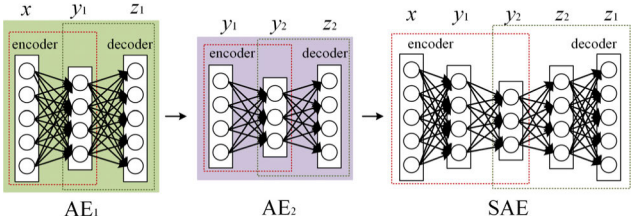

FIGURE 4. A demonstration of the stack of two AEs.

C. FEATURE EXTRACTION

Traditional feature extractions are mainly expert-based, where experts give feature extraction methods according to the analysis of certain kinds of data set. These methods are time-consuming, and feature representations might be different for various cases. Different from traditional expert-based ones, AE provides an unsupervised-learning way to extract features. The codes of three kinds of transients in HVDC overhead transmission are studied in this Section to test the feature extracting capability of SAE. Due to the speed requirement of protection response, only a short segment of post-fault transients can be adopted, for example,

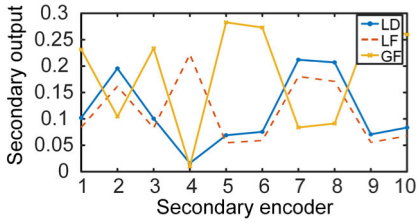

FIGURE 5. Secondary features of different transients.

a two-millisecond data segment. So the dimension of transient input is 200 when the sampling frequency is 10 kHz. A SAE with two hidden layers and a size of 200 - 50-10 is adopted to illustrate the feature extracting capability. Fig.5 shows the secondary features of different transients extracted by the designed SAE. For different transients, the distribution of secondary features should be as unique as possible. Compared with LD and LF, the secondary features of GF are entirely different. The secondary features of LF and LD are similar at most neutrons, but these two can still be distinguished at some neutrons such as neutron 4.

When using a SAE to encode the features of transients, it is necessary to consider the influences from grounding resistances and locations on both practical fault transient and disturbance transients. To demonstrate the coding stability of SAE, the effects of various factors on the secondary features are discussed. For the same transients, their feature distribution should be as similar as possible. As demonstrated by Fig.6, the features extracted by the SAE are insusceptible to distance variations.

Fig.7 shows the secondary features with different grounding resistances which vary from 1Ω to 20Ω. Although the change of grounding resistance can lead to slight variations at some neutrons, the trend of secondary features in Fig.7 is similar.

Wavelet energy is a commonly used tool in feature extraction. The transient signal is decomposed by wavelet to calculate the energy of each frequency band, and as in (10), where E_i is the wavelet energy of each frequency band, x_i is the wavelet coefficient of the calculated frequency band, and n is the number of decomposed level. The wavelet energy of all decomposed frequency bands forms a feature vector.

$$E_i = \sum_{k=1}^{n} |x_i|^2 \tag{10}$$

To demonstrate the stable performance of feature extraction, the features produced by both wavelet energy and SAE are compared. Here, 'db4' is selected to be the wavelet base, and the decomposition level is 10. Feature vectors $[E_1\ E_2\ \ldots\ E_{10}]$ of different kinds of transients are constructed and shown in Fig.8. They have the same dimensions with secondary features in Fig.8. The wavelet features of LD and LF are similar in general distribution. Compared with the SAE-features in Fig.5, where more unique distributions are

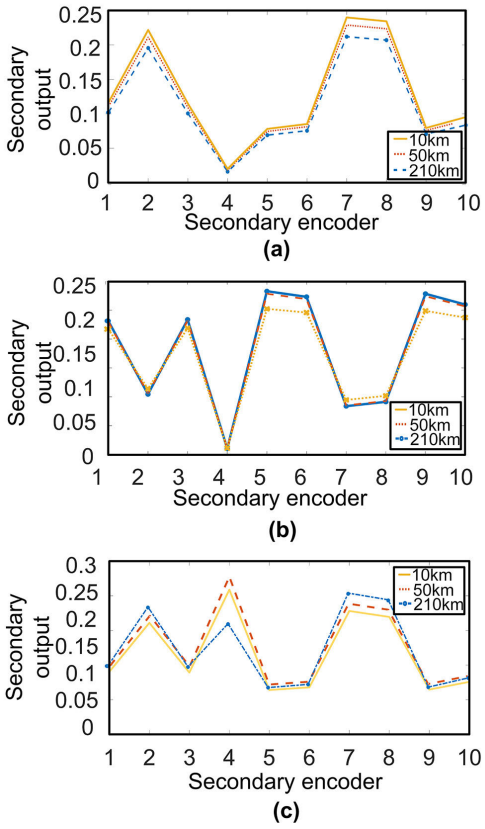

FIGURE 6. Effect of distance, (a) features of LD, (b) features of LF, (c) features of GF.

