Data-driven Fault Detection and Classification for MTDC Systems by Integrating HCTSA and Softmax Regression

Jiapeng Li, Guobing Song, *Senior Member, IEEE*, Jifei Yan, Yujun Li, *Member, IEEE*, Zhao Xu, *Senior Member, IEEE*

***Abstract*—The requirement of fast fault isolation poses a great challenge to the safe operation of multi-terminal direct current (MTDC) systems. In order to make a better tradeoff between the speed and reliability of the protection scheme, it is imperative to mine more valuable information from fault transient signals. This paper puts forward a data-driven framework capable of digging out and synthesizing multi-dimensional features to achieve fast and reliable DC fault detection and classification in MTDC systems. Highly comparative time-series analysis (HCTSA) is first adopted to extract extensive features with clear physical interpretations from fault current waveforms, and a few features valuable to fault identification are then selected utilizing the greedy forward search. Based on the reduced features, a softmax regression classifier (SRC) is further proposed to calculate the probability of each fault category with a relatively minor on-line computational burden. Numerical simulations carried out in PSCAD/EMTDC have demonstrated the proposed approach is effective under different fault conditions, robust against noise corruptions as well as abnormal samplings, and replicable in various DC grids. In addition, comprehensive comparison studies with conventional derivative-based protection methods and some typical artificial intelligence based (AI-based) methods have been conducted. It is verified that the proposed method has the advantages of higher fault identification accuracy over conventional protections and shallow structure AI-based methods, better interpretability as well as lower on-line computing complexity over the deep architecture AI-based approaches.**

***Index Terms*—Fault detection and classification, MTDC system, artificial intelligence, representation learning.**

I. Introduction

In the last decade, the voltage source converter (VSC) based MTDC system has been vastly adopted in the interconnection of asynchronous AC areas and integration of large bulk of renewable energies due to its high control flexibility and high reliability [1]-[3]. The emergence of the modular multilevel converter (MMC) reduces the switching losses of IGBTs and the harmonic level at the DC side [4]-[6], making the VSC-MTDC network a more promising solution for future power transmission and distribution. However, VSC-based converters are vulnerable to DC fault current surges, and to prevent the outage of the overall system, fault isolation should be achieved within several milliseconds after DC fault [7]. Therefore, fast and reliable fault detection has become a major challenge for the safe operation of MTDC systems.

A. Related works

The non-unit protection does not involve the communication delay, and thereby has been regarded as a more attractive scheme for fast fault isolation in MTDC systems than the differential approach. Single-end traveling wave protections that detect the reflection and refraction of traveling waves at the fault point or line boundaries are widely utilized in practical engineering [8]-[10]. To quantitatively describe the sudden change of the system status during DC fault, the rate of change of current and voltage (ROCOC and ROCOV) are respectively put forward in [8] and [9]. However, traveling wave protections suffer from the poor ability to tolerate fault resistance and the requirement of high-frequency sampling. To address these issues, transient protections that utilize the changes of stored energy in inductors and capacitors during fault are proposed in [11]-[13]. Reference [11] reports a transient average current based fault detection method for radial MTDC systems by overlooking the traveling wave process. In addition, numerous digital signal processing techniques such as Fast Fourier Transform (FFT) [12], Short Time Fourier Transform (STFT) [13], and Wavelet Transform (WT) [14]-[15], have been introduced into DC grid protection to describe the complex fault transient dynamics. Unfortunately, due to the lack of clear and accurate analytical expressions of transient short circuit current, it is difficult to determine which time-series analysis method is more suitable and effective for fault identification in a specified network. Thus, a rational feature selection from fault transient signals remains to be further explored for DC fault detection and classification in MTDC systems.

Recently, several artificial intelligence (AI) algorithms such as Principal Component Analysis [16], Naïve Bayes [17], K-means cluster [18], Support Vector Machine (SVM) [19], and Artificial Neural Network (ANN) [20]-[24] are utilized in DC grid protection. Compared with conventional approaches, AI algorithms make it possible to dig out and synthesize multi-dimensional features from fault transient signals, which may enhance the comprehensive performance of the protection scheme. Broadly speaking, AI-based protections can be classified into two categories: shallow structure methods and deep structure methods. The former class [16]-[23] normally requires proper features as the input of a specified type of fault

This work is supported by the National Natural Science Foundation of China (U1766209, 51807150). *(Corresponding author: Guobing Song.)*

J. Li and Y. Li are with the School of Electrical Engineering, Xi'an Jiaotong University, Shaan Xi, China. They are also with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong. (email: wennuanfrank@gmail.com, yujunlizju@gmail.com).

G. Song and J. Yan are with the School of Electrical Engineering, Xi'an Jiaotong University, Shaan Xi, China. (email: song.gb@mail.xjtu.edu.cn, jifeiyanhl@gmail.com).

Z. Xu is with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong. (email: eezhaoxu@polyu.edu.hk).

classifier, and then the parameters of the classification model can be tuned to achieve high-accuracy fault identification. For instance, in [19]-[22], FFT and WT are separately adopted to extract useful features from fault transient waveforms, and subsequently, AI algorithms such as SVM, ANN, and Fuzzy inference can be used to comprehensively analyze these features and predict the fault category. Still, manual feature selection cannot guarantee the extracted features are effective for fault detection. In MMC-MTDC systems, line boundaries are normally weak due to the small inductance of DC reactors and equivalent capacitance of converts, and thus more careful feature selection is in demand to capture the subtle differences between some internal and external faults. By contrast, the latter class [24] embeds the feature extraction into the deep neural network, so that the original transient waveforms can be directly inputted towards the data model, and useful features can be automatedly determined in the training process. In [24], a branch-structure convolutional neural network (CNN) is designed for fault identification in MMC-MTDC systems. Nevertheless, the deep learning model includes a large number of parameters, which significantly increases the computing complexity. Moreover, the extracted features in this end-to-end approach lack clear physical interpretations, making it hard to understand the operation mechanism of the deep learning scheme and ensure its effectiveness under various conditions.

B. Main contribution

The contribution of this paper mainly lies in two folds. For one thing, to enhance the accuracy and reliability of fault detection in MTDC networks, especially those with weak line boundaries, a data-driven protection framework is proposed, which can automatedly extract valuable features from fault transient signals and synthetically discriminate the fault category in high accuracy. For another, the proposed method retains good interpretability owing to clear physical meanings of the extracted features and solid mathematical implications of SRC. Compared with existing methods, the proposed one makes a better trade-off among fault identification accuracy, on-line computing complexity, interpretability, as well as robustness against noise corruptions and abnormal samplings.

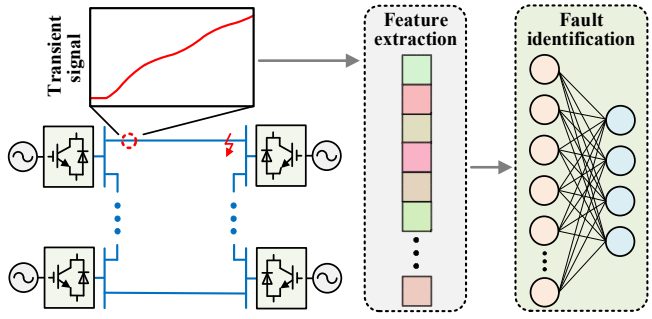


Fig. 1. Overall scheme of the proposed approach of DC fault detection and classification.

II. HCTSA-based Feature Extraction From DC Current Waveform

In the proposed approach of fault detection and classification for MTDC systems, feature extraction is firstly conducted to gain valuable information from the transient DC fault current waveform, and fault identification can be subsequently achieved via synthesizing the extracted features as illustrated by Fig. 1. In order to select features that are effective for the fault classification task, and simultaneously with clear physical interpretation, a state-of-the-art time-series analysis method, namely HCTSA, is adopted in this paper. The main procedures of the HCTSA-based feature extraction are depicted as follows.

A. Unified feature-based representation of current waveform

HCTSA is a supervised learning algorithm that specializes in the automated extraction of useful and interpretable features from the studied time series. For utilizing HCTSA, the training data should be first defined.

