IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 8, 2021, accepted July 22, 2021, date of publication July 26, 2021, date of current version August 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099866

Research on Double-Deck Traceability Identification Method of Commutation Failure in HVDC System

YUHONG WANG, (Member, IEEE), KEQIANG TAI, YUYAN SONG, (Student Member, IEEE), RAN KOU, ZONGSHENG ZHENG, (Member, IEEE), AND QI ZENG

College of Electrical Engineering, Sichuan University, Chengdu 610065, China

Corresponding author: Qi Zeng (zengqi-hk@163.com)

This work was supported in part by the Sichuan Science and Technology Program under Grant 2019YJ0114.

ABSTRACT In AC/DC hybrid power system, AC system failures and commutation valve trigger pulse disorder will lead to commutation failure, which may lead to DC voltage fluctuations, power transmission interruption and other serious consequences. In order to accurately and effectively identify the specific causes of commutation failures, a double deck traceability identification method is proposed in this paper. The surface identification based on wavelet entropy and affinity propagation (AP) algorithm can distinguish internal and external faults. The deep identification uses convolution neural network which can further lock the specific cause of commutation failures. In this paper, 1) the various factors leading to commutation failures are analyzed; 2) the fault feature space consists of the wavelet analysis components of DC voltage signal, and the AP algorithm is used to identify the surface source; 3) the DC current, AC voltage and current signals are added into the sample matrix of fault time-space, and the convolution neural network is used to identify the deep traceability. Finally, the accuracy of the method is verified by using the typical HVDC model.

INDEX TERMS Commutation failures, AP clustering algorithm, deep learning, fault cases identification, double-deck method.

I. INTRODUCTION

Commutation failure (CF) is one of the typical faults of traditional high-voltage direct current (HVDC) system [1]. If the cause of CF persists or is not properly handled, it will lead to subsequent CF, which will force the DC system to block and seriously affect the safe and stable operation of AC and DC systems [2]–[4]. The voltage drop of the inverter side and DC current rise are the main reasons for commutation failures [5]–[7]. The fault of internal trigger pulse circuit in commutation valve will also cause commutation failures. In [8], the commutation voltage time area was proposed and combined with the commutation angle to study the mechanism of commutation failures. Some scholars considered AC system faults, the change of AC voltage and phase shift to summarize the impact to CF in [9]–[11]. Some researchers focused on the prediction control and evaluation of CF, which could improve the stability of HVDC system in [12] and [13]. With the rapid development of HVDC project, commutation failure of multi infeed HVDC system has been studied in [9], [14], [15].

In view of CF fault identification, the existing literatures mainly focus on the identification indicators and threshold setting, which are the important part of methods. In [16], the DC voltage on the inverter side was selected as the monitoring signal. The wavelet energy of each layer was calculated and constructed as the inputs. The neural network based on back propagation (BP) was used to identify the fault types. However, the internal fault type of the converter valve was not considered and the specific type of DC line fault was also not specified. However, the internal fault type of the converter valve was not considered and the specific type of DC line fault was also not specified. A CF diagnosis method based on wavelet packet decomposition and

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

generalized regression neural network was proposed in [17]. AC current was monitored during the CF, and wavelet packet decomposition was carried out to construct a new energy spectrum. On this basis, the generalized return neural network in two modes was used to judge whether commutation failures occurs or not. It can also distinguish the specific fault cases. The decomposition of DC current and AC voltage was realized by wavelet analysis, and the specific fault types were distinguished and identified according to the wavelet energy statistics of each layer in [18]. However, the index definition and the threshold setting were mostly related to the system parameters. The structure setting was related to the general adaptability of the project. In [19], the grey comprehensive relationship degree and wavelet energy spectrum were introduced into the CF fault diagnosis. The inverter side DC current was selected as the original input signal. The wavelet energy spectrum was used to extract features. According to grey comprehensive relationship degree, the fault could be diagnosed. However, the specific failure type of the converter valve was not considered. The general idea of the above literatures is to extract the features of fault information and diagnose the fault by combining with identification algorithms. However, the parameters for the most of methods are required to be taken closely with the related to specific projects.

Accurately identifying the fault causes of commutation failures and taking corresponding measures in time could provide a powerful guarantee for suppressing the subsequent commutation failures, speeding up the fault recovery and ensuring the safety and stability of the system operation. In this paper, a double-deck tracing method for identifying the causes of CF in HVDC system is presented, which includes surface identification and deep identification. In surface identification, wavelet entropy is used to extract fault signal features and construct fault feature space. The AP algorithm is introduced to distinguish the converter valve faults from the AC system faults. In this process, the data's own characteristic is fully considered, which could avoid the interference of human factors, so as to make the surface identification unsupervised. In deep identification, sample matrix of fault time-space is constructed by auxiliary electrical signals, which could reflect the characteristics of the system during the fault more effectively. The convolution neural network is selected as the deep identification algorithm to identify the deep traceability of CF. Finally, the accuracy of the proposed method is verified by using the typical HVDC model.

The rest of the paper is organized as follows: Section II analyzes the CF mechanism and the different fault causes of HVDC system. Section III proposes the surface identification method and the composition of fault feature space.. Section IV proposes the deep identification method and the composition of sample matrix of fault time-space. In section V, the process of double-deck identification method is proposed. The case study is shown in Section VI, and the simulation is undertaken on the typical HVDC model. Based on the case study, conclusions are presented in Section VII.

II. ANALYSIS OF COMMUTATION FAILURES MECHANISM OF HVDC SYSTEM

The factors that cause commutation failures are complex. At present, the commonly used expression of the extinction angle of valves at the inverter side of HVDC transmission system is described as follows [20]:

$$\gamma = \arccos\left(\frac{\sqrt{2}I_d X_c}{U_{ac}} + \cos\beta\right) - \varphi \tag{1}$$

where I_d is the DC current, X_c represents the commutation reactance and it can be determined by $X_c = \omega L_c$. The commutating voltage is U_{ac} which is AC bus line voltage. β and φ are the leading firing angle and the forward angle of commutation voltage zero-crossing during the asymmetric grounding faults.

Generally, when the extinction angle of valves is less than its minimum angle, the commutation failures of the converter occur. It can be known from (1) that the failures of the AC system on the inverter side leads to the decrease of the AC side line voltage and the increase of the DC current, which are the main reason for the decrease of γ [7]. At the same time, the internal faults of the converter will also cause the converter valve to fail to conduct normally. In this paper, we consider the converter valve faults and AC system faults, which could cause CF directly. We use the disturbance of trigger pulse signals to simulate the faults of converter control system and firing circuit of converter.