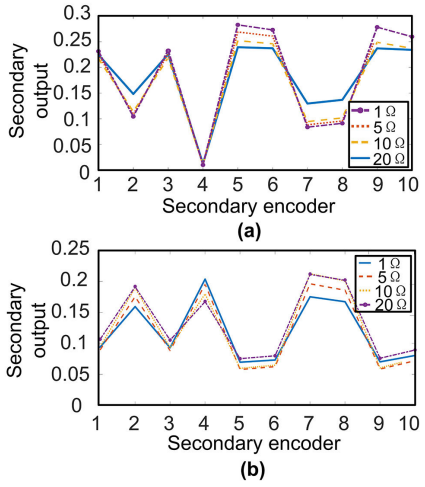

FIGURE 7. Effect of grounding resistances, (a) features of GF, (b) features of LF.

revealed, the wavelet-energy-features are more likely to produce incorrect recognitions if the same classifier is adopted.

The effect of distance and grounding resistance on wavelet-energy-features are studied and displayed in Fig.9 and Fig.10, respectively. Among the three types of transients, LDs appear to be more sensitive to distance than the others. The feature distributions of both LFs and GFs can be influenced by the

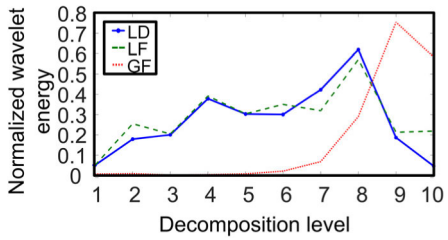

FIGURE 8. Wavelet energy features of different transients.

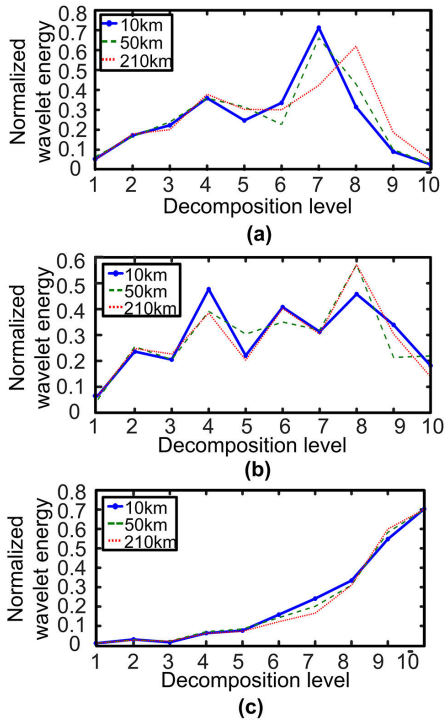

FIGURE 9. Effect of distance, (a) features of LD, (b) features of LF, (c) features of GF.

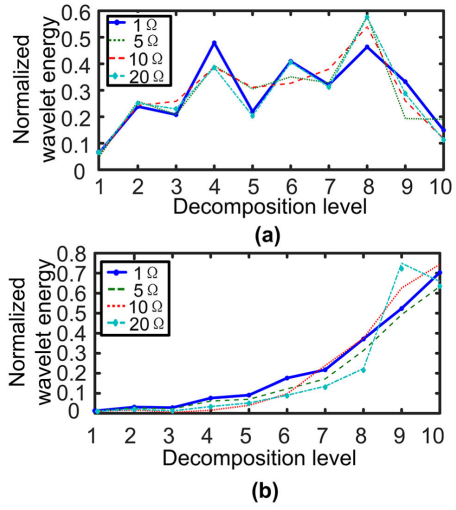

FIGURE 10. Effect of grounding resistance, (a) features of LF, (b) features of GF.

change of grounding resistances. Compared with SAE-based features, wavelet-energy-based ones are less robust and more sensitive to transient parameters.

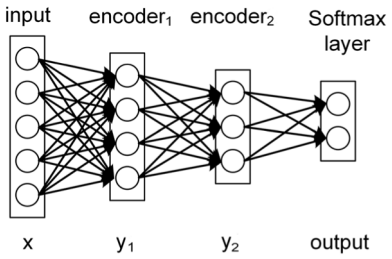

FIGURE 11. The structure of a SAE network with two hidden layers.

IV. SELECTION OF SAE PARAMETERS

With stable secondary features, a SAE can be used to recognize different transients if a Softmax layer is added. A structure of the SAE for classification is illustrated in Fig.11.