The input matrix $I \in \mathbb{R}^{M\times L}$ comprises quantities of DC current waveforms obtained under various conditions, which can be expressed as follows,

$$I=\begin{bmatrix} I^{(1)} \\ I^{(2)} \\ \vdots \\ I^{(M)} \end{bmatrix}=\begin{bmatrix} i_1^{(1)} & i_2^{(1)} & \cdots & i_L^{(1)} \\ i_1^{(2)} & i_2^{(2)} & \cdots & i_L^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ i_1^{(M)} & i_2^{(M)} & \cdots & i_L^{(M)} \end{bmatrix} \tag{1}$$

where each row of I noted as $I^{(i)}$ (i = 1, 2, ⋯, M) is a specified fault component current waveform measured by the positive-pole or negative-pole relaying, and M is the total number of samples in I. Further, each $I^{(i)}$ consists of L sequential sampling points denoted by $i_l^{(i)}$ (l = 1, 2, ⋯, L), and the sequence length L satisfies the following equation,

$$L = T \cdot f_s \tag{2}$$

where T and f_s are the length of the sampling window and the sampling frequency, respectively.

The output matrix $Y \in \mathbb{R}^{M\times 1}$ can be written as follows,

$$Y=\left[y^{(1)}, y^{(2)}, \cdots, y^{(M)} \right]^{\mathrm{T}} \tag{3}$$

where $y^{(i)}$ is an integer denoting the corresponding class label of $I^{(i)}$, and the superscript T is matrix transpose. For the feature extraction task, $y^{(i)}$ values in {1, 2, 3, 4}, which individually represent the pole-to-pole fault current, fault pole current under single-pole-to-ground fault, non-fault pole current under single-pole-to-ground fault, and current under disturbance.

For facilitating the analysis, an operation library offered by HCTSA is utilized to convert the original time-varying waveform $I^{(i)}$ into a static feature vector. More specifically, HCTSA totally selects 7749 kinds of time-series properties across the scientific disciplines, including basic statistics of the distribution, linear correlations, stationarity, wavelet methods, and so on. It is demonstrated in [25] that these features can handle almost all time-series analysis problems in the real world. To obtain the feature-based representation, the studied current waveform should be calculated by each time-series analysis function in the operation library separately,

$$\begin{aligned} G^{(i)} &= \Phi\left(I^{(i)}\right)=\left[\varphi_1\left(I^{(i)}\right),\varphi_2\left(I^{(i)}\right),\cdots,\varphi_F\left(I^{(i)}\right)\right] \\ &=\left[g_1^{(i)}, g_2^{(i)}, \cdots, g_F^{(i)}\right] \end{aligned} \tag{4}$$

where $G^{(i)} \in \mathbb{R}^{1\times F}$ is the feature vector corresponding to $I^{(i)}$;

$\Phi = \{\varphi_1, \varphi_2, \cdots, \varphi_F\}$ is the operation set (library) and φ_f (f = 1, 2, ⋯, F) is a time-series function that projects a time series of $1\times L$ to a scalar $g_f^{(i)}$. It should be noted that some operations may output a complex number or non-positive data for fitting positive-only distribution. These inappropriate operations are filtered out from the original operation library, and thereby the number of appropriate operations noted by F is less than 7749. Normally, F is as large as several thousand, so that $G^{(i)}$ can be referred to as the *extensive feature vector*.

Moreover, for guaranteeing each element of the feature vector could have an equal degree of effect on the fault classification task, the feature vector should be normalized by a scaled robust sigmoid function [26] given below,

$$\begin{cases} \widetilde{G^{(i)}} = \left[\widetilde{g_1^{(i)}}, \widetilde{g_2^{(i)}}, \cdots, \widetilde{g_F^{(i)}}\right] \\ \widetilde{g_f^{(i)}} = \left(1+\exp\left(-\dfrac{g_f^{(i)} - m_f}{1.35 r_f}\right)\right)^{-1}, f = 1,2,\cdots,F \end{cases} \tag{5}$$

where the tilde ~ on the head of symbols represents the normalized values; m_f and r_f are the median and interquartile range of the set $\left\{g_f^{(1)}, g_f^{(2)}, \cdots, g_f^{(M)}\right\}$, respectively.

After the normalization, each element of the feature vector has been scaled into the range of [0, 1], and the unified featured-based representation of DC current waveform is fully formulated.

B. Feature selection based on greedy policy

The extensive feature vector contains plenty of interpretable features, which, however, cannot be directly used in practical engineering due to its high dimensionality. On the one hand, from the perspective of machine learning, it is actually infeasible to collect sufficient training samples and calculate the distance in a high-dimension space, which is referred to as the *curse of dimensionality*. On the other hand, the extensive feature vector inevitably involves a large number of intermediate operations during the fault classification, thereby tremendously increasing the on-line computational burden. Considering the extensive features are highly redundant, it is necessary and feasible to select a few features useful to fault identification for the dimension reduction purpose.

However, selecting the optimal F_r time-series operations (features) as few as possible from the normalized operation set noted as $\widetilde{\Phi}$ for achieving the best classification performance is actually a complicated combinatorial optimization problem, and it is practically infeasible to find the global optimal solution using the exhaustive search. To address this issue, a commonly used approximation algorithm, namely, *greed policy* is adopted here and a local optimum can be figured out instead.

The detailed algorithm of greedy forward feature selection is stated as follows. Given the classifier, the misclassification rates by using all individual normalized feature operations $\widetilde{\varphi_f}$ are calculated, and $\widetilde{\psi_1}$, which is the operation with the minimal misclassification rate, is chosen as the first operation in the reduced set $\widetilde{\Psi}$. Next, the misclassification rates via utilizing all individual operations in combination with $\widetilde{\psi_1}$ are calculated, and the one with the lowest misclassification rate is chosen as $\widetilde{\psi_2}$. In this way, features to be added should be irrelevant from the existing ones, since the misclassification rate is more likely to decrease when additional features can complement more useful and different information. Repeat the procedure and choose the operation that conduces the maximum improvement of the classification performance in each iteration until the termination criterion is satisfied, yielding the final reduced normalized operation set denoted by $\widetilde{\Psi} = \left\{\widetilde{\psi_1}, \widetilde{\psi_2}, \cdots, \widetilde{\psi_{F_r}}\right\}$. The stop criterion is set as the point when the misclassification rate drops to zero, or its decrease by adding one additional operation is less than a sufficiently small number.

Initialization and Input:
Determine the classification learning algorithm $\mathcal{L}$.
Load the training data: the input I and the target output Y.
Set $\widetilde{\Phi} = \left\{\widetilde{\varphi_1}, \widetilde{\varphi_2}, \cdots, \widetilde{\varphi_F}\right\}$, $\widetilde{\Psi} = \varnothing$.

Process:
1: **Repeat**:
2: **For** each $\widetilde{\varphi_f}$ in $\widetilde{\Phi}$ **do**:
3: Operation set update: $O_f = \widetilde{\Psi} \cup \widetilde{\varphi_f}$.
4: Training set update: $D_f = \left\{\left(O_f(I), Y\right)\right\}$.
5: Performance assessment:
calculate the misclassification rate of algorithm $\mathcal{L}$ in the training set D_f, noted as ε_f.
6: **End for**
7: Optimal choice: set φ^* as the $\widetilde{\varphi_f}$ with the minimal ε_f.
8: Operation set update: $\widetilde{\Phi} \leftarrow \complement_{\widetilde{\Phi}}\left(\varphi^*\right)$, $\widetilde{\Psi} \leftarrow \widetilde{\Psi} \cup \varphi^*$.
9: **Until** the termination criterion is satisfied.

Output: The reduced feature operation set $\widetilde{\Psi}$.

Algorithm 1. Greedy forward search algorithm for selecting features.

The HCTSA-based feature extraction provides a unified and reduced feature vector representation of transient DC current waveform, which retains the interpretability of features and simultaneously ensures good performance of the followed fault identification.

III. Fault Detection and Classification Method

A. Start-up criterion

In order to reduce the risk of protection maloperations as well as the on-line computational burden, the fault detection and classification should be initialized only if the start-up criterion is satisfied. In this paper, the start-up element is chosen as the transient average increment of current, which is defined as the difference between the cumulative sum of the current sampling value in two consecutive intervals. The discrete form of transient average increment of current marked as $\overline{\Delta i(t)}$ can be written as follows,

$$\overline{\Delta i(t)} = \frac{1}{L_s^2}\sum_{l=1}^{L_s}\left(i\left(t+l-L_s\right) - i\left(t+l-2L_s\right)\right) \tag{6}$$

where L_s is the number of sampling points in each interval and

it is set as 5 in this paper. The main advantage of the transient-average-increment-based start-up criterion is the ability to filter out some of the measurement noise as well as low computational complexity.