This expression can theoretically analyze which factors can lead to the decrease of γ, but in practical engineering applications, it is very difficult to judge which fault lead to CF. The causes of CF of HVDC systems can be roughly divided into two categories, which can be expressed as two surface fault sets V and W, where V is the converter valve fault set and W is the AC side fault set. Each type of surface fault set can be subdivided into different deep fault types,, $\mathrm{V} = [\mathrm{V}_1, \mathrm{V}_2, \cdots, \mathrm{V}_n]$, $\mathrm{W} = [\mathrm{W}_1, \mathrm{W}_2, \cdots, W_n]$ and the details are shown in Table 1.

III. SURFACE IDENTIFICATION METHOD

The double-deck traceability identification method proposed in this paper is divided into surface identification and deep identification. AP algorithm is selected as the core algorithm for surface identification to identify which kind of surface fault the unknown fault signal belongs to.

A. FAULT FEATURE SPACE

When commutation failures occurs in HVDC system, the fault electrical signals often show complex and disordered characteristics. Establishing a fault identification framework with the electrical signals directly could increase the complexity of calculation. At the same time, the single dimension electrical signals cannot well represent the whole characteristics of the fault signal. Therefore, it is particularly important to reasonably extract the characteristic information contained in the fault signal from multiple dimensions.

TABLE 1. Fault causes classification.

Surface fault	Deep fault	
	Types	Explanation
Converter valve faults V	V_1	Single bridge single trigger pulse loss of converter valve
	V_2	Multi bridge single trigger pulse loss of converter valve
	V_3	Single bridge multi trigger pulse loss of converter valve
	V_4	Multi bridge multi trigger pulse loss of converter valve
AC system faults W	W_1	Single phase to grounding fault at AC system of converter valve
	W_2	Double phases to grounding fault at AC system of converter valve
	W_3	Phase to phase short circuit fault at AC system of converter valve
	W_4	Three phases to grounding fault at AC system of converter valve

The different types of wavelet entropy are used to show the characteristics of electrical signals to construct the fault feature space in this paper, which includes the wavelet energy entropy (WEE), wavelet singular entropy (WSE), wavelet distance entropy (WDE) [21]–[23]. The energy distribution characteristics of the signal in time domain and frequency domain are shown by WEE. WSE is used to measure the complexity and uncertainty of signals, and WDE is used to describe the distance between different coefficient matrices, which reflects the internal relation characteristics of fault signals. The expressions are shown as follows:

$$WEE = -\sum_{\varepsilon=1} (E_\varepsilon/E) \log (E_\varepsilon/E) \tag{2}$$

$$WSE = \sum_{\varepsilon=1} \left[-(\lambda_\varepsilon/\lambda) \log (\lambda_\varepsilon/\lambda) \right] \tag{3}$$

$$\begin{cases} WDE = -\sum_{s=1} D_s \ln D_s \\ D_s = \|x_a - x_b\| / \sum_{a \neq b} \|x_a - x_b\| \end{cases} \tag{4}$$

where E_ε represents the energy of wavelet layer ε and E is the sum of the energy, λ_ε represents the singular value of the coefficient matrix of wavelet layer ε and λ are the sum of the singular values. D_s represents the distance between the wavelet coefficients of the s-th pair of wavelet layers. The parameters x_a and x_b are the wavelet coefficients of wavelet layer a and layer b.

Various faults could cause CF. According to the above theory, three different wavelet entropies are used to construct the fault feature space T. It not only can show the characteristic information contained in the complex and disordered fault signal from different angles, but also can reduce the complexity of the signal. So that the operation speed of the subsequent algorithm could be improved

$$T = [T_1, T_2, \cdots, T_i]^T, i = 1, 2, \cdots, m \tag{5}$$

$$T_i = [WSE_i, WEE_i, WDE_i] \tag{6}$$

where i represents a specific fault which cause CF, and m represents the number of all fault types.

B. AP ALGORITHM

AP algorithm is a new unsupervised clustering algorithm [24]. The algorithm regards all sample points as potential clustering center points. It introduces the concepts of attraction and attribution among samples, and realizes sample clustering through iterative updating. Compared with the traditional algorithms, this algorithm is insensitive to the selection of initial values and does not need human intervention [25].

Firstly, the fault feature space T is constructed by processing the data according to section A, and then the similarity matrix $S_{m\times m}$ can be calculated to initialize the algorithm

$$S(i,j) = -\left\| T_i - T_j \right\|, (i, j = 1, 2, \cdots, m). \tag{7}$$

In the matrix S, the diagonal element is called the bias parameter, and its value is used as the standard to select which sample point could become the center point.

Then, the attractiveness and the attribution among the sample points are calculated iteratively

$$\begin{cases} R_{t+1}(i,j) = \begin{cases} S(i,j) - \max\limits_{k\neq i,j} A_t(i,k) \\ \quad + S(i,k)\}, i \neq j \\ S(i,j) - \max\limits_{k\neq i,j} \{S(i,k)\}, i = j \end{cases} \\ A_{t+1}(i,j) = \begin{cases} \min\{0, R_{t+1}(j,j) \\ \quad + \sum\limits_{k\neq i,j} \max\{0, R_{t+1}(k,j)\}\}, i \neq j \\ \sum\limits_{k\neq j} \max\{0, R_{t+1}(k,j)\}, i = j \end{cases} \\ i, j, k = 1, 2, \cdots, m \\ t = 0, 1, \cdots, n \end{cases} \tag{8}$$

where $R_t(i,j)$ is the attractiveness between the point i and point j at the t-th iteration. $A_t(i,j)$ is the attribution between the point i and point j at the iteration t, it indicates the support of point i for point j to become itself center. $S(i,j)$ is the similarity between point i and point j. The initial condition is $A_0(i,j) = 0$. In (8), the attractiveness $R_t(i,j)$ and the attribution $A_t(i,j)$ are updated iteratively, and the parameter t denotes the t-th iteration.

