IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 10, 2021, accepted February 24, 2021, date of publication February 26, 2021, date of current version March 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062703

Single Pole-to-Ground Fault Location System for MMC-HVDC Transmission Lines Based on Active Pulse and CEEMDAN

JIAN-YU WU[1], SHENG LAN[2], (Member, IEEE), SI-JIE XIAO[2], AND YONG-BIN YUAN[2]
[1]Fuzhou Minjiang Park Office, Fuzhou 350002, China
[2]College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
Corresponding author: Sheng Lan (lansheng@fzu.edu.cn)

ABSTRACT To solve the problem of the location of the fault point of single-pole-to-ground faults in the transmission lines of MMC-HVDC systems, this paper designs a fault location system based on support vector machine (SVM). The waveform of the traveling wave after the fault occurs is collected as a feature, and the regression mechanism of the SVM is utilized to achieve fault location. Because it is very difficult to locate high-resistance ground faults, this paper first analyzes the waveform characteristics of high-resistance ground faults. Next, three steps are proposed to reduce the influence of grounding resistance on fault location. These steps include using the active pulse waveform as a new feature, classifying the samples according to ground resistance values before training regression models, and a method for adaptively extracting fault distance features is proposed. Finally, a complete location system design is proposed, and its workflow is illustrated. After the simulation test, the proposed location system only needs to obtain a single-ended fault voltage waveform at a fault recording frequency of 20 kHz to achieve an accurate location of single-pole-to-ground faults for different values of grounding resistance.

INDEX TERMS Active pulse, fault location, high-resistance fault, MMC-HVDC system, SVM, single-end measurement.

I. INTRODUCTION

Modular multilevel converter (MMC) technology has gradually replaced two-level or three-level voltage source converters (VSCs) as a new direction for the development of flexible DC transmission technology. The MMC-HVDC system has many advantages, such as a low switching frequency, low operating loss, high efficiency, and high reliability. It has broad application prospects in the fields of new energy generation through grid connections, urban distribution network capacity expansion, and long-distance load power supply [1], [2].

The transmission lines of MMC-HVDC are generally long and pass through various complicated terrains along the way, which tends to result in various line faults. It is usually built in the wilderness and exposed to unpredictable environment, so its failure rate will be higher than other power systems [3]. In general, disconnection faults and pole-to-pole short-circuit faults are usually caused by external mechanical stresses, usually producing permanent faults. Ensuring the safe and reliable operation of the transmission line will ensure the stable operation of the system and the power system in neighboring areas and even the entire grid skeleton [4]. Therefore, once a fault is detected, the converter stations at both ends are isolated and go out of operation. However, when a single-pole ground fault with a small current ground occurs, the converter station is not isolated, and the line can continue to transmit power [5]. Therefore, when a single-pole ground fault occurs in the MMC-HVDC system, if the fault point can be quickly and accurately found, the fault can be eliminated in time, the line can be repaired, and normal operation of the system can be resumed in a short time.

At present, fault location in HVDC transmission lines is mainly based on the traveling wave theory. The traveling wave method can be divided into single-ended method and double-ended method. In general, the measurement accuracy and reliability of the double-ended methods are higher than those of the single-ended method, but the double-end method needs to maintain recording time synchronization at both

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

ends of the line [7], which increases equipment cost and technical requirements.

Although the traveling wave method is currently the most mainstream line fault detection and line fault location method, the traveling wave method also has some inherent defects [8]. The traveling wave method has some inherent defects. The accuracy of fault location depends on the sampling frequency of the fault recorder. Whether the arrival time of the traveling wave can be accurately detected is the key to successful location [9]. Accurate location using the traveling wave method also relies on accurate measurement of the traveling wave velocity, which is not easy to implement in the MMC-HVDC transmission line, especially in the case where the transmission line is a cable [10]. In addition, the influence of the grounding resistance cannot be ignored. In the event of a high-resistance ground fault, the transient signal decays faster, and the fault characteristics are not clear, which leads to a large location error.