Based on (6), the fault detection is activated when either of the transient average increment of positive or negative current exceeds the threshold value denoted by Δi_{set},

$$\max\left\{\overline{\Delta i_p(t)},\ \overline{\Delta i_n(t)}\right\} > \Delta i_{set} \tag{7}$$

The threshold of the start-up element should exceed the maximum transient average increment of current during the change of load current, which is denoted by $\overline{\Delta i_L}$. Meanwhile, it should be less than the transient average increment of current under the remotest internal fault with the maximum tolerant fault resistance, marked as $\overline{\Delta i_R}$. Therefore, the threshold value can be determined by the following inequality constraint,

$$K_{r1}\overline{\Delta i_L} < \Delta i_{set} < K_{r2}\overline{\Delta i_R} \tag{8}$$

where K_{r1} and K_{r2} are two reliability coefficients. K_{r1} should be larger than 1, and K_{r2} should be in the range of (0, 1).

B. Softmax regression classifier

Softmax regression is the extension of *linear regression* to multi-classification tasks, which has been proven simple but powerful by large amounts of industrial practice. In this part, a *softmax regression classifier* (SRC) is trained for identifying the fault type via the features of transient DC current waveform.

The input of SRC is the combination of reduced feature vectors of positive and negative fault component current, which is denoted by $X^{(i)} \in \mathbb{R}^{1\times 2F_r}$. The output of SRC marked as $y^{(i)}$ values in {1, 2, 3, 4}, representing the positive-to-negative fault (P2N), the positive-to-ground fault (P2G), the negative-to-ground fault (N2G), and the disturbance, individually. The SRC utilizes the following equation known as *softmax regression function* to model the posterior probability of $y^{(i)}$ under the given observation $X^{(i)}$,

$$\begin{bmatrix} P\left(y^{(i)}=1\middle|X^{(i)};W,B\right) \\ P\left(y^{(i)}=2\middle|X^{(i)};W,B\right) \\ \vdots \\ P\left(y^{(i)}=K\middle|X^{(i)};W,B\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K}\exp\left(W_j^{\,T}X^{(i)}+b_j\right)} \begin{bmatrix} \exp\left(W_1^{\,T}X^{(i)}+b_1\right) \\ \exp\left(W_2^{\,T}X^{(i)}+b_2\right) \\ \vdots \\ \exp\left(W_K^{\,T}X^{(i)}+b_K\right) \end{bmatrix} \tag{9}$$

where K is the number of categories and it is 4 in this fault classification task; $W \in \mathbb{R}^{K\times 2F_r}$ and $B \in \mathbb{R}^{K\times 1}$ are the weight matrix and bias matrix, respectively,

$$W = \begin{bmatrix} W_1^{\,T} \\ W_2^{\,T} \\ \vdots \\ W_K^{\,T} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1(2F_r)} \\ w_{21} & w_{22} & \cdots & w_{2(2F_r)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1} & w_{K2} & \cdots & w_{K(2F_r)} \end{bmatrix},\ B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix} \tag{10}$$

Generally, the value of W and B can be determined by the classical *maximum likelihood estimation* (MLE) [27]. Based on (9), the likelihood function is calculated as follows,

$$\begin{aligned} L(W,B) &= \prod_{i=1}^{N} P\left(y^{(i)}\middle|X^{(i)};W,B\right) \\ &= \prod_{i=1}^{N}\prod_{j=1}^{K} P\left(y^{(i)}=j\middle|X^{(i)};W,B\right)^{\mathbb{I}\left\{y^{(i)}=j\right\}} \end{aligned} \tag{11}$$

where N is the number of samples in the training set; $\mathbb{I}\{\cdot\}$ is the index function satisfying the following equation,

$$\mathbb{I}\left\{y^{(i)}=j\right\} = \begin{cases} 1, & y^{(i)}=j \\ 0, & y^{(i)}\neq j \end{cases} \tag{12}$$

The objective of MLE is to find the parameters W and B that maximize the likelihood function. For facilitating the calculation, the focus should be put on maximize the log-likelihood function expressed as follows,

$$\begin{aligned} \ln L(W,B) &= \ln\prod_{i=1}^{N}\prod_{j=1}^{K} P\left(y^{(i)}=j\middle|X^{(i)};W,B\right)^{\mathbb{I}\left\{y^{(i)}=j\right\}} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{K}\mathbb{I}\left\{y^{(i)}=j\right\}\ln\frac{\exp\left(W_j^{\,T}X^{(i)}+b_j\right)}{\sum_{k=1}^{K}\exp\left(W_k^{\,T}X^{(i)}+b_k\right)} \end{aligned} \tag{13}$$

Based on (13), the loss function for the training of SRC can be chosen as follows,

$$J(W,B) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}\mathbb{I}\left\{y^{(i)}=j\right\}\ln\frac{\exp\left(W_j^{\,T}X^{(i)}+b_j\right)}{\sum_{k=1}^{K}\exp\left(W_k^{\,T}X^{(i)}+b_k\right)} \tag{14}$$

Correspondingly, the estimations of W and B respectively noted as $\widehat{W}$ and $\widehat{B}$ can be obtained by solving the minimization problem below,

$$\max_{W,B}\ln L(W,B) \Leftrightarrow \min_{W,B} J(W,B) \tag{15}$$

Gradient descent, as a commonly used optimization algorithm, is adopted here to train the parameters. The update rule in each iteration can be expressed as follows,

$$\begin{cases} W_j \leftarrow W_j - \eta\cdot\nabla_{W_j}J(W,B) \\ b_j \leftarrow b_j - \eta\cdot\nabla_{b_j}J(W,B) \end{cases} \quad (j=1,2,\cdots,K) \tag{16}$$

where η represents the learning rate, and ∇ is the gradient operator. Based on (14), the gradient of $J(W, B)$ with respect to W_j and b_j can be directly calculated as follows,

$$\begin{cases} \nabla_{W_j}J(W,B) = -\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{I}\left\{y^{(i)}=j\right\} - \frac{\exp\left(W_j^{\,T}X^{(i)}+b_j\right)}{\sum_{k=1}^{K}\exp\left(W_k^{\,T}X^{(i)}+b_k\right)}\right)X^{(i)} \\ \nabla_{b_j}J(W,B) = -\frac{1}{N}\sum_{i=1}^{N}\left(\mathbb{I}\left\{y^{(i)}=j\right\} - \frac{\exp\left(W_j^{\,T}X^{(i)}+b_j\right)}{\sum_{k=1}^{K}\exp\left(W_k^{\,T}X^{(i)}+b_k\right)}\right) \end{cases} \tag{17}$$

SRC actually builds a probabilistic model between the selected features and all possible fault categories. Once the off-line training is completed, the trained SRC can be directly utilized in on-line fault detection and classification. For a new feature-based input denoted by $X^{(*)}$, the trained SRC will output the category $y^{(*)}$ with the maximum posterior probability, as written in the following equation,

$$y^{(*)} = \arg\max_{j} P\left(y^{(*)}=j\middle|X^{(*)};\widehat{W},\widehat{B}\right) \tag{18}$$

C. Process of the proposed fault detection and classification scheme

The overall framework of the proposed fault detection and

classification scheme for MTDC systems is illustrated in Fig. 2. Generally speaking, the scheme can be implemented into two phases: the off-line training and the on-line application.

In the first phase, plenty of training data under various conditions (different fault distances, resistances, and types) should be prepared for the following feature selection and SRC training. Then, HCTSA is carried out to extract thousands of features from the original current waveforms and the greedy forward search is utilized to select a small number of key features according to their contribution to the improvement of fault classification rate. Finally, the SRC for fault identification is constructed to build the probability relation model between the selected features and the fault categories. Based on the *maximum likelihood principle*, the objective function is determined and the parameters of the SRC can be optimized using the gradient descent algorithm.