Finally, the damping coefficient λ is introduced to avoid iterative oscillation

$$R_{t+1}(i,j) \leftarrow \lambda R_t(i,j) + (1-\lambda) R_{t+1}(i,j) \tag{9}$$

$$A_{t+1}(i,j) \leftarrow \lambda A_t(i,j) + (1-\lambda) A_{t+1}(i,j) \tag{10}$$

where $\lambda \in [0.5, 1)$. When the maximum number of iterations or the number of fixed iterations is reached, the iterative update of the algorithm should be stopped.

$$C = argmax\{Q\} = argmax\{A(i,j) + R(i,j)\} \tag{11}$$

The result matrix Q is the sum of the attractiveness matrix and the attribution matrix and the position of the maximum value of each row is found in turn. When $i = j$, it determines point j as surface identification center. When $i \neq j$, the point i is divided into the category which is represented by point j. Through above iterative process, it can achieve surface identification of unknown fault signals

IV. DEEP IDENTIFICATION METHOD

In order to further determine the detailed cause of CF, the convolution neural network (CNN) is selected for deep identification to achieve the accurate identification of the fault tracing based on the surface identification results.

A. CNN AND STRUCTURE OPTIMIZATION

CNN is one of the deep learning modes. The core idea is to alternately use convolution layer and pooling layer to extract and pool the input data layer by layer. Compared with other artificial neural networks, CNN has the characteristics of local connection, regional weight sharing and feature refinement transmission [26]. It can reduce the training parameters of the model, improve the training cost and the accuracy of the algorithm, and is suitable for dealing with large-scale data. In order to build a network with excellent structure, excellent training and test results and strong generalization ability, this paper optimizes the CNN network structure. The optimized structure adopted in this paper is described as follows:

1) ADAM OPTIMIZATION ALGORITHM

In the training of neural network, the setting of learning rate is related to the speed of parameter updating, which plays an important role in network training. Adam optimization algorithm has the invariance of gradient diagonal scaling and adaptive learning rate, which has unique advantages in dealing with large-scale and high-latitude electrical signal data. It has good performance for noisy electrical fault data and non-steady-state problems with sparse gradient as well [27]. Moreover, the overshoot parameters are simple and clear, and the default parameters have a wide range of applications without manual adjustment. In this paper, Adam optimization algorithm is selected as the learning rate adjustment optimization algorithm to improve the accuracy of deep identification.

2) CODING

The training samples coding of deep identification adopts commonly used one-hot coding, which gives each kind of deep fault type the single label. In the coding, only one bit is valid at any time. The output target of test samples is judged according to the value of different targets. The maximum value corresponding to the label determines the category of the test samples. The encoding formula is

$$R_V = R_W \in \begin{cases} R_1 \,|1, 0, 0, 0| \\ R_2 \rightarrow |0, 1, 0, 0| \\ R_3 \rightarrow |0, 0, 1, 0| \\ R_4 \rightarrow |0, 0, 0, 1| \end{cases} \tag{12}$$

where R_V and R_W are the labels of surface fault samples respectively, both of which include four deep fault types of $R_1 \sim \mathrm{R}_4$. Each of them is only label corresponding to it.

3) DROPOUT MECHANISM

CNN has unique advantages in dealing with massive data. However, when the training data contains noise and the sample size is insufficient, the phenomenon of over-fitting often occurs, which shows that it has a good ''memory'' ability for the training set, but does not ''learn'' the features contained in the training set, resulting in poor performance in identifying unknown data. Considering the interference of noise and uncertain signal in fault electrical signal, this paper adopts the dropout mechanism which is widely used to improve the network generalization ability [28].

B. SAMPLE MATRIX OF FAULT TIME-SPACE

According to the learning characteristics of CNN, it is necessary to construct a high-dimensional sample matrix, so that CNN can fully and effectively analyze the state characteristics of the system during the fault period. Considering the variation law of electrical signals at AC and DC sides of converter station during the fault period, and the operation state characteristics of AC and DC sides. The voltage and current values of DC line, current value and AC side voltage of inverter valve are selected as fault information acquisition signals to construct the sample matrix of fault time-space (SMOFTS).

In this paper, the sliding high frequency data acquisition window is used to collect data points of four kinds of electrical signals. The window width is set to 7, that is, each window contains 7 data points, and the data matrix size is 7×1. Then, the four groups of window matrices are rearranged in sequence, and the size of SMOFTS is transformed into 28×1. In order to meet the requirements of CNN inputting samples and give full play to its advantages in the field of image recognition, it is necessary to transform the data form of SMOFTS. In this paper, the matrix is reconstructed in the form of binary data conversion. The specific steps are given as follows: Firstly, the information points in the matrix are divided into integer and decimal parts. Secondly, the values of the two parts expressed in decimal system are converted into 14-digit binary numbers respectively. Finally, the two parts are combined into a twenty-eight-bit binary number according to the principle that the integer part comes before the decimal part, so that the size of the SMOFTS is converted into 28×28. The schematic diagram of the process is shown in Fig. 1.

V. DOUBLE-DECK IDENTIFICATION METHOD

The process of double-deck traceability identification method is shown in Fig. 2. It includes three parts: data processing, surface identification and deep identification. The digital simulation results are used as the data source for the proposed method. It is composed of the original data set, which could be used to train the convolutional neural networks.

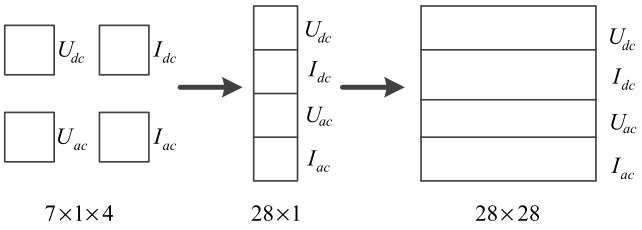

FIGURE 1. Sample matrix of fault time-space construction process.