When used for classification, a good way to obtain proper parameters for a SAE is to use greedy layer-wise training. The training procedure is as follows:

1) Pre-training: the encoding layer of AEs are trained separately, and their optimal trained network parameters θ are kept. The features of AE is used as the input of the next one.

2) Network construction: These encoded features are treated as ''raw input'' to a Softmax classifier, which is trained to map encoded features to labelled outputs.

3) Network initialization: The parameters θ are initialized with the optimal network parameters of self-encoders after pre-training.

4) Fine-tuning: the entire network is fine-tuned. With labeled outputs, the SAE parameters are updated with back-propagation algorithms to achieve global optimization.

A feedforward network with a linear output layer and at least one hidden layer with any ''squashing'' activation function (such as the logistic sigmoid activation function) can approximate any Borel measurable function from one finite-dimensional space to another with any desired nonzero amount of error [28]. For different applications, the architecture and the functions should be properly selected to speed up the training procedure, to improve the generalization capability, and to reduce training errors. Here, the number of layers and the size of hidden layers are discussed when the architecture of SAE is designed. The functions which include activation function, loss function, and regulation function are also discussed.

A. SELECTION OF LAYER NUMBER

The size of input and output layers are the number of raw inputs and the number of transient types, respectively. The performance of different SAE networks are evaluated by their training time and errors. Since weights and thresholds of SAE are initialed randomly, the average performance of 5 trained networks is shown in Table 1. Except for the architecture, all training parameters were the same in this comparison. Obviously, the training time of SAE increases with the increase of the layer numbers. As the number of layers

TABLE 1. The training time and training errors with different number of layers.

Structure of net	Training time	MSE
200-50-3	11.39s	0.0144
200-100-10-3	17.98s	0.0032
200-150-100-50-3	39.84s	0.0106
200-150-100-50-10-3	48.56s	0.609
200-150-100-50-30-10-3	176.98s	0.572

increases from 3 to 4, the recognition rate increases and reaches its peak. The mean square errors (MSE) between target outputs and trained outputs are also used for evaluation. But when the number of layers continues increasing, the MSE decreases significantly, which is resulted from the overfitting that is generated by too many hidden layers. A SAE works best when its calculating capacity matches the complexity of the task and the amount of training data. The 4-layer network, which includes two AEs, has the lowest MSE and shorter training time, and it is thus selected in this paper.

B. SELECTION OF HIDDEN LAYER SIZES

As aforementioned, two hidden layers are chosen for the SAE. The size of hidden layers: the second layer S2 and the third layer S3, are discussed. To reduce the dimension of extracted features layer by layer, the size of the hidden layer should be smaller than that of its former one. The average training errors of SAE networks are adopted to help to set the architecture parameters. According to the errors shown in Fig.12, the SAE network has better performance when S2=13 and S3=85. So, the SAE architecture is selected to be 200-130-85-3 in this paper.

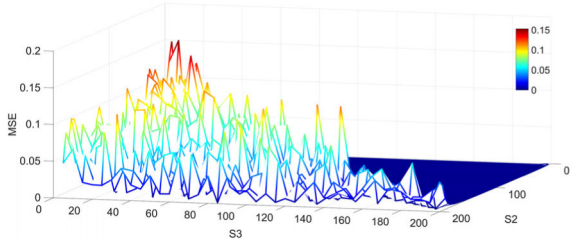

FIGURE 12. The distribution of training errors with different size of hidden layers.

C. SELECTION OF ACTIVATION FUNCTION

The expression ability of deep network is improved by using the nonlinear activation function as the activation function. The quality of an activation function can influence the performance of a neural network. To choose the appropriate activation function for this task, the network structure is ignored, but the saturation problems and the rate of convergence are discussed. During the training process, a saturation of the activation function which may lead to learning difficulties should be prevented. The activation values should approach their stable limits as fast as they can. Activation values of three common functions: rectifier (ReLU), sigmoid, and tangent sigmoid (Tansig) are discussed. Fig.13 shows the trend of the activation values of different layers during supervised learning. Both the mean values (bold lines) and the standard deviations (vertical bars) of 320 test samples at each epoch are demonstrated. Layer 1 and layer 2 stand for the outputs of the first and the second hidden layer, respectively.