In the second phase, the transient average increment of real-time sampling current is updated by the start-up element that detects any possible DC fault. Once the start-up criterion is satisfied, only a few selected features of the observed current are calculated. By combining the reduced feature-based representations of positive and negative current, the input vector of the SRC can be formed. Subsequently, the trained SRC will cope with the input vector to obtain the posterior probabilities of all individual fault categories. Lastly, the fault category with the maximum probability, which shows whether there is a fault as well as the fault type, is outputted by the SRC so that DC circuit breakers (CBs) can operate accordingly.

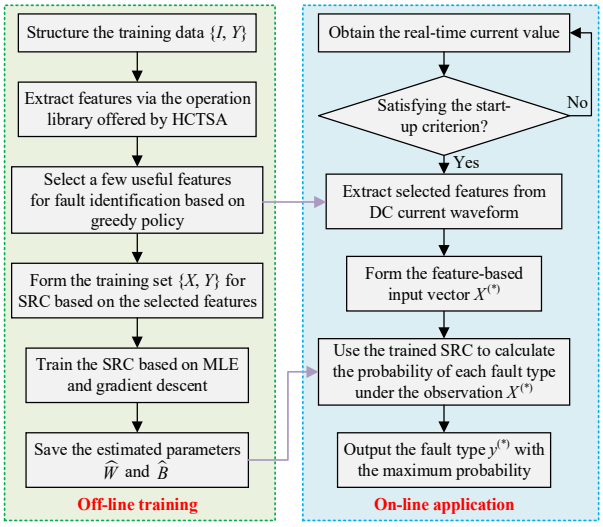


Fig. 2. Overall flow chat of the proposed fault detection and classification scheme for MTDC systems.

IV. Validations and Study Results

A. Introduction of the test system

In this section, a four-terminal MTDC system as shown in Fig. 3 is built in PSCAD/EMTDC to verify the effectiveness of the proposed fault detection and classification scheme. The test system is a symmetrical monopolar half-bridge MMC-based system that is commonly used in engineering due to the economical construction cost, where the AC-side neutral-points of converter transformers are non-directly grounded. Four 51-level MMC stations with the detailed switching model are interconnected by the ±200 kV DC power transmission cables with the frequency-dependent (phase) representation. To suppress DC fault current, supplementary reactors of 50 mH are implemented at both ends of each cable. As depicted in Fig. 3, the supplementary reactor and the hybrid DC CB [8] located at the *m*-side of Cable *mn* are denoted by L_{Tmn} and B_{mn}, individually. The MMCs adopt the master-slave control that MMC_2 regulates its reactive power and the DC system voltage, while MMC_1, MMC_3, and MMC_4 regulate their active and reactive power separately. The length of the time window is selected as 2 ms to utilize as much information as possible and simultaneously satisfy the requirement of fast fault isolation. According to (8), the threshold of the start-up element is set as 0.145 kA/ms. Other related parameters of the four-terminal test system are provided in Table I.

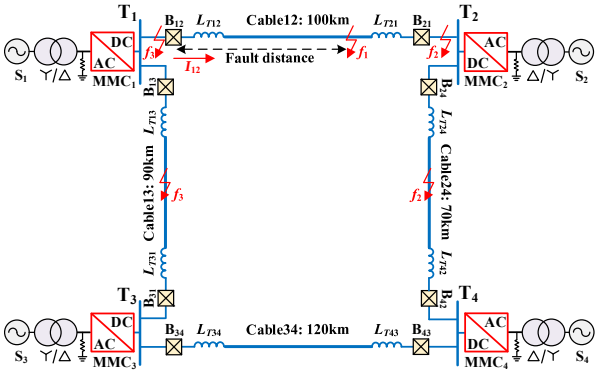


Fig. 3. Outline of the four-terminal VSC-MTDC test system.

TABLE I Concerned Parameters of the Four-Terminal System

Symbol	Item	Value
N_{SM}	Number of SM per arm	50
C_{SM}	Capacitance of SM capacitor	9000 μF
R_{arm}	Bridge arm resistance	0.1 Ω
R_{on}	On-state resistance of a diode / IGBT	5.445 mΩ
L_{arm}	Bridge arm inductance	30 mH
L_{Tmn}	Supplementary inductance	50 mH
U_{dc}	DC system voltage	±200 kV
U_s / U_c	Transformer ratio	110kV / 220kV
SCR	AC system short circuit ratio	5.6
T	Time window of the signal	2 ms
f_s	Sampling frequency	50 kHz
Δi_{set}	Threshold of the start-up element	0.145 kA/ms

TABLE II Fault Scenarios for the Four-Terminal Test System

Fault distance (km)	f_1	1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99
	f_2	100, 101, 102, 103, 104, 105, 110, 120, 130, 140, 150, 160, 165, 166, 167, 168, 169
	f_3	0, -1, -2, -3, -4, -5, -10, -20, -30, -40, -50, -60, -70, -80, -85, -86, -87, -88, -89
Fault resistance (Ω)		0, 5, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500
Fault type		P2N, P2G, N2G

In order to enhance the generalization ability of the proposed fault detection and classification algorithm, a large number of simulations under various fault conditions have been carried out. Taking B_{12} for instance, DC faults can be broadly classified into three categories: the internal fault (f_1), the forward external fault (f_2), and the reverse external fault (f_3). Apparently, both of the latter two categories belong to the disturbance with respect to B_{12}. Define the fault distance as the orientation distance from

the relaying to the fault point, and DC faults at different locations can be simply represented as shown in Table II. In addition, the ranges of fault resistance and fault type are also presented in Table II, and totally 2640 fault scenarios have been taken into account. Notably, the start-up criterion expressed in (7) will always hold when encountering internal faults, while may not be triggered under some high-resistance external fault cases, so these cases can be pre-excluded from the training set.

B. Feature extraction results

Fig. 4 shows the process of the HCTSA-based feature selection choosing SRC as the classifier. It can be clearly seen that the misclassification rate of the classifier drops drastically with the increase of the number of features when it is relatively small, which indicates that utilizing more features can effectively enhance the performance of fault detection and classification. However, when the number of features exceeds 6, simply adding new features cannot bring about apparent improvement of the classification rate. The difference of misclassification rate with 7 features and that with 6 features is only 0.0011, which is sufficiently small so that the termination criterion of the feature selection can be satisfied. Therefore, for this fault classification task, the number of features is determined as 6 marked by the red point in Fig. 4. As illustrated by Table III, the selected features derive from a wide range of scientific subjects including the visibility graph analysis [28], crucial statistical parameters, stationarity properties, and wavelet transform. These features describe various time series characteristics from different perspectives, and thereby render a comprehensive picture of transient DC fault current.

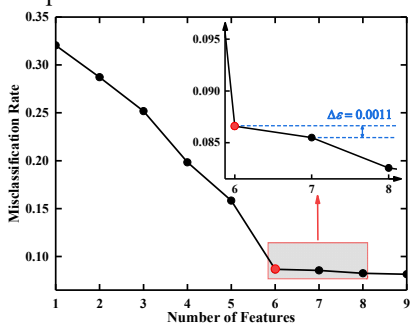


Fig. 4. Change of misclassification rate versus the number of features.

TABLE III SELECTED FEATURES FOR FAULT IDENTIFICATION

Symbol	Interpretation	Category
Feature 1	Average degree of all vertexes	Visibility graph
Feature 2	Skewness	Distribution
Feature 3	Ratio of mean-stationarity of original and resampled series	Stationarity
Feature 4	Median of top-20%-absolute-value data	Stationarity
Feature 5	Maximum 4th-layer detail coefficient with Symlet2 wavelet	Wavelet
Feature 6	Fifth-order central moment	Distribution

Fig. 5 provides an intuitive visualization of the selected features. Fig. 5 (a) and (b) are respectively the boxplot of values of Feature 1 and Feature 2 under different fault conditions, where the white dash line denotes the median value and the black square dot represents the mean value. It can be seen from Fig. 5 (a) that the distributions of Feature 1 amongst the four classes are quite different, which well interprets why Feature 1 is an effective index for DC fault detection and classification. Meanwhile, it is also indicated that utilizing a single feature can hardly distinguish the DC fault class accurately, since there are obvious overlaps between the four classes. Comparing Fig. 5 (a) and (b), it is observed that Feature 2 has different distribution characteristics from Feature 1. In other words, Feature 2 can to some extent provide complementary information for Feature 1, which explains the phenomenon presented in Fig. 4 that fault identification based on the combination of multi-features tends to possess a higher classification rate than the single-feature-based approach. In order to present the overall distribution properties of the selected 6 features, a mainstream visualization technique, namely the parallel coordinates, is utilized as illustrated in Fig. 5 (c), where the reduced features of any single current waveform are represented by the sequential connection of 6 coordinate points respectively in 6 parallel-arranged axes. It can be seen that the feature-based representations of different categories marked by varied colors have distinct patterns from each other, which demonstrates the feasibility of accurate fault classification using the selected 6 features.