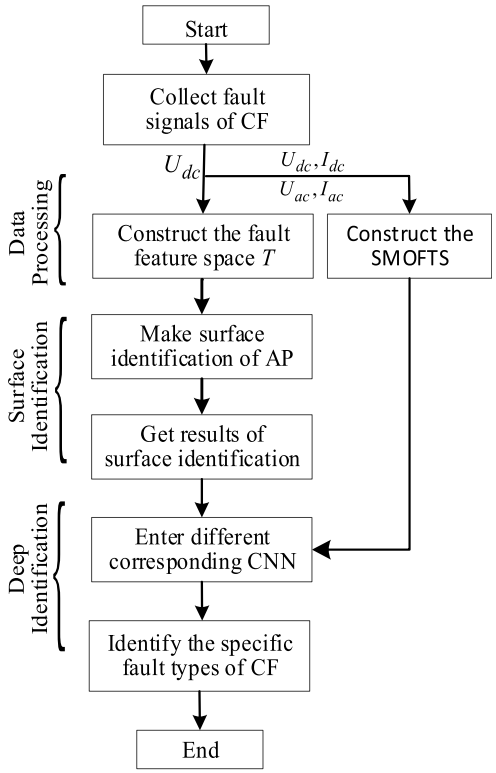

FIGURE 2. Double-Deck identification method.

DC voltage is an important representation of HVDC system transmission stability, and its fault waveform contains rich information. In this paper, DC voltage is selected as the input data source signal for surface identification method. The process of the method is shown as follows.

1) Firstly, data processing. In the converter station, various electrical signals such as U_{dc}, I_{dc}, U_{ac}, and I_{ac} are collected locally during the fault period. According to the above content, the fault feature space T and the SMOFTS are constructed respectively, which are prepared to identify the fault type.

2) Secondly, surface identification. In this part, the T of step 1) is put into the AP algorithm to surface identify. According to the result matrix Q, the surface fault category of the T could be clear.

3) Thirdly, deep identification. In this part, the CNN corresponding to the surface fault category is selected according to the result of step 2). Putting the SMOFTS of test sample into the CNN, the deep fault type of CF is identified by the output results. The label with the maximum value is determined as the failure type of the test sample.

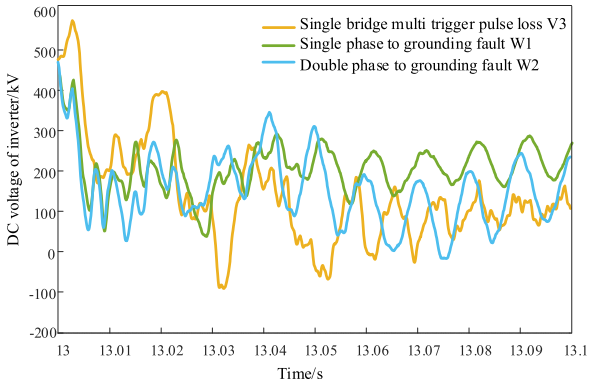

FIGURE 3. Curves of DC fault voltage.

VI. CASE STUDY

This section is to verify the validity of the proposed method. According to an actual operated HVDC system in South China power grid, a HVDC model is built in PSCAD/EMTDC. The rated voltage and capacity of DC system are respectively 500kV and 3000MW. The digital simulation results are used as the numerical source for the verification of this method, in which the simulation step is set to 50μs and the drawing step is set to 100μs. According to Table 1, all faults are set to the converter bus and converter valve of the inverter side. Failures duration is 0.1s (from 13s to 13.1s), and the electrical fault data is recorded to form surface identification reference data group. Taking the faults V_3, W_1 and W_2 as examples, the curve of DC voltage is shown in Fig. 3.

It can be seen from Fig.3 that during the fault duration, the DC voltage waveform drops to different degrees under the three fault conditions with the fault duration, the DC voltage fluctuates in different amplitudes. Among them, the variation trend of DC voltage curves from 10ms to 40ms is basically identical, ranging from 100kV to 300kV. The curve changes between different DC voltages were observed at the later stage of the fault. The DC voltage waveforms of fault V_3 and fault W_2 oscillated around 100kV, and the different voltage waveforms were slightly different during the fault period, but the discrimination was not great. It was not easy to judge the specific cause of commutation failures only from the DC voltage waveforms.

A. SURFACE IDENTIFICATION

The two-phase short circuit fault is set at the converter bus of the inverter side with grounding resistance of 10Ω and multi trigger pulse loss of No.6 valve fault are chosen as test sample, which are numbered S_1 and S_2 respectively. Fault feature space T is shown in Table 2. According to reference [24], the maximum number of iteration AP algorithm is 500, the number of fixed iterations is 50, the damping coefficient is 0.5 and the bias parameter P is the median value of data. The result of surface identification is shown in Table 3.

In Table 3, the maximum value of each row is positive value. It can be determined that [V_1,W_1] is the center of the

TABLE 2. Fault feature space T.

Number	WSE	WEE	WDE
T_{V1}	0.1528852	0.6534602	12.674902
T_{V2}	0.1446651	0.6172726	12.648114
T_{V3}	0.1877956	0.7497342	12.692359
T_{V4}	0.1451577	0.6267915	12.669075
T_{W1}	0.1347421	0.6145465	12.591860
T_{W2}	0.1422106	0.5818794	12.582178
T_{W3}	0.1404732	0.5998851	12.502838
T_{W4}	0.1334957	0.5745727	12.654896
T_{S1}	0.1377453	0.6219496	12.580569
T_{S2}	0.1810814	0.6282569	12.680849

fault cause category [V,W] of surface identification, and it is the representative fault of each surface fault cause category. And the surface fault category of other data could be judged in turn according to the number of columns in which the positive elements of each row are located. According to the maximum values of each element in the second and third rows of the result matrix Q appear in the first column, V_2, V_3 and V_1 can be classified into the same category, which belong to the surface fault category represented by V_1. Similarly, it can be known that the attribution results of other reference sample conforms to the expected design. Furthermore, the results of surface identification of test signals reveal that: the test signal S_1 is identified as AC system fault of the inverter side, and the test signal S_2 is identified as converter valve fault. The surface identification result is accurate.

In this part, the simulation uses the fuzzy-clustering algorithm as the comparison algorithm to recalculate the same data. The number of clustering center is set to 2, and the membership degree is taken as the basis of fault type division. The comparison algorithm set the Euclidean distance as based distance formula. The results are shown in Table 4. For classification of test signals, the result of fuzzy-clustering algorithm accords with pre-design, but on the result of reference data, the comparison algorithm has poor performance. The reference signal V_2 is misjudged as the AC system faults W, and the reference signal W_1 is misjudged as converter valve faults V. Therefore, it can be seen that the accuracy of AP algorithm is better than that of fuzzy-clustering algorithm, and the surface identification of unknown fault causes can be realized accurately. In order to better demonstrate the surface identification results, this paper takes the fault feature space values as the coordinates of the fault signal samples, visualizes the identification results of the surface identification, and establishes a three-dimensional diagram. The results are shown in Fig. 4.