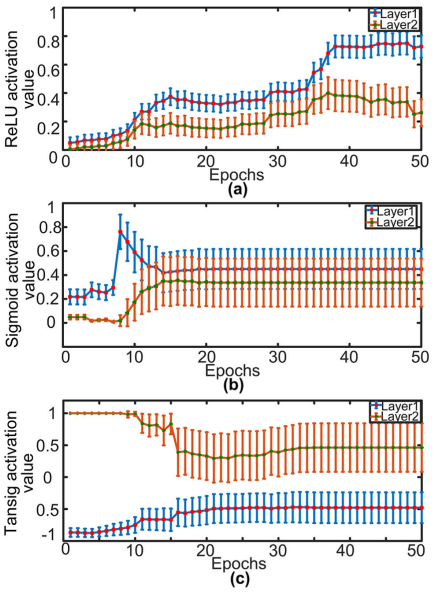

FIGURE 13. Mean and standard deviation of the activation values, (a) rectifier function, (b) sigmoid function, (c) tangent sigmoid function.

The activation values of the rectifier are shown in Fig.13(a). The saturation still exists in the first 10 epochs. But rectifier function can escape saturation state quicker than sigmoid function. But rectifier function needs more than 50 epochs to realize convergence. Fig.13(b) shows the activation values of sigmoid function. The output of the layer 2 fluctuates around its saturation limit 0 at the beginning of training, then jumps out of the saturation state after 10 epochs, and finally approaches to the stable state around 20 epochs. Fig.13(c) displays the outputs of tangent sigmoid function. The output of layer 2 completely saturated at the beginning of training, which prevents gradient return and hinders the deep layers from learning useful characteristics. The activation values of layer 2 escape saturation gradually and converge at around 35 epochs.

All errors decrease with the increase of training epochs. Among the three functions, sigmoid function performs best. Its error drops more quickly, and its error value is smaller than those of the other two. Therefore, sigmoid function is selected as the activation function in this paper.

D. SELECTION OF COST FUNCTION

A cost function $J(\theta)$ represents the difference between the predicted value and the actual one. The training process of a model is the process of optimizing the value of cost function. MSE and cross-entropy are the most commonly used cost functions.

1) MSE

The definition of MSE is as in (11). $f(\mathrm{x};\theta)$ is the predicted value of the neural network and y is the actual value. x is the input, and θ is the neural network parameter value. The gradient will become very small as output neuron reaches saturation. It is more suitable for regression rather than classification.

$$J(\theta) = \frac{1}{2} E_{x,y\sim\hat{p}data} \|y - f(x;\theta)\|^2 \tag{11}$$

2) CROSS-ENTROPY

If the maximum likelihood theorem is used to train network, it is necessary for a network to learn a complete probability distribution $p(y \mid x;\theta)$. With the maximum likelihood theorem, the cross entropy between the actual training data and the model prediction is taken as a cost function. Cross-entropy as the cost function eliminates the exponential effect in the output neuron and avoids saturation. The cost function is shown as in (12).

$$J(\theta) = -E_{x,y\sim\hat{p}data} \log p_{\mathrm{mod}\,el}(y|x) \tag{12}$$

Since the training procedure depends on how the gradient is updated, the cost functions can affect the performance of a neural network. To choose an effective cost functions for proposed SAE, the training errors of SAEs with same settings but different cost functions are listed in Table 2. The training errors are the average of 5 trained SAEs. It can be clearly concluded from the table that the SAEs perform better with cross-entropy. Cross-entropy is thus used as the cost function in the proposed SAE.

TABLE 2. Errors of trained network with different cost functions.

Cost Function	Cross entropy	MSE
Error	0.0090	0.0101

E. L^2 PARAMETER REGULARIZATION

In order to prevent overfitting and to improve the generalization ability of the neural network, L^2 regularization method is adopted; namely, the original cost function is directly added with a regularization term, as in (13):

$$J = J_0 + \frac{\lambda}{2n} \sum_{\omega} \omega^2 \tag{13}$$

where, J_0 is the original cost function, the rest is the regularization term of L^2, ω represents all of the weights that should be affected by a norm penalty, n denotes the size of the training set, and λ is the regulation coefficient. The recognition rate of network can be affected by the regulation coefficient. Fig.14 illustrates the relationship between the regularization coefficient and recognitions, which is the average result of 5 trained SAEs. Clearly, when regulation coefficient is selected around 0.02, a better recognition rate will be achieved. Thus, the regulation coefficient is selected to be 0.014.

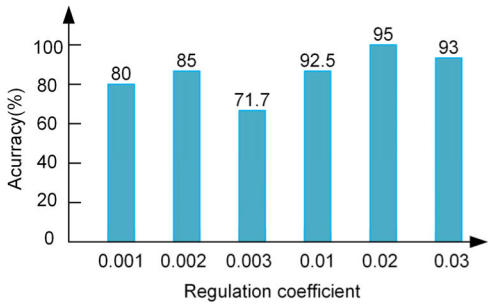

FIGURE 14. The relationship between the regulation coefficient λ and recognition rate.