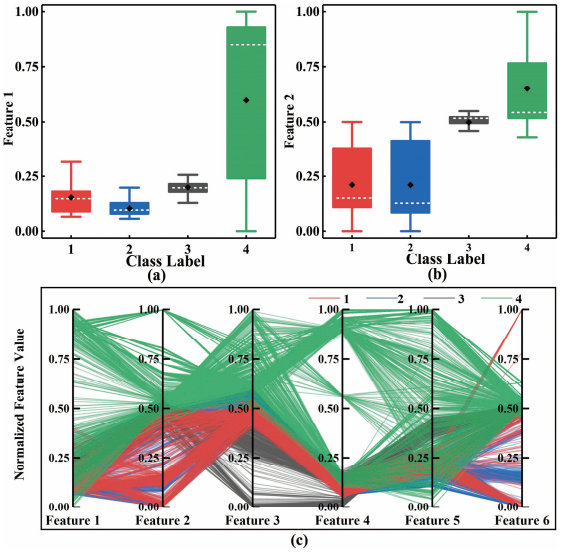


Fig. 5. Visualization of the selected features on the training data.

C. Performance of SRC for fault identification

The training process of the SRC for fault identification using the selected 6 features as the input is presented in Fig. 6 (a). It can be clearly seen that the loss function value in the training set denoted by the black solid line and that in the test set denoted by the green dash line reach their minimum within 35 iterations, which manifests the efficiency of the *gradient descend* algorithm for training the classifier. Moreover, the accuracy in the training set marked as the red solid line and that in the test set marked as the blue dash line are almost the same, indicating that the simple structure of SRC can effectively avoid overfitting. Fig. 6 (b) shows the confusion matrix of the trained SRC on the test set, where each row represents the predicted category, and each column denotes the true category. It is illustrated in Fig. 6 (b) that all samples in the test set are classified correctly by the SRC, which is better than the classification performance as shown in Fig. 4. It can be explained by the fact that, compared with the feature selection stage, broader information from both the positive relaying and

the negative relaying has been utilized in the fault identification stage. In order to thoroughly evaluate the performance of the proposed SRC, classical 10-times 10-fold cross-validation is carried out here, meaning the fault cases for training and test are randomly and repeatedly assigned. The average classification accuracy is 99.84%, which again verifies the effectiveness of the proposed SRC for fault detection and classification.

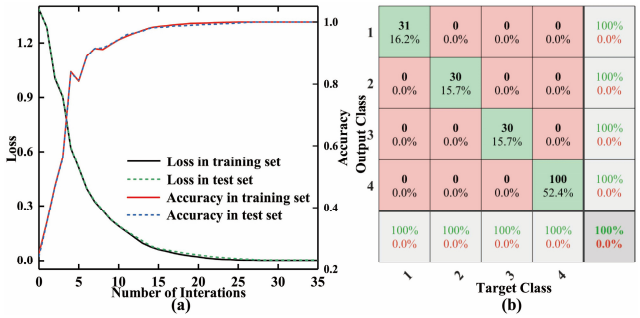


Fig. 6. Training process and result of the SRC for fault identification.

D. Computational complexity analysis

In the proposed method, off-line training and on-line application are separately conducted. Since HCTSA involves a large number of feature extraction, the computational complexity of off-line training is rather high. However, this is a one-off computational cost: once suitable features and SRC parameters for fault identification have been determined in the off-line training, only a few selected features are required to be computed in the on-line fault identification. Also, it is feasible to complete the off-line training over a relatively long time period with computers rather than protection devices, while the on-line computational burden should be the main concern.

TABLE IV ON-LINE COMPUTATIONAL BURDEN OF THE PROPOSED SCHEME

Stage		Number of operations	
		Multiplication	Exponentiation
Start-up judgment		2	0
Feature extraction	Feature 1	2	0
	Feature 2	$4\times L+4$	0
	Feature 3	$3\times L/5+10$	0
	Feature 4	0	0
	Feature 5	$\frac{31}{4}\times 2^{\mathrm{ceil}(\log_2 L)}$	0
	Feature 6	$8\times L+4$	0
	Normalization	24	12
SRC judgment		49	4
Total		$\frac{31}{4}\times 2^{\mathrm{ceil}(\log_2 L)}+\frac{63}{5}L+95$	16

The on-line computation burden of the proposed method mainly comes from three aspects: the start-up judgment, feature extraction, and SRC judgment. Based on (6), the start-up element only includes 2 comparisons, $4\times L_s$ additions, and 2 multiplications for the bipolar DC current. The feature extraction requires $2\times F_r$ time-series operations, which is normally the majority of the on-line computational cost. Based on (9), the SRC judgment requires (K-1) comparisons, (2×K-1) additions, (2×K×F_r+1) multiplications, and K exponentiations. Detailed on-line computation requirements of the proposed method are shown in Table IV, where only multiplications and exponentiations are presented since these two operations are more time-consuming than comparison and addition operations. Supposed that L equals 100, the proposed method requires 2347 multiplications and 16 exponentiations in total for realizing fault identification. It can be seen that the proposed fault detection and classification scheme has the advantage of relatively minor on-line calculation complexity, which is promising for practical applications.

E. Cooperation with DC CBs

Fig. 7 shows the change of DC current when a metallic P2N fault occurs at the middle of Cable 12. In Fig. 7 (a), the fault line current suddenly surges when suffering DC fault, resulting in the start-up element initiating the protection, and the fault transient signal will be recorded for the following fault identification, corresponding to Stage B. Then, in Stage C, feature extraction and SRC-based fault identification will be conducted using the measured waveform. It should be noted that this process only takes around 47 μs in the personal computer, while an additional time delay of 0.7 ms has been set to mimic the computing speed of practical protection devices. Subsequently, according to the fault identification output, tripping signals are sent to CBs so that they can fast cut off the fault current as shown in Stages D and E, where the breaking time of CB is around 2.2 ms. Additionally, the transient average increment of current measured at B_{42} is depicted in Fig. 7 (b). It can be seen that the operation of DC CBs may cause a slight transient process on other healthy lines, but their transient average increment is far less than the threshold so that the protection will not mal-operate. The simulation results demonstrate the proposed fault detection and classification scheme can well cooperate with DC CBs to realize fast and reliable DC fault isolation.

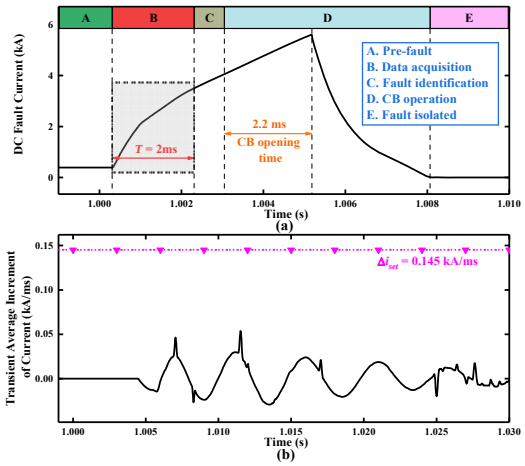


Fig. 7. Cooperation with DC CBs.