In response to the inherent flaws of the traveling wave method, scholars have proposed different solutions. Reference [11] proposed a method for determining fault points based on natural frequency values. This method does not need to detect the arrival time of the surge, thus reducing the requirements for the sampling frequency, but the influence of the grounding resistance on the frequency component cannot be avoided. Liang *et al.* proposed a nonsynchronous double-ended fault location method based on Bergerian theory [12], which can achieve accurate ranging without double-end time synchronization. According to the relationship between the natural frequency value and the distance to the fault and the traveling wave speed. A single-ended positioning method is proposed to avoid the requirement of high sampling frequency [13]. Bi *et al.* proposed an inactive pulse generation method based on MMC-HVDC submodule control [14]. The continuous pulse traveling wave generated by such artificial control is less affected by the grounding resistance than the fault traveling wave, but its measurement accuracy still depends on the arrival time of the high sampling rate detection pulse.

In recent years, some scholars have tried to use various intelligent algorithms to implement line fault location as a pattern recognition problem. Farshad and Sadeh [15] proposed to realize fault location based on k-nearest neighbor classification and discrete Fourier transform, but this method is not suitable for current transformer saturation. Reference [16] used a neural network for fault location. Lan *et al.* [17] proposed a high-voltage DC fault location method based on a one-dimensional convolutional neural network. However, CNN model training requires large-scale training samples, and the accuracy of the method is not verified at low sampling frequencies. SVM has good performance in fault detection and location of transmission lines because of its advantages in handling small-scale samples at high latitudes [18]. Reference [19] proposed a method of grounding fault detection using wavelet decomposition combined with convolutional neural network. Johnson and Yadav [20] used the support vector machine (SVM) regression mechanism to achieve high location accuracy when the grounding resistance factor is not considered. The accuracy of location can be improved by improving the performance of the SVM [21] or optimizing the process of manually extracting features in combination with other algorithms [22].

This paper analyzes the waveform characteristics of the voltage fault traveling wave and the active pulse proposed in [14] when different grounding resistance fault values occur. In this paper, a single-ended comprehensive fault location system based on the regression mechanism of SVM is proposed. According to the fault characteristics of the waveform, this paper proposes a method for extracting distance features using phase-mode transformation and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) decomposition. Then, SVM classification and regression mechanisms are used to achieve fault location, and finally, a comprehensive system architecture suitable for practical engineering is designed. In this paper, an MMC-HVDC single pole-to-ground fault location method based on active pulses and CEEMDAN is proposed. It is designed to achieve accurate location of fault points when single-pole ground faults with different fault grounding resistance values occur. In view of the various links in the training process of the SVM regression model, this paper proposes different methods to reduce the influence of grounding resistance on fault location. This paper uses PSCAD/EMTDC software to establish a 200-level MMC-HVDC model to simulate faults and obtain the faulty pole-to-ground voltage waveforms. The simulation proves that the system can accurately locate the ground fault point when the fault recording frequency is 20 kHz, and the average error of the ranging is controlled within 0.5 km.

II. RESEARCH ON FAULT WAVEFORM CHARACTERISTICS

A. STRUCTURE TOPOLOGY OF THE MMC-HVDC SYSTEM

The bipolar MMC-HVDC structure is shown in Fig. 1. The converter station has three phases, each phase consisting of an upper arm and a lower arm. Each phase unit is composed of 2N cascaded submodules, and the upper and lower bridge arms each have N cascaded submodules. In addition, a bridge arm reactance is connected in series with each bridge arm. Most MMC-HVDC systems that have been put into operation generally use the DC-side grounding method, that is, the grounding method of the DC-side clamped large resistor or clamp capacitor, as shown in Fig. 1.

The submodule is the basic component of MMC-HVDC. Without considering the redundant submodule, there are 2N cascade submodules for each phase and N for each of the upper and lower arms. The submodule of the half-bridge structure is shown in Fig. 1. T1 and T2 represent IGBTs, D1 and D2 represent antiparallel diodes, and C represents a DC-side capacitor of the submodule. The submodule in the working state has two working states of input and bypass. As shown in Fig. 2, controlling the turn-on and turn-off of IGBT1 and IGBT2 can switch the working state of

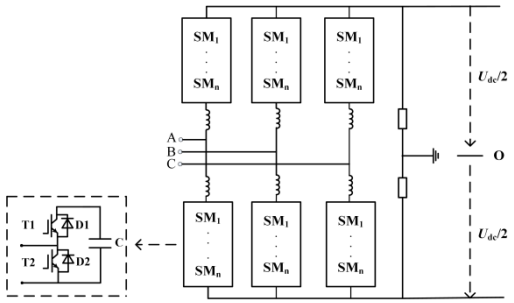

FIGURE 1. Topology of MMC-HVDC.