V. SIMULATION AND COMPARISON

To demonstrate the performance of proposed SAE-based transient recognition, a two-terminal point-to-point VSC-HVDC transmission system, as shown in Fig.1 is modelled on the platform of PSCAD. VSC1 is the rectifier, and VSC2 is the inverter. The transmission capacity of this system is 500 MW. The DC bus voltage is ±500kV, and the midpoint of the supporting capacitor is grounded. The transmission line is modeled with the frequency-dependent model, and its total length is 250 km. The ground wire type is GL-70, and the transmission wire type is 4*LGJQ-300 [29].

For VSC-HVDC transmission, the protection unit must make the judgment in only a few milliseconds to avoid large-amplitude fault currents, which are hard to isolate. The protection algorithm will start when the fault current exceeds 1.2 times rated current value, and only a data segment of 2 milliseconds will be collected for transient recognition. All transient currents are normalized with equation (2) and used as the input of SAE. Since the sampling frequency is 100 kHz, the size of the raw input of SAE is 200 ($2\times10^{-3}\mathrm{s}\times100\times10^{3}\mathrm{samples/s}=200$).

A. RECOGNITION RESULTS

The three kinds of transients introduced in Section 2: single-pole-to-ground faults (GF), lightning faults (LF), and lightning disturbances (LD) are simulated to test the performance of proposed SAE. For each kind of transients, 200 samples are generated. Among them, 80% of samples are used for training (480 samples in total), and the rest 20% are used for testing (120 samples in total). According to the high voltage insulation and lightning protection [30], the negative lightning current waveform with 2.6μs rising time and 50μs decaying time is adopted in simulations. The peak current amplitude varies from 10kA to 30kA for LDs, and from 25 kA to 100kA for LFs. The grounding resistance, both for GFs and LFs, varies between 1Ω and 50Ω.

Table 3 shows the simulation settings and recognition results. All GFs are correctly recognized with only 2 millisecond post-fault current signals. But misjudgments occur when recognizing LDs and LFs. One LD is misjudged to be LF, while one LF is misjudged to be LD. However, the overall

TABLE 3. Recognition result of proposed SAE.

Type	Grounding resistance	Lighting current	Samples number	Recognitions	Misjudgment number
LD	/	10kA-30kA	40	97.5%	1 (LF)
LF	1Ω-50Ω	25kA-100kA	40	97.5%	1 (LD)
GF	1Ω-50Ω	/	40	100%	0

recognition rate of three transients is 98.3 %, which is effective enough when only 2 millisecond signals are used.

B. EFFECT OF NOISES

In practical applications, transient signals detected by measuring devices are often polluted by noises. There are a lot of noise sources, for example, operations of power electronic devices and measuring background noises. Measuring background noises are usually known as Gaussian white noises, and they are commonly encountered in online measurements. To test performance of proposed SAE-based transient recognition, Gaussian white noises with different signal-to-noise ratios (SNR) are thus added to the test samples, and the recognition results are shown in Table 4. All GFs can be correctly recognized, no matter how serious noise pollution is. When the SNR continues decreasing, more misjudgments occur. For example, 4 LDs are recognized to be LFs when the SNR is 60dB, and 6 LFs are classified to be LDs when SNR is 40dB. But when the SNR reaches 30 dB, the overall recognition rate is 95%, which is acceptable for most applications. So, the proposed SAE-based method is still effective in the presence of noises.

TABLE 4. Recognition results of proposed SAE with different SNRs.

Type of transients	SNR	60dB	50dB	40dB	30dB
LD	Recognitions	90%	90%	100%	100%
	Misjudgments	4 (LF)	4 (LF)	/	/
LF	Recognitions	100%	100%	85%	85%
	Misjudgments	/	/	6 (LD)	6 (LD)
GF	Recognitions	100%	100%	100%	100%
	Misjudgments	/	/	/	/
Recognitions		96.7%	96.7%	95%	95%

C. EFFECT OF CONDUCTORS

The transient characteristics will vary with the parameters of conductors, for instance, radius, resistance, material and, so on. The simulated transmission line model in PSCAD can only work as a kind of particular conductor as close as possible. But in practical HVDC systems, the conductors are different. It is necessary to discuss the performance of the proposed method on different conductors. To demonstrate the identification capability of the proposed method, transients collected from HVDC systems with four different conductors are analyzed. The parameters of the four conductors are listed in Table 5.

TABLE 5. Parameters of 4 conductors.