F. Impacts of noise and abnormal sampling

Fig. 8 depicts the impact of noise and abnormal samplings on the proposed fault identification approach. The system dynamic response when a P2G fault with fault resistance of 400 Ω occurs at the end of Cable 12 is shown in Fig. 8 (a), where the black solid line and green dash line separately denote the original signal and that with data drop, while the blue line and red line represent the noisy signals with the signal to noise ratio (SNR) of 36 dB and 26 dB, respectively. It can be seen that noise corruptions and abnormal samplings may induce significant distortion into the transient current waveform. However, the

features with noise corruption or data drop are quite similar to those of the original signal as illustrated by Fig. 8 (b), which indicates that the selected time-series features to some extent contribute to improving the anti-interference capability of the fault identification. Also, the change principles of fault classification performance versus the intensity of noise are further investigated. Considering it is an objective reality that there are some differences between training data and true fault data, both training and test data are contaminated with noise. In each noise test, Gaussian noise with a certain SNR has been added to all the data, and 10-times 10-fold cross-validation is conducted to output the classification accuracy of this test. Then, re-contaminate the data and repeat the noise test 10 times for each SNR level. The results of the 10-times noise test are presented in Fig. 8 (c). It is shown that the classification rate of SRC gradually decreases as the intensity of noise increases, and greater uncertainty of the classification rate will appear as the noise becomes stronger. Nevertheless, the accuracy of the SRC is still above 92% even subjected to rather strong noise corruptions. Moreover, Fig. 8 (d) shows the individual average accuracy of each fault category when the peak pointing of every current waveform is missing. The classification rate suffering data loss is still at a high level, and the average accuracy over all fault categories is 97.85%. It is well-demonstrates that the proposed fault detection and classification scheme is robust against noise corruptions and abnormal samplings.

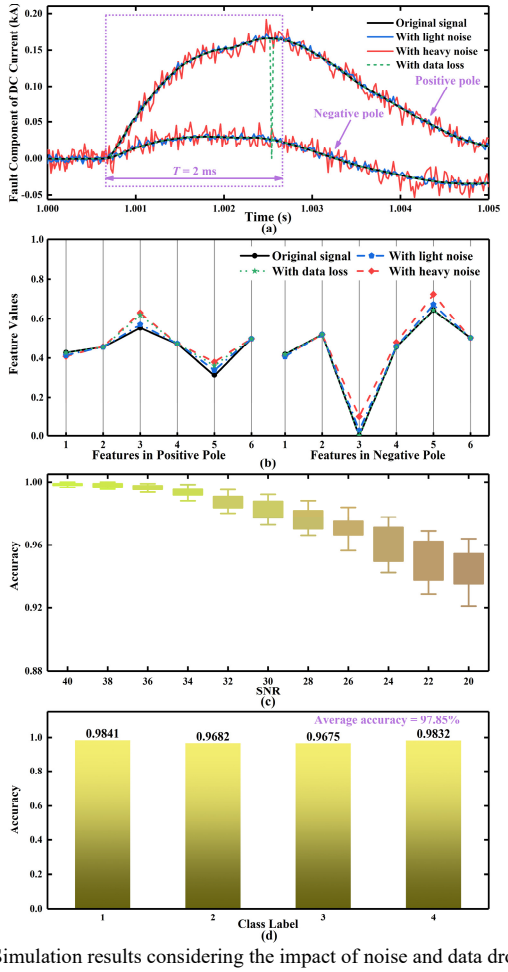


Fig. 8. Simulation results considering the impact of noise and data drop.

G. *Comparison with existing methods*

1) *Conventional derivative-based method*

Derivative-based methods that examine the ROCOC [8] or ROCOV [9] are commonly used for achieving fast fault detection. To test the performance of these methods, three P2N fault scenarios have been considered here: Internal Fault 1 happens at the end of Cable 12 with the fault resistance of 500 Ω, External Fault 2 occurs between the supplementary reactor L_{T21} and the DC bus T_2 with the fault resistance of 50 Ω, and External Fault 3 is a backward metallic fault occurring between the supplementary reactor L_{T13} and DC bus T_1. Fig. 9 (a) and (b) respectively depict the fault component of current and the corresponding ROCOC under Internal Fault 1 and External Fault 2. It can be seen from Fig. 9 (b) that the maximum ROCOC under external fault can exceed that under internal fault, indicating that merely utilizing ROCOC cannot completely distinguish between the internal fault and disturbance. Fig. 9 (c) and (d) illustrate the fault component of voltage and the corresponding ROCOV under Internal Fault 1 and External Fault 3, individually. As shown in Fig. 9 (d), ROCOV under External Fault 3 is a little bit larger than that under Internal Fault 1, implying the ROCOV-based algorithm may not be able to well identify the backward external faults. More importantly, derivative-based methods are very sensitive to noise. Although the current and voltage waveforms might have some very subtle changes under slight noise with the SNR of 36 dB as depicted in Fig. 9 (a) and (c), the corresponding ROCOC and ROCOV curves could deform severely as shown in Fig. 9 (b) and (d). As a result, the absolute value of the threshold of derivative-based methods should be large enough to prevent noise interference, but simultaneously, its capability of tolerating fault resistance will decrease significantly.

To overcome the shortcomings of pure derivative-based methods, an integrated-criteria method using both current and voltage signals has been proposed in [10]. The attributes of three conventional methods (ROCOC-based, ROCOV-based, and integrated-criteria algorithms) and the proposed method are categorized and summarized in terms of the classification accuracy, fault detection speed, tolerance for low sampling frequency (TFSF), and tolerance for noise corruptions (TFN), as shown in Table V. TFSF and TFN are respectively defined as the corresponding sampling frequency and the SNR of noise when the average fault classification rate decreases to a certain level (below 95% of the original accuracy in this paper). It can be seen that the proposed method possesses significantly higher classification accuracy than the conventional approaches due to the powerful ability of AI algorithms to dig out and synthesize multi-dimensional information. Besides, the average accuracy of the proposed method reduces to 93.24% at 10 kHz sampling frequency and 94.16% with the noise of 20 dB as shown in Table V, which manifests the proposed method is more prominent than others in terms of the relatively loose requirement for sampling frequency as well as robustness against noise corruptions. As for the fault detection speed, the proposed algorithm is much slower than derivative-based algorithms since it involves a data acquisition of real-time signal sampling for 2 ms. This implies a higher requirement of

interrupting large currents for DC CBs, which is the main drawback of the proposed method. Therefore, designing AI-based protection algorithms with a shorter sampling window and detection time will be our future work.

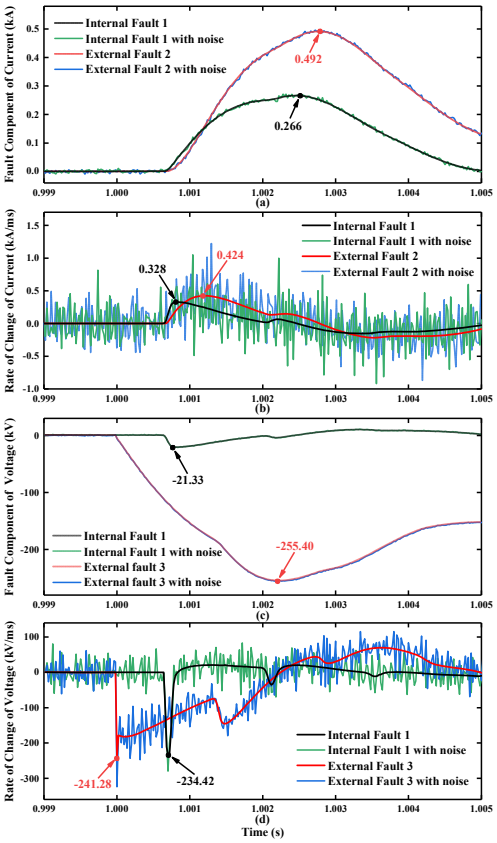


Fig. 9. Simulation results under different fault conditions.

TABLE V COMPARISON BETWEEN CONVENTIONAL PROTECTIONS AND THE PROPOSED METHOD

Method	Average accuracy	Average detection time	TFSF	TFN
ROCOC-based	83.78%	0.221 ms	16.7 kHz (76.21%)	34 dB (77.79%)
ROCOV-based	91.07%	0.146 ms	16.7 kHz (77.28%)	32 dB (77.50%)
Integrated-criteria	94.63%	0.168 ms	16.7 kHz (89.05%)	36 dB (89.00%)
Proposed method	99.84%	2.047 ms	10.0 kHz (93.24%)	20 dB (94.16%)

2) AI-based methods

Six state-of-the-art AI-based fault identification methods are considered here for the comparison, including five traditional shallow data-driven methods (i.e., ANN with FFT of the DC fault current as the input [20], ANN with Daubechies4 WT of the fault current as the input [21], ANN with the first three principal components of the fault current as the input, SVM with WT entropy as the input [19], and fuzzy-neural patter recognizer [22]), and one deep architecture method (i.e., CNN utilizing the original current signal as the input [24]). The 10-times cross-validation results of the above-mentioned methods are presented in Table VI. Compared with traditional shallow structure AI-based methods, the proposed approach gives significantly higher classification accuracy, which should be attributed to the meticulous feature selection. Both the CNN-based method and the proposed one have a rather high classification accuracy (above 99%), while the proposed approach stands out for its better interpretability, fast training convergency speed, and minor on-line computational burden. In addition, the average computing time for on-line fault detection is also provided in Table VI, where all these methods are conducted in a personal computer with Intel(R) Core(TM) i7-8565U at 1.8 GHz and 8 GB RAM using MATLAB R2019a. It can be seen that the on-line computing time of the proposed method is a little bit longer than that of ANN combined with PCA, and much shorter than those of other methods, which again demonstrates the on-line computational efficiency of the proposed method.