B. DEEP IDENTIFICATION

In this paper, the TensorFlow architecture based on Python is selected to build the CNN model of deep identification method. The CNN structure parameters are shown in Table 5.

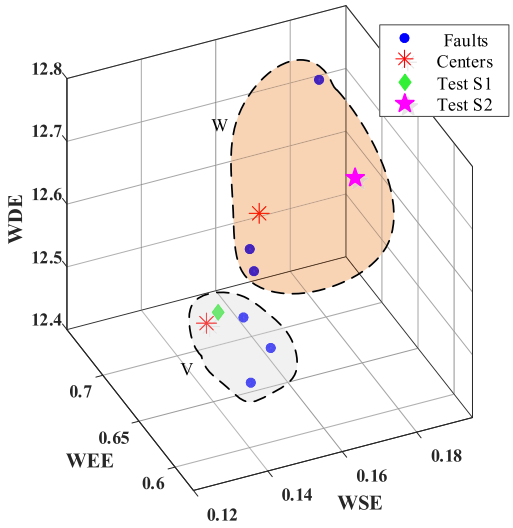

FIGURE 4. Three-dimensional fault feature space of surface identification.

According to the characteristic that the training of CNN needs a large number of samples as support, this paper expands the data samples under different fault types to supplement different fault states of the system under the same fault condition. For converter valve fault, this paper considers different bridges and different combined bridges of converter valve, set single trigger pulse loss and multi-trigger pulse loss faults during the simulation. For AC side faults, this paper uses different fault resistances ($0 \sim 50\Omega$) to represent different fault locations of transmission lines under the different AC side fault types. The simulation settings are the same as above. According to the construction method of SMOFTS, the sliding sampling is carried out on the simulation data within the failure duration of 0.1s, and a total of 21600 sample data are generated. In this paper, the training set and the test set are intentionally divided. The random sub-sampling verification method is adopted, and 2,400 samples are randomly selected as test sets. Therefore, there are 9,600 training sets and 1,200 test sets for converter valve faults. There are 9600 training sets and 1200 test sets for AC system faults.

In the simulation of test signals, the deep identification results of part test samples are shown in Tables 6 and 7, where the label with the maximum value is determined as the failure type. According to the results, the five samples which are randomly selected from the test signal S_1 are identified by the deep algorithm as double phase to grounding fault at AC system of converter valve W_2, and the test samples of test signals S_2 are identified as single bridge multi-trigger pulse loss of converter valve V_3. The values of the correct target of the identification results are much larger than other targets.

The results of the deep identification method are shown in Fig. 5, where principal component analysis (PCA) is used to reduce the dimension of the output vector. The test samples of fault types are separated and clustered successfully. Fig.6 shows the confusion matrix of CNN in identifying various deep faults in the test samples. The data on the symmetry axis represents the rate of judgment results consistent with

TABLE 3. Matrix Q of surface identification.

NO.	V_1	V_2	V_3	V_4	W_1	W_2	W_3	W_4	S_1	S_2
V_1	0.013327	-0.022734	-0.026588	-0.022480	-0.013327	-0.037500	-0.051046	-0.027914	-0.034881	-0.026588
V_2	0.001176	-0.016052	-0.030477	-0.017087	-0.001176	-0.022588	-0.033157	-0.016326	-0.022710	-0.020958
V_3	0.011208	-0.026623	-0.011208	-0.023770	-0.020398	-0.050678	-0.063647	-0.040568	-0.040711	-0.024677
V_4	0.005416	-0.016052	-0.027878	-0.017087	-0.005416	-0.027849	-0.041403	-0.018585	-0.027279	-0.021161
W_1	-0.024542	-0.035124	-0.058200	-0.039110	0.024542	-0.036096	-0.037802	-0.037701	-0.036166	-0.046591
W_2	-0.012619	-0.020441	-0.052384	-0.025447	0.012619	-0.019077	-0.019233	-0.020527	-0.020584	-0.032716
W_3	-0.024458	-0.029304	-0.063647	-0.037295	0.013827	-0.019077	-0.013827	-0.031965	-0.020172	-0.046504
W_4	-0.001426	-0.016052	-0.040668	-0.017087	0.001426	-0.019077	-0.032065	-0.016326	-0.022386	-0.020781
S_1	-0.009930	-0.020492	-0.042347	-0.024808	0.009930	-0.020514	-0.020172	-0.023922	-0.020172	-0.032317
S_2	0.008789	-0.017104	-0.024677	-0.017087	-0.008789	-0.031009	-0.046504	-0.020681	-0.030681	-0.020534

TABLE 4. Membership degree matrix of fuzzy-clustering algorithm.

Types \ Number	V_1	V_2	V_3	V_4	W_1	W_2	W_3	W_4	S_1	S_2
V	0.968943	0.197677	0.870251	0.898742	0.697398	0.049053	0.095986	0.092253	0.353772	0.903767
W	0.031057	0.802323	0.129749	0.101258	0.302602	0.950947	0.904014	0.907747	0.646228	0.096233

TABLE 5. Deep identification parameters.

Layers	Parameters
1	Input layer
2	Convolution layer: ReLU activation function (Convolution kernel: 5X5, step size: 1)
3	Pool layer (pool core: 2X2, step size: 2) Dropout coefficient: 1.0
4	Convolution layer: ReLU activation function (Convolution kernel: 5X5, step size: 1)
5	Pool layer (pool core: 2X2, step size: 2) Dropout coefficient: 1.0
6	Full connection: ReLU activation function Dropout coefficient: 1.0
7	Softmax layer
8	Output layer (output identification result)

TABLE 6. Deep identification results of some examples of test signal S_1.

Signal	No.	W_1	W_2	W_3	W_4	Results
S_1	1	0.7837	3.0211	-3.3638	-0.8301	W_2
	2	0.5765	2.6123	-2.1164	-1.3221	
	3	1.2619	3.4955	-1.6034	-3.2765	
	4	1.9045	4.6788	-3.3137	-3.5388	
	5	0.8030	3.2204	-3.0856	-1.3737	

TABLE 7. Deep identification results of some examples of test signal S_2.