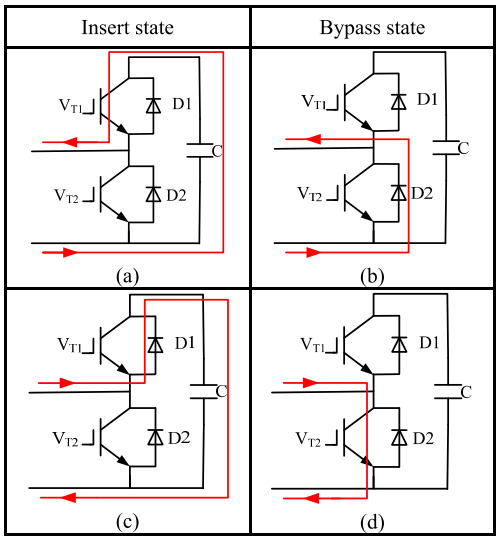

FIGURE 2. Two states of SM. (a) Insert state. (b) Bypass state. (c) Insert state. (d) Bypass state.

the submodule. The input state is shown in Fig. 2(a)(c). At this time, T1 is turned on, T2 is turned off, current flows through the submodule capacitor, and the output voltage of the submodule is the capacitor voltage U_c. The charging or discharging of the submodule capacitor depends on the flow direction of the submodule current. The bypass state is shown in Fig. 2(b)(d). At this time, T1 is turned off, T2 is turned on, the current does not flow through the submodule capacitor, and the submodule output voltage is zero.

B. PRINCIPLE OF ACTIVE PULSE GENERATION

If several submodules in the input state are converted to the bypass state and then restored to the input state in a short time, a voltage drop occurs in the bridge arm for a short period of time. This voltage drop propagates along the transmission line, which is equivalent to sending a low-voltage pulse down the line. This process can be achieved by changing the T1 and T2 control signals. For example, to cut off several submodules of the upper arm, at time t1, the control signal of the submodule that needs to be cut is changed to a bypass signal and is restored to the input signal at time t2. The series submodule voltage on this bridge arm appears as shown in Fig. 3(a).

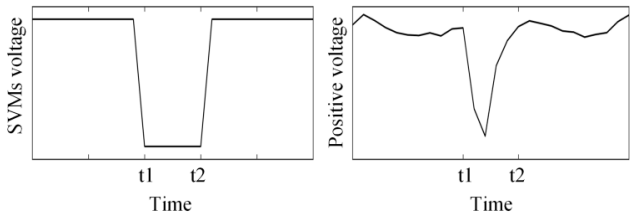

FIGURE 3. Production of the active pulse. (a) SM voltage drop. (b) Line voltage drop.

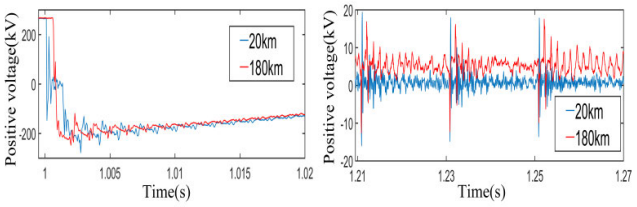

FIGURE 4. Waveform of different fault points.

The active pulse actually propagating in the line is shown in Fig. 3(b).

C. COMPARISON OF FAULT WAVEFORM AND ACTIVE PULSE WAVEFORM

A positive ground fault occurs at 1 s, and an active pulse occurs at 1.2 s. Faults are simulated at different fault points and different grounding resistances, and the voltage fault waveform of the fault pole is recorded, as shown in Fig. 4 and Fig. 5. The fault point position feature mainly mentions the position of the extreme value of the high-frequency component in the waveform, and the different grounding resistance values do not change the position where the extreme value occurs. For fault voltage waveforms, the grounding resistance has a greater impact on the overall waveform than the fault point distance. As the grounding resistance increases, the decay rate of the voltage waveform slows down, and the high-frequency components in the waveform decrease. Different transition resistances and different fault locations are two important factors that affect the fault traveling wave. How to reduce the impact of transition resistance factors on the fault voltage waveform to better extract the fault location characteristics from the waveform has become the key to accurate fault location. Although the fault pole voltage drops to 0 V, due to the effect of the traveling wave process of the transmission line, the active pulse at the time of the fault contains abundant frequency components and has an obvious transient process. Therefore, fault detection based on the characteristics of the active pulse can be deeply explored to achieve accurate and reliable unipolar ground fault location. In contrast, the grounding resistance has a relatively small influence on the waveform of the active pulse, so the waveform difference can reflect more fault point distance characteristics. Even so, the effect of the grounding resistance cannot be ignored.