TABLE VI PERFORMANCE OF SOME MAINSTREAM AI-BASED METHODS

Method	Average accuracy	Average iterations	Average computing time
ANN combined with FFT	83.85%	532	0.632 ms
ANN combined with WT	89.81%	458	0.161 ms
ANN combined with PCA	96.27%	130	0.033 ms
SVM combined with WT	91.87%	376	0.658 ms
Fuzzy-neural recognizer	92.92%	271	0.406 ms
CNN	99.25%	1100	1.452 ms
Proposed method	99.84%	33	0.047 ms

3) Comprehensive evaluations based on ROC

Apart from the classification accuracy and convergence property, classical receiver operating characteristics (ROC) curve [29] is utilized to comprehensively evaluate the performance of the above-mentioned methods. The ROC curve illuminates how the test approach makes a tradeoff between the false positive rate (FPR) and the true positive rate (TPR), and a larger area under the ROC curve (AUC) means the test method has better performance. As depicted in Fig. 10, the CNN-based method and the proposed method provide the best performance among all the methods, showing feature selection can have vital influence on fault identification. The ROC evaluation results well-demonstrate the superiority of the proposed fault detection and classification method.

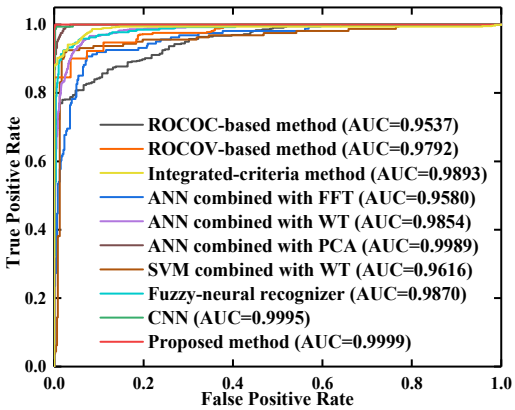


Fig. 10. ROC curves of various fault identification methods.

H. Adaptability to various MTDC systems

To demonstrate the replicability of the proposed method, a six-terminal HVDC system adopting the symmetrical bipolar structure with hybrid transmission lines consisting of overhead lines (OHLs) and underground cables has been established in PSCAD/EMTDC as shown in Fig. 11 [22]. Related parameters

are given in Table VII, and 2475 fault scenarios have been considered in this test.

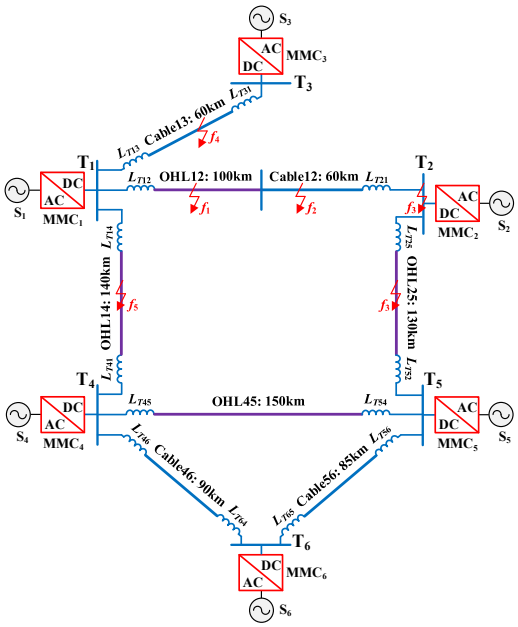


Fig. 11. Outline of the test six-terminal VSC-HVDC system.

TABLE VII RELATED PARAMETERS OF THE SIX-TERMINAL SYSTEM

Symbol	Item	Value
N_{SM}	Number of SM per arm	200
C_{SM}	Capacitance of SM capacitor	8000 μF
R_{arm}	Bridge arm resistance	0.1 Ω
R_{on}	On-state resistance of a diode / IGBT	5.445 mΩ
L_{arm}	Bridge arm inductance	70 mH
L_{Tmn}	Supplementary inductance	100 mH
U_{dc}	DC system voltage	±400 kV

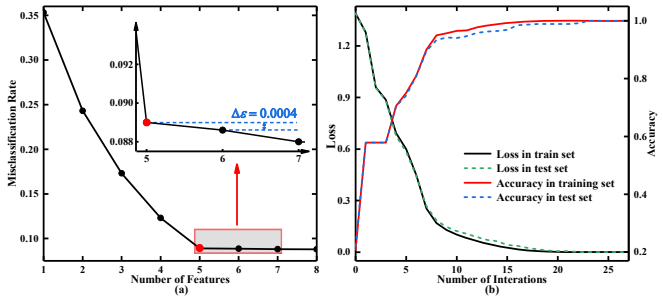


Fig. 12. Performance of the proposed method in the six-terminal test MTDC system.

TABLE VIII SELECTED FEATURES FOR FAULT IDENTIFICATION IN THE SIX-TERMINAL TEST SYSTEM

Symbol	Interpretation	Category
Feature 1	Autocorrelation coefficient with 3 points shifting	Auto-mutual information
Feature 2	Autocorrelation coefficient with 2 points shifting	Auto-mutual information
Feature 3	Intercept of the linear fitting of logarithmic power spectrum	Power spectrum
Feature 4	Skewness	Distribution
Feature 5	Decay rate of the 1-st layer detail coefficients with Daubechies3 wavelet	Wavelet

Fig. 12 (a) and (b) depict the feature selection process and the training process of SRC, respectively. It can be seen from Fig. 12 (a) that the misclassification rate drops dramatically as the number of features increases from 1 to 5, while after determining 5 features, the misclassification rate decays slowly with adding new features. Therefore, 5 features are selected for the fault identification of the six-terminal test system, as presented in Table VIII. Notably, the selected features are different from those selected in the four-terminal test system, and it is reasonable because fault characteristics may vary with the change of system topology and transmission materials so that the proposed method is supposed to select more proper features according to the specified system. Fig. 12 (b) shows the training process of the SRC utilizing the selected 5 features of bipolar current waveforms as the input. It can be seen that the classification accuracy grows to 1.0 within 30 iterations, which again demonstrates the training convergency speed of the proposed SRC is rather fast. Moreover, 10-times 10-fold cross-validation has been carried out and the average classification accuracy reaches 100%, which well verifies the effectiveness of the proposed method with respect to various MTDC systems.

V. CONCLUSION

This paper puts forward a data-driven fault detection and classification framework for MTDC systems. HCTSA is first utilized to automatedly select effective and interpretable features from the transient DC current waveform. The softmax regression model is then adopted to build the relationship between the values of the selected features and the probability of the fault category. After parameters tuning and optimizing in the off-line phase, the proposed algorithm can fast and accurately identify DC faults with a relatively low on-line computing burden. Numerical simulation results show that for MMC-MTDC systems with weak line boundary, high-speed non-unit protections using a single feature fail to distinguish the fault category accurately, while synthesizing multi-dimensional features as well as rationally selecting useful features can effectively boost the fault classification accuracy. Compared with existing conventional derivative-based methods and shallow structure AI-based methods, the proposed one stands out mainly for its high fault identification accuracy and high anti-interference capability. Besides, it has the advantages of good interpretability and minor on-line computational burden over deep architecture AI-based approaches. Considering there is room for improving the accuracy of the proposed algorithm under noise corruptions and abnormal samplings, exploring how to comprehensively utilize voltage and current signals to form features with superior performance, and how to design more robust intelligent algorithms with lower computational complexity will be our future work.