Type of conductor	Out radius	Total number of strands	Total number of outer strands	Strands radius	DC resistance
Stranded wire 1	0.0126m	37	18	0.0018	0.11195Ω/km
Stranded wire 2	0.0091m	19	7	0.0018	0.27966Ω/km
Stranded wire 3	0.01383m	61	24	0.00154	0.0854Ω/km
Stranded wire 4	0.0098m	37	18	0.0014	0.1889Ω/km

TABLE 6. Recognition results of proposed SAE with different conductors.

Type of transient	LD	LF	GF
Recognitions	97.5% (2LF)	97.5% (2LD)	100%

20 samples are generated for each kind of transient and each type of conductor. So, 240 transient samples (20 × 3 transients × 4 conductors=240) in total are collected and tested by the proposed SAE. The results are shown in Table 6. Only 2 LDs and 2 LFs are misjudged in 80 samples. The recognition results of the proposed SAE are unchanged when the line parameters change. The total recognition rate is 97.5% for LD and LF, and 100% for GF. It can be concluded the proposed SAE-based transient recognition is robust to the changes of the conductor, and it has great potential to be used in practical applications.

D. COMPARISONS

To demonstrate the effectiveness of the proposed SAE-based recognition method, the transients used in Section C are also recognized by three existing recognition methods: a practical one: differential integral method, a signal processing based one: wavelet transform method, and a traditional machine learning based one: neural network method.

1) DIFFERENTIAL INTEGRAL METHOD

In practical applications, simple calculations such as increment are used to discriminate faults and interferences. A type used one is the differential integral method [30], [31]. The differential of current measurement is integrated to obtain the current increment ΔI, whose definition is as in (14). Here, TD is the integral time constant; TI is the differential time constant; I_{dH} is the current input.

$$\Delta I = \Delta I_{(n-1)} + \frac{T_D}{T_I}(\Delta I_{dH(n)} - \Delta I_{dH(n-1)}) \tag{14}$$

Since interferences do not lead to the variation of system models, their increment is smaller than those of faults. A threshold is usually adopted, and 15 is chosen in this comparison according to the settings of VSC-HVDC model.

2) WAVELET TRANSFORM METHOD

Wavelet transform is the most widely used tool in analyzing the multi-resolution time-frequency spectrums of transient

signals [32]–[34]. Since grounding fault can be mathematically modeled by a step function, the energy of its current response should be concentrated in lower frequency bands while most energies of interferences are contained in the high frequency range. The distribution of the wavelet energy spectrum is often used in distinguishing faults and disturbances. Here, the energy ratio of a 5-layer wavelet energy spectrum is adopted in comparison when sampling frequency 100kHz is used. The definition of energy ratio is as in (15):

$$k = \frac{E_h}{E_l} \tag{15}$$

where, E_h represents the high-frequency energy of detail coefficients, E_l is the low-frequency energy of approximate coefficients. The parameter k is selected to be 0.6. If the energy ratio k is larger than 0.6, the transient is classified to be interference. Otherwise, it is regarded to be a fault.

3) NEURAL NETWORK METHOD

Back-propagation (BP) neural network is a classical machine learning method and was commonly used in transient recognition of power systems. In order to eliminate the influences from architecture and functions, the parameters of the BP network are set as the same as the proposed SAE. The BP neural network is also set as a 4-layer network: 1 input layer, 2 hidden layers and 1 output layer, and its size is 200-135-85-3, which is the same as the proposed SAE. The sigmoid function is adopted as an activation function for both the hidden layers and the output layer. Gradient descent back-propagation is used as the training algorithm. All the same samples used in Section C are used to train and test the designed BP network.

The performance of traditional methods and proposed SAE based methods are illustrated in Table 7. The duration of input suggests the data length of transient signals. The computation duration is the time that each method needs to complete the calculation on a computer with 2.2GHz-CPU. For all these methods, 80 transient samples composed of LD (40 samples) and LF (40 samples) are recognized.

TABLE 7. Comparisons between different recognition methods.

Recognition	Duration of input	Computation duration	Recognition rate (misjudgments)
Differential integral	10 ms	0.307 ms	65% (25)
Wavelet transform	5 ms	0.46 ms	78.75% (17)
P neural network	2 ms	0.247 ms	87.5% (10)
Proposed SAE-based	2 ms	0.241 ms	97.5% (2)

In order to avoid the transient influences from grounding capacitances, integral methods usually need to wait until the decay of oscillations. Among all these methods, the differential integral method requires the longest transient inputs. Its total processing time is more than 10 times that of other methods. Wavelet transform includes a lot of convolution calculations, which requires more time. So, the computation duration of the wavelet transform method is the longest. The training of neural network which require a large amount of computation are performed offline, while the online recognition is very quickly. As shown in Table 7, the online computation durations of neural networks are a bit more than 0.2 milliseconds.