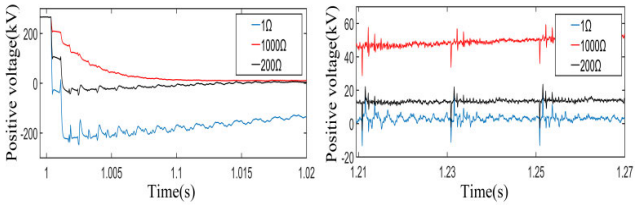

FIGURE 5. Waveform of different grounding resistance values.

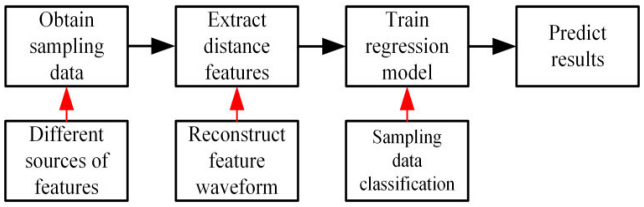

FIGURE 6. Training process and optimization methods of the SVM regression model.

III. FAULT LOCATION METHOD

From the analysis in the previous part, it can be seen that the fault distance feature is contained in the high-frequency component of the fault traveling wave, so this part will perform deep location feature extraction for the active pulse. Therefore, an MMC-HVDC single pole-to-ground fault location method based on active pulses and CEEMDAN is proposed. This paper will optimize the three key steps in the process. First, phase-to-mode transformation is used to decouple the fault waveform and extract the line-mode component. Then, each fault line modulus component is decomposed through the CEEMDAN algorithm, and the characteristic waveform of the decomposition result is reconstructed. Finally, the reconstructed waveform is sent to the SVM regression model to obtain the final fault location result. It should be noted that the fault voltage waveforms with large differences in transition resistance values have great differences. If the feature vectors with large differences in resistance values are directly sent to the SVM for training, the positioning results will be seriously interfered with by transition resistance factors. Therefore, prior to fault location, the feature vector is classified according to a certain range of transition resistance to reduce the influence of transition resistance factors on the location result.

In this section, the method of extracting the distance feature after feature waveform reconstruction using phase-mode transformation and CEEMDAN is described in detail, and the working principle of the support vector machine is briefly described. The conventional process of regression prediction is shown in Fig. 6.

A. PHASE-MODE TRANSFORMATION

For actual large-scale bipolar high-voltage direct current transmission lines, most of the time is spent in the bipolar operating state, and the coupling between the bipolar transmission lines cannot be ignored. The coupling phenomenon is especially noticeable when the line fails. Therefore, the phase mode conversion must be performed on the line before using the traveling wave information. The phase-mode transformation in the transmission line refers to the introduction of the decoupling matrix, i.e., Eq. (1) into the differential equation of the transmission line, i.e., Eq. (2) [23]. Therefore, the original voltage signal is decoupled into a line mode component and a ground mode component. The ground mode component is greatly affected by the geographical environment and frequency, while the line mode component is less affected by the frequency, so the line mode component is more stable than the ground mode component. In this paper, the collected voltage is first subjected to phase-mode transformation, and the linear modulus component is extracted as the preliminary extraction of the traveling wave characteristics.

$$S = \begin{cases} \dfrac{\sqrt{2}}{2} \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} \\ S^{-1} = S^T \end{cases} \tag{1}$$

$$\begin{aligned} \frac{\partial u}{\partial x} &= -L\frac{\partial i}{\partial t} - Ri \\ \frac{\partial i}{\partial x} &= -C\frac{\partial u}{\partial t} - Gi \end{aligned} \tag{2}$$

B. EMPIRICAL MODE DECOMPOSITION AND ITS IMPROVED ALGORITHM

Empirical mode decomposition is a decomposition method proposed by Huang to address nonlinear and nonstationary signals. The method adaptively decomposes the original signal into multiple intrinsic mode functions (IMFs) from high frequency to low frequency. Compared with other signal processing methods, such as wavelet analysis, EMD decomposition does not need to set the basis function in advance, which overcomes the characteristics of relying on subjective experience. The distance feature is mainly reflected in the extreme point position of the waveform, so use of the EMD method is considered to remove the residual of the original waveform decomposition and several low-frequency components, highlighting the extreme point characteristics of the waveform.