REFERENCES

[1] N. R. Chaudhuri, R. Majumder, and B. Chaudhuri, "System frequency support through multi-terminal DC (MTDC) grids," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 347-356, Feb. 2012.

[2] T. Ding, Z. Zeng, B. Qin, et al., "Quantifying cyber attacks on industrial MMC-HVDC control system using structured pseudospectrum," *IEEE Trans. on Power Electron.*, to be published.

[3] M. Wang, S. Yan, S. Tan, et al., "Decentralized control of DC electric springs for storage reduction in DC microgrids," *IEEE Trans. Power Electron.*, vol. 35, no. 5, pp. 4634-4646, May 2020.

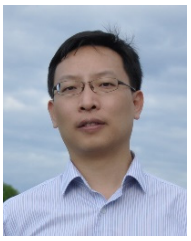
[4] Q. Tu, Z. Xu, and L. Xu, "Reduced switching-frequency modulation and circulating current suppression for modular multilevel converters," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 2009-2017, Jul. 2011.
[5] Y. Li, J. Li, G. Wu, et al., "DC fault analysis models of three converter topologies considering control effects," *IEEE Trans. Ind. Electron.*, vol. 67, no. 11, pp. 9480-9491, Nov. 2020.
[6] B. Qin, W. Liu, R. Zhang, et al., "Small-signal stability analysis and optimal control parameters design of MMC-based MTDC transmission systems," *IET Gener. Transm. Distrib.*, vol. 14, no. 21, pp. 4675-4683, Oct. 2020.
[7] Y. Li, J. Li, L. Xiong, et al., "DC fault detection in meshed MTDC systems based on transient average value of current," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 1932-1943, Mar. 2020.
[8] J. I. Marvik, S. D'arco, and K. Sharifabadi, "Protection scheme for multi-terminal radial VSC HVDC system without communication between terminals," *CIGRE Int. Symp.-Across Borders-HVDC Syst. Mark. Integr.*, pp. 1–10, May 2015.
[9] J. Sneath, and A. D. Rajapakse, "Fault detection and interruption in an earthed HVDC grid using ROCOV and hybrid DC breakers," *IEEE Trans. Power Del.*, vol. 31, no. 3, pp. 973–981, Jun. 2016.
[10] W. Leterme, J. Beerten, and D. Van Hertem, "Nonunit protection of HVDC grids with inductive DC cable termination," *IEEE Trans. Power Del.*, vol. 31, no. 2, pp. 820-828, Apr. 2016.
[11] J. Li, Y. Li, L. Xiong, et al., "DC fault analysis and transient average current based fault detection for radial MTDC system," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1310-1320, Jun. 2020.
[12] Z. Zheng, N. Tai, J. S. Thorp, et al., "A transient harmonic current protection scheme for hvdc transmission line," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2278-2285, Oct. 2012.
[13] 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 Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4080-4091, May 2018.
[14] K. De Kerf, K. Srivastava, M. Reza, et al., "Wavelet-based protection strategy for DC faults in multi-terminal VSC HVDC systems," *IET Gener. Transm. Distrib.*, vol. 5, no. 4, pp. 496-503, Apr. 2011.
[15] B. Li, Y. Li, J. He, et al. "An improved transient traveling-wave based direction criterion for multi-terminal hvdc grid," *IEEE Trans. Power Del.*, vol. 35, no. 5, pp. 2517-2529, Oct. 2020.
[16] R. Bertho, V. A. Lacerda, R. M. Monaro, et al., "Selective nonunit protection technique for multiterminal VSC-HVDC grids," *IEEE Trans. Power Del.*, vol. 33, no. 5, pp. 2106-2114, Oct. 2017.
[17] A. Raza, A. Akhtar, M. Jamil, et al., "A protection scheme for multi-terminal VSC-HVDC transmission systems," *IEEE Access*, vol. 6, pp. 3159-3166, Dec. 2017.
[18] M. Farshad, "Detection and classification of internal faults in bipolar HVDC transmission lines based on K-means data description method," *International Journal of Electrical Power & Energy Systems*, vol. 104, pp. 615-625, Jan. 2019.
[19] G. Luo, C. Yao, Y. Tan, et al., "Transient signal identification of HVDC transmission lines based on wavelet entropy and SVM," *Journal of Engineering*, vol. 2019, no. 16, pp. 2414-2419, Apr. 2019.
[20] Q. Yang, S. L. Blond, R. Aggarwal, et al., "New ANN method for multi-terminal HVDC protection relaying," *Elect. Power Syst. Res.*, vol. 148, pp. 192–201, Jul. 2017.
[21] S. Vasanth, Y. M. Yeap, and A. Ukil, "Fault location estimation for VSC-HVDC system using artificial neural network," *2016 IEEE Region 10 Conference (TENCON) IEEE*, pp. 501-504, Nov. 2016.
[22] A. Hossam-Eldin, A. Lotfy, M. Elgamal, at al., "Artificial intelligence-based short-circuit fault identifier for MT-HVDC systems," *IET Gener. Transm. Distrib.*, vol. 12, no. 10, pp. 2436-2443, May 2018.
[23] W. Xiang, S. Yang, and J. Wen, "ANN-based robust DC fault protection algorithm for MMC high-voltage direct current grids," *IET Renewable Power Generation*, vol. 14, no. 2, pp. 199-210, Mar. 2020.
[24] J. Mei, R. Ge, Z. Liu, et al., "An auxiliary fault identification strategy of flexible HVDC grid based on convolutional neural network with branch structures," *IEEE Access*, vol. 8, pp. 115922-115931, Jun. 2020.
[25] B. D. Fulcher, and N. S. Jones, "hctsa: A computational framework for automated time-series phenotyping using massive feature extraction," *Cell syst.*, vol. 5, no. 5, pp. 527-531, Nov. 2017.
[26] B. D. Fulcher, and N. S. Jones, "Highly comparative feature-based timeseries classification," *IEEE Trans. Knowl. Data Eng.*, vol. 26, pp. 3026–3037, Dec. 2014.
[27] J. Fu, G. Song, and B. De Schutter, "Influence of Measurement Uncertainty on Parameter Estimation and Fault Location for Transmission Lines," *IEEE Trans. Autom. Sci. Eng.*, to be published.
[28] B. Luque, L. Lacasa, F. Ballesteros, et al., "Horizontal visibility graphs: Exact results for random time series," *Phys. Rev.*, E 80, 046103, Oct. 2009.
[29] H. He, and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263-1284, Jun. 2009.

Jiapeng Li received the B.Eng. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2017, and is currently pursuing the Ph.D. degree at both Xi'an Jiaotong University and The Hong Kong Polytechnic University. His research interests mainly include protection and fault location of HVDC system.

Guobing Song (M'10, SM'20) received the Ph.D. degree in electrical engineering at the Xi'an Jiaotong University, Xi'an, China, in 2005. Currently, he works in the Xi'an Jiaotong University. His research interests include transmission line fault location and protections.

Jifei Yan received the B.Eng. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2020, and is currently pursuing the M.Sc. degree at Xi'an Jiaotong University. His research interests include transmission line protection.

Yujun Li (S'15, M'20) received the B.Sc. degree from Xi'an Jiaotong University, Xi'an, China, in 2011, the M.Sc. degree from Zhejiang University, Hangzhou, China, in 2014, and the Ph.D. degree from Hong Kong Polytechnic University, Hung Hom, Hong Kong, in 2017, all in electrical engineering. In 2017, he joined the School of Electrical Engineering, Xi'an Jiaotong University, where he was a Lecturer and is currently an Associate Professor with the School of Electrical Engineering, Xi'an Jiaotong University. His main fields of interest include grid integration of renewable energy and high-voltage directing current modeling, fault analysis and detection of power systems.



Zhao Xu (M'06, SM'13) received the Ph.D. degree in electrical engineering from The University of Queensland, Brisbane, Australia, in 2006. From 2006 to 2009, he was an Assistant and later Associate Professor with the Centre for Electric Technology, Technical University of Denmark, Lyngby, Denmark. Since 2010, he has been with Hong Kong Polytechnic University. His research interests include demand side, grid integration of wind power, electricity market planning and management, and AI applications. He is an Editor of the Electric Power Components and Systems journal.