Signal	No.	V_1	V_2	V_3	V_4	Results
S_2	1	0.8154	-1.3664	3.1315	-2.5584	V_3
	2	0.8408	-2.2687	3.0289	-1.7205	
	3	-0.3216	-2.4016	2.7114	-0.2183	
	4	-0.2251	-3.2842	4.6727	-1.3288	
	5	0.3446	-1.5654	4.0410	-2.7078	

TABLE 8. Identification accuracy of different methods.

Methods	Accuracy		Time(P./ms)
	Fault V	Fault W	
SVM	93.10%	93.80%	3.571
KNN	94.25%	95.42%	8.324
DT	84.68%	87.45%	1.036
CNN	96.50%	98.75%	1.749

the real category in the test set. The rest of the data in each row is the proportion of other types of data. Except for a few test samples that are misidentified, different deep fault types can be accurately identified. The identification accuracy of fault test set W is better than that of fault test set V. The experimental results show that the proposed method could be used to identify the cause of commutation failure.

To verify the effectiveness of the deep identification method, the proposed method is compared with the traditional machine learning methods such as support vector machine (SVM) [17], k-nearest neighbor (KNN) [29] and decision tree (DT) [30]. All methods share the same training process and testing samples, and Table 8 shows the identification accuracy.

It can be seen from Table 8 that the identification accuracy of the proposed method is higher than that of the traditional

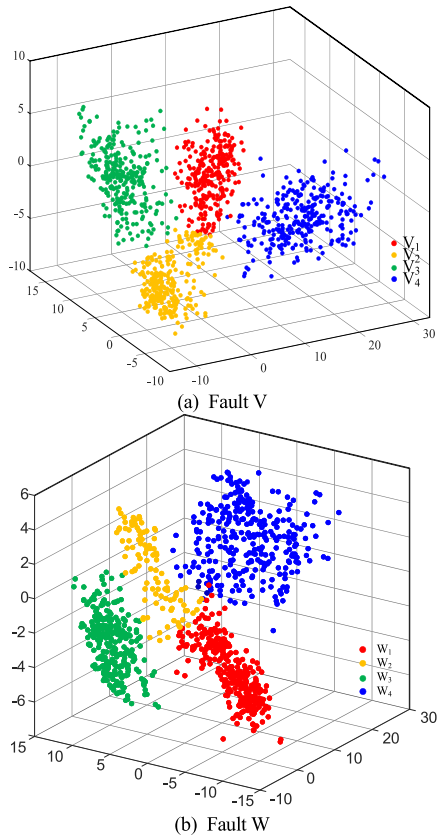

FIGURE 5. Three-dimensional feature visualization of deep identification outputs.

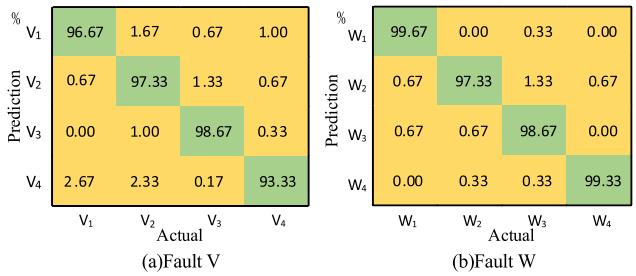

FIGURE 6. Confusion matrix of deep identification.

methods. The identification accuracy of converter valve faults and AC system faults are 96.50% and 98.75%, respectively. The identification time for single sample is in 2ms. For test samples, receiver operating characteristic (ROC) curves under different comparison methods can be drawn by taking false positive rate and true rate as abscissa and ordinate, respectively. According to the area under the ROC curves, the performance of identification method can be evaluated. The larger the area is, the better the performance is. Here, we take fault W test set as an example to draw ROC curves corresponding to different methods, as shown in Fig.7. The area under the ROC curves of the proposed deep identification method is the largest, which means the performance of deep identification method is the best.

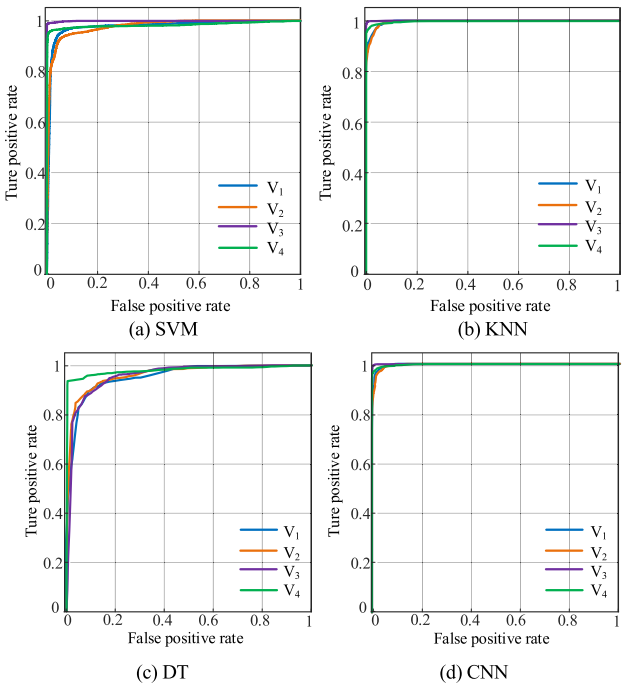

FIGURE 7. Curves of ROC for different methods.

TABLE 9. Input eigenvector of test samples.

Signals	E_5	E_6	E_7	E_8	E_9	E_{10}
$W_1(20\Omega)$	2953.96	120.08	89.251	218.20	30.042	7.856
$W_2(25\Omega)$	1436.34	285.05	396.52	54.485	39.232	-0.876
$W_3(20\Omega)$	1858.29	417.98	1379.11	78.783	58.776	18.743
$W_4(25\Omega)$	1948.64	582.81	1270.7	384.85	56.985	13.840

C. VERIFICATION OF COMPARISON METHOD

In this paper, taking the AC system fault simulation data of typical HVDC model as an example, the effectiveness of the proposed method is verified and compared to the identification method in [16]

According to the method in [16], the fault signal is decomposed and reconstructed by 10 layers based on db10 wavelet. The wavelet energy of each layer, that is, E_ε in equation (2), could be calculated. Wavelet energy from the fifth layer to the tenth layer were selected to form input eigenvector [E_5 E_6 E_7 E_8 E_9 E_{10}]. Some test signals and the eigenvector elements, from E_5 to E_{10}, are shown in Table 9. The contents of brackets in signals indicate the grounding resistances. For example, $W_1(20\Omega)$ means the W_1 fault type with 20Ω grounding resistance.