Besides the advantages in the aspect of processing time, the recognition results of the proposed method are much better than the others. The machine learning-based methods show better performance than thresholding-based ones. With more stable feature representations, the SAE-based method generates higher recognition rates than the supervised learning network.

VI. CONCLUSION

In this paper, a SAE with the proper design is used to classify the transient currents of LD, LF and GF, which are commonly encountered in overhead HVDC transmissions. The architecture and functions of SAE are discussed and carefully selected. The recognition performance of the proposed method is tested by various scenarios. The test results demonstrate SAE is effective in identifying different transient currents and is immune to the effect of grounding resistance, locations, type of conductors, and noises. Compared with traditional methods, SAE performs better in recognition rates. Based on the analysis and discussions mentioned in this paper, SAE reveals excellent performance in feature extraction and transient recognition and is highly potential to be used in practical applications.

REFERENCES

[1] N. Flourentzou, V. Agelidis, and G. Demetriades, "VSC-based HVDC power transmission systems: An overview," *IEEE Trans. Power Electron.*, vol. 24, no. 3, pp. 592–602, Mar. 2009.
[2] Z. Yao, Q. Zhang, P. Chen, and Q. Zhao, "Research on fault diagnosis for MMC-HVDC systems," *Protection Control Mod. Power Syst.*, vol. 1, no. 1, Jun. 2016, Art. no. 8.
[3] S. Li, Z. Wu, and J. Huang, "Power flow modelling to UHVDC line and its hierarchical connection mode," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 7, pp. 1554–1564, Apr. 2018.
[4] T. Wu and W. U. Bajwa, "Revisiting robustness of the union-of-subspaces model for data-adaptive learning of nonlinear signal models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 3390–3394.
[5] G. Luo, Q. Lin, L. Zhou, and J. He, "Recognition of traveling surges in HVDC with wavelet entropy," *Entropy*, vol. 19, no. 5, May 2017, Art. no. 184.
[6] J. He, Y. Luo, M. Li, Y. Zhang, Y. Xu, Q. Zhang, and G. Luo, "A high-performance and economical multi-port hybrid DC circuit breaker," *IEEE Trans. Ind. Electron.*, to be published.
[7] S. R. Mohanty, N. Kishor, P. K. Ray, and J. P. S. Catalao, "Comparative study of advanced signal processing techniques for islanding detection in a hybrid distributed generation system," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 122–131, Jan. 2015.
[8] B. Patel, "A new FDOST entropy based intelligent digital relaying for detection, classification and localization of faults on the hybrid transmission line," *Electr. Power Syst. Res.*, vol. 157, pp. 39–47, Apr. 2018.
[9] N. Roy and K. Bhattacharya, "Detection, classification, and estimation of fault location on an overhead transmission line using S-transform and neural network," *Electr. Power Compon. Syst.*, vol. 43, no. 4, pp. 461–472, Feb. 2015.