However, when dealing with intermittent signals, pulse signals, and noise-containing signals, modal aliasing is prone to occur. Modal aliasing refers to the feature time scales that contain extreme differences in an IMF or similar feature time scales being distributed in different IMFs, resulting in aliasing of adjacent IMF waveforms, mutual influence, and illegibility.

Ensemble empirical mode decomposition (EEMD) improves the EMD method by eliminating the effects of modal aliasing in EMD by adding auxiliary white noise multiple times. On the basis of EEMD, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and with adaptive white noise is proposed, which adds adaptive white noise at each stage of decomposition and obtains each IMF by calculating a unique residual signal. Compared with the EEMD method, regardless of the number

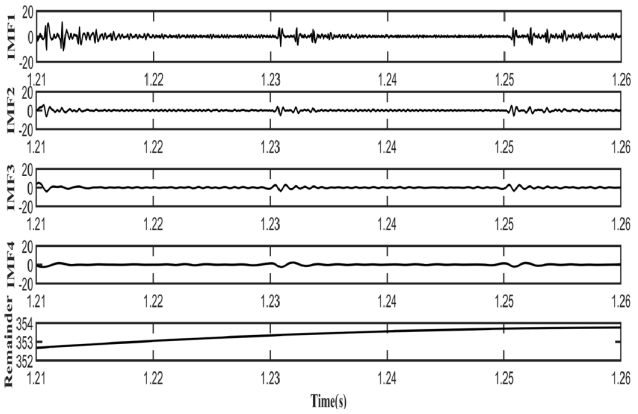

FIGURE 7. CEEMDAN decomposition results.

of ensembles, the reconstruction error is almost zero, and the decomposition process has completeness, which overcomes the problem of the low efficiency of EEMD decomposition [24], [25]. The main difference between EMD and its improved algorithm is reflected in the method of adding auxiliary noise. The basic structure can be expressed as Eq. (3), where r(t) is the residual.

$$s(t) = \sum_{i=1}^{n} imf_i(t) + r(t) \tag{3}$$

Taking a single-pole grounding fault of 100Ω occurred at 100km from the rectifier side as an example, the result of CEEMDAN decomposition of the fault voltage waveform is shown.

C. SUPPORT VECTOR MACHINE

Support vector machine is a supervised machine learning model based on VC dimension theory and the structural risk minimization principle. It has unique advantages in solving small-sample, nonlinear or high-dimensional problems and is often used for high-dimensional pattern recognition. As a mature algorithm, support vector machine is widely used in classification and regression problems in various fields.

SVM is essentially a linear classifier. However, in solving practical problems, most of the problems we encounter are linear indivisible problems. For nonlinear cases, the SVM processing method is used to select a kernel function by mapping the data to a high-dimensional space [26]. Then, the optimal classification surface is found to solve the problem of linear inseparability in the original space, as shown in Fig. 8. The Gaussian kernel function is the most widely used kernel function, and it can achieve ideal results in various problems. While using the Gaussian RBF kernel, there are two parameters that can be varied: c and g, where parameter c is the penalty factor, which is the tolerance for the error. The parameter g is associated with the RBF function as the selected kernel, which reflects the distribution of the support vector after the data are mapped to the new feature space.

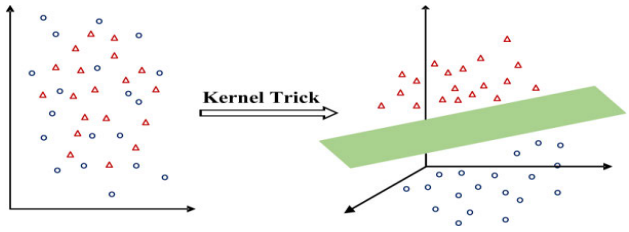

FIGURE 8. Kernel trick.

Finding the optimal c and g parameters is required when using the Gaussian RBF kernel.

The method of support vector regression was proposed by Drucker. Its goal is to find a y_1 function f(x) that has the greatest deviation ϵ from the actual target a for all training data [27], [28]. Given a set of training samples, (x_1, y_1), $1 = 1, \ldots, 1$, where the input is $\curvearrowleft \in \mathbb{R}^n, m_t \in \mathbb{R}^n$ is the target output, and the standard form of support vector regression is shown by Eqs. (4–7) [29].