According to [16], the parameters and structure of the neural network settings were carried out, and the simulation example data used in the paper is trained. AC system faults with different grounding resistance are selected as test signals. The results are shown in Table10. In Table 10, the failure type of the test signals can be determined according to the label with the maximum value, which is with the same definition of failure and success of the proposed method.

TABLE 10. Results of method in [16].

Signals	W_1	W_2	W_3	W_4	Results
$W_1(5\Omega)$	0.2794	0.6552	0.0925	-0.0503	W_2
$W_1(10\Omega)$	0.5797	0.4521	0.1909	-0.1140	W_1
$W_2(5\Omega)$	0.0511	0.8559	-0.0297	0.0098	W_2
$W_2(10\Omega)$	0.1767	0.8767	-0.1500	0.1048	W_2
$W_3(5\Omega)$	0.0332	-0.0547	1.0035	-0.0252	W_3
$W_3(10\Omega)$	0.3063	-0.2857	0.7835	0.1443	W_3
$W_4(5\Omega)$	0.0098	-0.4593	1.0516	0.1301	W_3
$W_4(10\Omega)$	-0.2584	0.2530	0.1563	0.8026	W_4

TABLE 11. Identification accuracy of the method in [16] and our method.

Methods	Samples	Accuracy
WBPNN	Fault W	66.24%
	Fault V	63.58%
CNN	Fault W	98.75%
	Fault V	96.50%

The results of the method presented in [16] are shown in Table 10, where the label with the maximum value is determined as the failure type. As can be seen, the test samples W_2 and W_3 are identified correctly. However, the test sample $W_1(5\Omega)$ is misidentified as W_2, and the test sample $W_4(5\Omega)$ is misidentified as W_3. Although the test sample $W_1(10\Omega)$ is identified correctly, the label value of W_1 is 0.5797 and the label value of W_2 is 0.4521, that is, the difference between them is only 0.1276. In Tables 6 and 7, the proposed method can correctly identify all the failure types of the test samples S_1 and S_2. Moreover, the result distinction is also obvious. Taking the first sample of S_1 as an example, the difference between the label values of W_1 and W_2 is 2.2374.

There are two reasons behind failure of the method to detect W_1 and W_4 faults. First, we uses the AP algorithm to achieve surface classification, which could reduce the difficulty of the method for sample identification tasks. However, the AP algorithm is not used in [16]. It is easy to cause identification confusion between samples and lead to failure. Second, [16] used wavelet energy with back propagation neural network (WBPNN), which has a weak ability to extract sample features and has poor performance in identifying similar samples, so failures may occur. In contrast, we utilizes CNN to extract the features of the samples, which could make the method more sensitive to the features contained in the samples and clearly identify samples with similar features. The comparison of these method is given in Table 11, and the superiority of our method is demonstrated.

VII. CONCLUSION

In this paper, based on wavelet entropy AP clustering algorithm theory and CNN deep learning theory, a double-deck traceability identification method for HVDC system commutation failures fault reason is proposed, which is verified by simulation experiments and draws the following conclusions:

1) The surface identification method based on wavelet entropy fault feature space and AP algorithm can use DC voltage signal to quickly and effectively judge whether the fault signal belongs to converter valve fault or AC side fault, and complete shallow identification.

2) On the basis of surface identification, DC current, AC voltage and current signals are supplemented, and convolution neural network can be used to realize the deep identification of commutation failures and lock the real cause of commutation failures. This method has superior performance and high recognition rate.

3) The experimental verification in this paper takes the simulation data as the signal source. In practical application, the on-line monitoring data and fault recording signals can be used to accurately determine the specific fault types of commutation failures and expand the identification database. How to apply the proposed method in HVDC system and extend the potential fault identification in control system need to be further studied.

REFERENCES

[1] Q. Wang, C. Zhang, X. Wu, and Y. Tang, "Commutation failure prediction method considering commutation voltage distortion and DC current variation," *IEEE Access*, vol. 7, pp. 96531–96539, 2019.
[2] Z. Wei, Y. Yuan, X. Lei, H. Wang, G. Sun, and Y. Sun, "Direct-current predictive control strategy for inhibiting commutation failure in HVDC converter," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2409–2417, Sep. 2014.
[3] J. Wu, H. Li, G. Wang, and Y. Liang, "An improved traveling-wave protection scheme for LCC-HVDC transmission lines," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 106–116, Feb. 2017.
[4] Y. Yang and P. Zhang, "Study on the influence of inconsistent valve parameters on LCC-HVDC commutation and operation," *IEEE Access*, vol. 7, pp. 109015–109025, 2019.
[5] Y. Shao and Y. Tang, "Fast evaluation of commutation failure risk in multi-infeed HVDC systems," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 646–653, Jan. 2018.
[6] Y. Xue and X.-P. Zhang, "Reactive power and AC voltage control of LCC HVDC system with controllable capacitors," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 753–764, Jan. 2017.
[7] J. Ouyang, Z. Zhang, T. Tang, M. Pang, M. Li, and D. Zheng, "Fault overload control method for high-proportion wind power transmission systems based on emergency acceleration of doubly-fed induction generator," *IEEE Access*, vol. 7, pp. 32874–32883, 2019.
[8] C. V. Thio, J. B. Davies, and K. L. Kent, "Commutation failures in HVDC transmission systems," *IEEE Trans. Power Del.*, vol. 11, no. 2, pp. 946–957, Apr. 1996.
[9] G. Li, S. Zhang, T. Jiang, H. Chen, and X. Li, "A method of detecting commutation failure in multi-infeed HVDC systems based on critical failure impedance boundary," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2017, pp. 1–5.
[10] H. Xiao, Y. Li, J. Zhu, and X. Duan, "Efficient approach to quantify commutation failure immunity levels in multi-infeed HVDC systems," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 4, pp. 1032–1038, Mar. 2016.
[11] S. Mirsaeidi, X. Dong, D. Tzelepis, D. M. Said, A. Dysko, and C. Booth, "A predictive control strategy for mitigation of commutation failure in LCC-based HVDC systems," *IEEE Trans. Power Electron.*, vol. 34, no. 1, pp. 160–172, Jan. 2018.
[12] W. Yao, C. Liu, J. Fang, X. Ai, J. Wen, and S. Cheng, "Probabilistic analysis of commutation failure in LCC-HVDC system considering the CFPREV and the initial fault voltage angle," *IEEE Trans. Power Del.*, vol. 35, no. 2, pp. 715–724, Apr. 2020.