[10] Y. Kong, T. Y. Wang, and F. L. Chu, "Meshing frequency modulation assisted empirical wavelet transform for fault diagnosis of wind turbine planetary ring gear," *Renew. Energy*, vol. 132, pp. 1373–1388, Mar. 2019.
[11] S. Wang, X. Meng, Y. Yin, Y. Wang, X. Yang, X. Zhang, X. Peng, W. He, G. Dong, and H. Chen, "Optical image watermarking based on singular value decomposition ghost imaging and lifting wavelet transform," *Opt. Lasers Eng.*, vol. 114, pp. 76–82, Mar. 2019.
[12] F. Deng, X. Li, and X. Zeng, "Single-ended travelling wave protection algorithm based on full waveform in the time and frequency domains," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 15, pp. 3680–3691, Aug. 2018.
[13] J. Chen, Y. Dou, Y. Li, J. Li, and G. Li, "A transient fault recognition method for an AC-DC hybrid transmission system based on MMC information fusion," *Energies*, vol. 10, no. 1, p. 23, Dec. 2016.
[14] L. Zhang and Z.-Q. Lang, "Wavelet energy transmissibility function and its application to wind turbine bearing condition monitoring," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1833–1843, Oct. 2018.
[15] S. Lin, S. Gao, Z. He, and Y. Deng, "A pilot directional protection for HVDC transmission line based on relative entropy of wavelet energy," *Entropy*, vol. 17, no. 12, pp. 5257–5273, Jul. 2015.
[16] F. B. Costa, B. A. Souza, N. S. D. Brito, J. A. C. B. Silva, and W. C. Santos, "Real-time detection of transients induced by high-impedance faults based on the boundary wavelet transform," *IEEE Trans. Ind. Appl.*, vol. 51, no. 6, pp. 5312–5323, Nov. 2015.
[17] T. Gush, S. B. A. Bukhari, R. Haider, S. Admasie, Y.-S. Oh, G.-J. Cho, and C.-H. Kim, "Fault detection and location in a microgrid using mathematical morphology and recursive least square methods," *Int. J. Elect. Power Energy Syst.*, vol. 102, pp. 324–331, Nov. 2018.
[18] M. A. Barik, A. Gargoom, M. A. Mahmud, M. E. Haque, H. Al-Khalidi, and A. M. T. Oo, "A decentralized fault detection technique for detecting single phase to ground faults in power distribution systems with resonant grounding," *IEEE Trans. Power Del.*, vol. 33, no. 5, pp. 2462–2473, Oct. 2018.
[19] T. Wu, G. Polatkan, D. Steel, W. Brown, I. Daubechies, and R. Calderbank, "Painting analysis using wavelets and probabilistic topic models," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013.
[20] E. Ghazizadeh, D. Abbasi-Moghadam, and H. Nezamabadi-Pour, "An enhanced two-phase SVM algorithm for cooperative spectrum sensing in cognitive radio networks," *Int. J. Commun. Syst.*, vol. 32, no. 2, p. e3856, Jan. 2019.
[21] A. Tsoupos and V. Khadkikar, "A novel SVM technique with enhanced output voltage quality for indirect matrix converters," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 832–841, Feb. 2019.
[22] M. Oleskovicz, D. V. Coury, O. D. Felho, W. F. Usida, A. A. Carneiro, and L. R. Pires, "Power quality analysis applying a hybrid methodology with wavelet transforms and neural networks," *Int. J. Elect. Power Energy Syst.*, vol. 31, no. 5, pp. 206–212, Jun. 2009.
[23] K. Silva, B. Souza, and N. Brito, "Fault detection and classification in transmission lines based on wavelet transform and ANN," *IEEE Trans. Power Del.*, vol. 21, no. 4, pp. 2058–2063, Oct. 2006.
[24] S. Khan and S. Bhowmick, "A comprehensive power-flow model of multi-terminal PWM based VSC-HVDC systems with DC voltage droop control," *Int. J. Elect. Power Energy Syst.*, vol. 102, pp. 71–83, Nov. 2018.
[25] G. Luo, C. Yao, Y. Liu, Y. Tan, J. He, and K. Wang, "Stacked auto-encoder based fault location in VSC-HVDC," *IEEE Access*, vol. 6, pp. 33216–33224, 2018.
[26] I. M. Dudurych, T. J. Gallagher, J. Corbett, and M. V. Escudero, "EMTP analysis of the lightning performance of a HV transmission line," *IEE Proc.-Gener., Transmiss. Distrib.*, vol. 150, no. 4, pp. 501–506, Jul. 2003.
[27] D. Rajicic and M. Todorovski, "Two-component current waveform for lightning simulation," *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 5, pp. 1062–1069, Oct. 2015.
[28] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.
[29] Z. Liu, *Typical Design of Transmission and Transformation Project of State Grid Corporation of China: 500kV Transmission Line*, 1st ed. Beijing, China: China Electric Power Press, 2007.
[30] F. Kiessling, P. Nefzger, J. F. Nolasco, and U. Kaintzyk, "Lightning protection," in *Overhead Power Lines: Planning, Design*. Berlin, Germany: Springer, 2003, pp. 99–113.
[31] F. Kong, Z. Hao, S. Zhang, and B. Zhang, "Development of a novel protection device for bipolar HVDC transmission lines," *IEEE Trans. Power Del.*, vol. 29, no. 5, pp. 2270–2278, Oct. 2014.
[32] J. Liu, N. Tai, and C. Fan, "Transient-voltage-based protection scheme for DC line faults in the multiterminal VSC-HVDC system," *IEEE Trans. Power Del.*, vol. 32, no. 3, pp. 1483–1494, Jun. 2017.
[33] X. Zheng, T. Nengling, Y. Guangliang, and D. Haoyin, "A transient protection scheme for HVDC transmission line," *IEEE Trans. Power Del.*, vol. 27, no. 2, pp. 718–724, Apr. 2012.
[34] C. Fan, K. Li, W. Chan, and W. Yu, "Study of protection scheme for transmission line based on wavelet transient energy," *Int. J. Elect. Power Energy Syst.*, vol. 28, no. 7, pp. 459–470, Sep. 2006.