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i + C \sum_{i=1}^{l} \xi_i^* \tag{4}$$

$$\text{s.t.} \left[w^T \phi(x_i) + b - y_i \right] \leq \epsilon + \xi_i \tag{5}$$

$$y_i - w^T \phi(x_i) - b \leq \epsilon + \xi_i^* \tag{6}$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots, l \tag{7}$$

In this paper, a single-pole ground fault of 0-1000 Ω is simulated. The collected fault waveforms are divided into four classes according to the grounding resistance, and a regression model is trained in each class. The classification criteria are shown in Table 1. It is necessary to classify the samples before training the regression model. The waveform shapes in the same class are similar, which avoids the under fitting problem caused by the large difference in waveforms during training and makes the determination of c and g parameters more accurate. Since the grounding resistance span of different samples is large, after performing CEEMDAN decomposition, the same IMF component may be a waveform with a large frequency difference. Through classification, the same IMF component frequency is not much different after waveform decomposition in the same class, so the number of IMFs in the decomposition or reconstruction process can be selected according to the waveform characteristics of each type. To increase the fault tolerance of the classification and better adapt to the actual situation that may occur in the actual project, the sampling ranges of adjacent categories overlap when training the regression model, as shown in Table 2. Taking a single-pole grounding fault with a transition resistance value of 30 Ω as an example, after extracting the characteristics of the fault traveling wave, the characteristic vector is sent to the SVM classification model. Although the resistance value is in the range not covered by each category, the output category may be classified as Class 1 or Class 2 in the model, but the SVM regression models corresponding to these two categories have learned the characteristic

TABLE 1. Classification of training samples.

Class	Classification resistance/Ω	Regression resistance/Ω
Class 1	0.1, 1, 10, 20	0.1, 1, 10, 20, 50
Class 2	50, 100, 150, 200	20, 50, 100, 150, 200
Class 3	250, 350, 450, 500	200, 250, 350, 450, 500
Class 4	600, 700, 800, 1000	500, 600, 700, 800, 1000

TABLE 2. Detailed parameters of MMC-HVDC system.

Parameter	Value
Rated DC pole voltage/kV	±250
Rated power/MW	400
DC line length/km	200
Number of SMs per arm	200
Capacitance of each SM/mF	10
Arm inductance/m H	50
Neutral grounding resistance/k Ω	100
Winding line voltage/kV	380/220
Winding connection	Yn/D
Number of cutoff SMs	50
Pulse duration time/m s	0.1

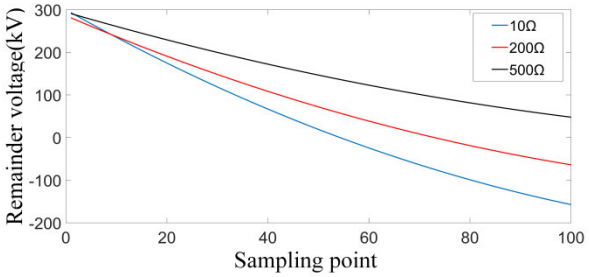

FIGURE 9. Remainder voltage.

information of Class 1 or Class 2 in the regression model containing the resistance value, so accurate fault location can still be achieved.

D. RECONSTRUCTION OF CHARACTERISTIC WAVEFORM

For the DC voltage of the fault pole, when CEEMDAN is performed, the waveform of the residual decreases linearly in the early stage, and the falling speed slows down as the grounding resistance increases, as shown in Fig. 9. For a fault voltage waveform $s(t)$, the falling value of the residual waveform of 100 sample points is defined as d. The average value of d of the lowest ground resistance samples in the class of $s(t)$ is calculated, and this average value is defined as D. The correction factor h is defined according to Eq. (8).

$$h = \frac{D}{d} \tag{8}$$

h is generally a real number greater than 1, and each fault voltage waveform can calculate an h value. In the same class, the larger the ground resistance value is, the larger the h value. The purpose of introducing h is to highlight the extreme value of the waveform that is attenuated due to the increase in

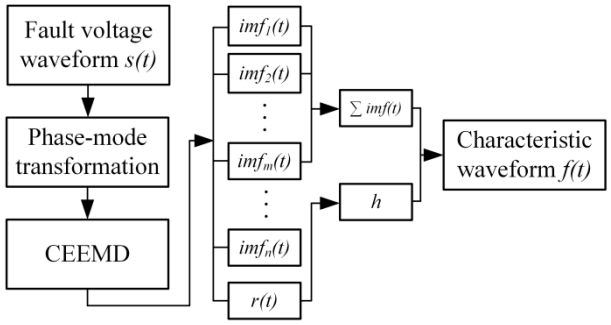

FIGURE 10. Characteristic waveform construction process.