[13] Y. Xue, X.-P. Zhang, and C. Yang, "Commutation failure elimination of LCC HVDC systems using thyristor-based controllable capacitors," *IEEE Trans. Power Del.*, vol. 33, no. 3, pp. 1448–1458, Jun. 2018.
[14] X. Chen, A. M. Gole, and M. Han, "Analysis of mixed inverter/rectifier multi-infeed HVDC systems," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1565–1573, Jul. 2012.
[15] E. Rahimi, A. M. Gole, J. B. Davies, I. T. Fernando, and K. L. Kent, "Commutation failure analysis in multi-Infeed HVDC systems," *IEEE Trans. Power Del.*, vol. 26, no. 1, pp. 378–384, Jan. 2011.
[16] C. Liu, F. Wang, F. Zhuo, and Z. Zhang, "Fault diagnosis of HVDC transmission system using wavelet energy entropy and the wavelet neural network," in *Proc. 22nd Eur. Conf. Power Electron. Appl. (EPE ECCE Eur.)*, Sep. 2020, pp. 1–8.
[17] C. Gao, Z. Liao, and S. Huang, "Fault diagnosis of commutation failures in the HVDC system based on wavelet singular value and support vector machine," in *Proc. Asia–Pacific Power Energy Eng. Conf.*, Mar. 2009, pp. 1–4.
[18] L. Lin, Y. Zhang, Q. Zhong, and F. Wen, "Identification of commutation failures in HVDC systems based on wavelet transform," in *Proc. Int. Conf. Intell. Syst. Appl. Power Syst.*, Nov. 2007, pp. 1–5.
[19] L. Zhiwei and L. Long, "Research on fault diagnosis of HVDC commutation failure," in *Proc. IEEE Int. Conf. High Voltage Eng. Appl. (ICHVE)*, Sep. 2016, pp. 1–4.
[20] C. Wang, C. Zhang, X. Kong, P. Li, and Y. Yuan, "Procedure analysis of UHVDC commutation failure," *J. Eng.*, vol. 2019, no. 16, pp. 3132–3134, Mar. 2019.
[21] Y. Deng, S. Lin, L. Fu, K. Liao, L. Liu, Z. He, S. Gao, and Y. Liu, "New criterion of converter transformer differential protection based on wavelet energy entropy," *IEEE Trans. Power Del.*, vol. 34, no. 3, pp. 980–990, Jun. 2019.
[22] Z. He, L. Fu, S. Lin, and Z. Bo, "Fault detection and classification in EHV transmission line based on wavelet singular entropy," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2156–2163, Oct. 2010.
[23] F. Ling, H. Zhengyou, and B. Zhiqian, "A novel algorithm for power fault diagnosis based on wavelet entropy and D-S evidence theory," in *Proc. 43rd Int. Universities Power Eng. Conf.*, Sep. 2008, pp. 1–4.
[24] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[25] S. Subedi, H.-S. Gang, N. Y. Ko, S.-S. Hwang, and J.-Y. Pyun, "Improving indoor fingerprinting positioning with affinity propagation clustering and weighted centroid fingerprint," *IEEE Access*, vol. 7, pp. 31738–31750, 2019.
[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[27] L. Zhu, D. J. Hill, and C. Lu, "Hierarchical deep learning machine for power system online transient stability prediction," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2399–2411, May 2020.
[28] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: http://arxiv.org/abs/1207.0580
[29] S. Naik and E. Koley, "Fault detection and classification scheme using KNN for AC/HVDC transmission lines," in *Proc. Int. Conf. Commun. Electron. Syst. (ICCES)*, Jul. 2019, pp. 1131–1135.
[30] J. Yan, C. Li, and Y. Liu, "Insecurity early warning for large scale hybrid AC/DC grids based on decision tree and semi-supervised deep learning," *IEEE Trans. Power Syst.*, early access, Apr. 8, 2021, doi: 10.1109/TPWRS.2021.3071918.

YUHONG WANG (Member, IEEE) received the B.S. and M.S. degrees in power system and its automation from Chongqing University, Chongqing, China, in 1993 and 1995, respectively, and the Ph.D. degree in power system and its automation engineering from Southwest Jiaotong University, Chengdu, China, in 2008. She is currently a Professor with Sichuan University.

Her research interests include power system stability and control, HVDC and FACTS, renewable energy integration, and artificial intelligence.

KEQIANG TAI received the B.S. degree in power system and its automation engineering from the Taiyuan University of Technology, Taiyuan, China, in 2017. He is currently pursuing the M.S. degree with Sichuan University, Chengdu, China.

His current research interests include HVDC system stability, power system operation and control, and artificial intelligence.

YUYAN SONG (Student Member, IEEE) was born in Xinjiang, China, in 1994. She received the B.S. degree in power system and its automation engineering from Sichuan University, in 2017. She is currently pursuing the M.S. and Ph.D. degrees in electrical engineering with Sichuan University, Chengdu, China.

Her research interests include power system stability and control, HVDC and FACTS, and renewable energy integration.

RAN KOU received the bachelor's degree from Sichuan University, Chengdu, China, in 2016, where he is currently pursuing the M.S. degree.

His current research interests include power system operation and control and HVDC system stability analysis.

ZONGSHENG ZHENG (Member, IEEE) received the B.S. degree in bioinformatics and the Ph.D. degree in electrical engineering from Southwest Jiaotong University, Chengdu, China, in 2013 and 2020, respectively. From 2018 to 2019, he was a Visiting Scholar with the Virginia Tech-Northern Virginia Center, Bradley Department of Electrical and Computer Engineering, Falls Church, VA, USA. He is currently an Associate Research Fellow with the College of Electrical Engineering, Sichuan University.

His research interests include uncertainty quantification, parameter, and state estimation.

QI ZENG received the B.S. and M.S. degrees in power system and its automation from Sichuan University, Chengdu, China, in 2000 and 2003, respectively, and the Ph.D. degree in power system and its automation engineering from Sichuan University, Chengdu, China, in 2017. She is currently an Associate Professor with Sichuan University.

Her research interests include power system stability and control and HVDC.