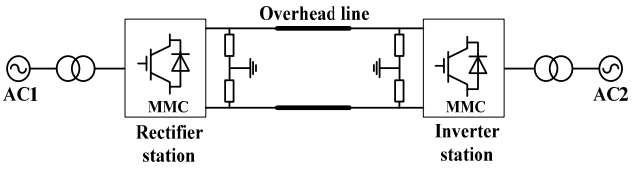

FIGURE 11. MMC-HVDC system model.

grounding resistance. In addition, because the samples used in the training model cannot include all grounding resistance values, h can also improve the generalization performance of the regression model to some extent. According to the decomposition result of CEEMDAN, several high-frequency IMFs are selected accordingly to obtain the characteristic waveform $f(t)$. The complete feature waveform reconstruction process is shown in Fig. 10.

$$f(t) = h * \sum_{i=1}^{m} imf_i(t) \tag{9}$$

IV. SIMULATION EXPERIMENT RESULTS

A. MMC-HVDC SYSTEM SIMULATION MODELS AND PARAMETERS

In the PSCAD/EMTDC simulation environment, a 200-level MMC-HVDC system based on NLM modulation is built according to the graph topology. The system parameters are shown in Table 2. The system topology is shown in Fig. 11. The neutral point grounding method used in this system is DC side clamp resistor grounding. The rectifier side adopts a constant DC voltage and constant reactive power control, and the inverter side adopts a fixed active power and fixed reactive power control. The structure and parameters of the rectifier station and the inverter station are the same. The length of the transmission line is 200 km. To simulate long-distance DC transmission more accurately, the DC line in the simulation adopts the frequency-variable overhead line model.

The fault waveform used in this paper was obtained by simulation in PSCAD with a sampling frequency of 20 kHz. The single-pole ground fault below refers to the positive ground, and the generation of the active pulse is based on the removal of the submodule from the upper arm of the A phase. The fault distance refers to the distance from the fault point to the rectifier end. Therefore, 16 faults with different grounding

TABLE 3. Location results of different waveforms.

Class	Fault voltage								Active pulse							
	s(t)				*f(t)*				*s(t)*				*f(t)*			
	Maximum error/km/%		Average error/km/%		Maximum error/km/%		Average error/km/%		Maximum error/km/%		Average error/km/%		Maximum error/km/%		Average error/km/%	
Class 1	0.7963	0.4%	0.2019	0.1%	0.8312	0.42%	0.1723	0.09%	0.6408	0.32%	0.1520	0.08%	0.7681	0.38%	0.1636	0.08%
Class 2	1.2535	0.63%	0.5067	0.25%	1.4123	0.71%	0.3145	0.16%	0.9298	0.47%	0.3142	0.16%	0.9532	0.48%	0.2093	0.11%
Class 3	3.5681	1.78%	0.9613	0.48%	1.3566	0.69%	0.4443	0.22%	1.2735	0.64%	0.6345	0.32%	1.3002	0.65%	0.4111	0.21%
Class 4	4.9986	2.5%	1.1213	0.56%	1.4512	0.73%	0.4102	0.21%	1.4103	0.71%	0.6098	0.31%	1.2513	0.63%	0.4672	0.23%
Not classified	9.7969	4.9%	4.2152	2.1%	9.0236	4.52%	4.1333	2.07%	5.4464	2.72%	2.7419	1.37%	4.0256	2.01%	1.7865	0.89%

resistance values were simulated at each fault point, and a total of 3,184 original fault waveforms were obtained.

A single-pole ground fault is simulated every 1 km on the transmission line. The grounding resistance value is set as shown in Table 1. A single pole ground fault occurs at 1 s. At 1.3 s, active pulses are emitted by the rectifier, with a period of 0.2 s. The sampling window has a length of 200 sampling points, which is composed of the first 5 sampling points at the moment that the fault occurs and the following 194 sampling points.

After the original waveform is obtained, it is reconstructed into a characteristic waveform. In the process of CEEMDAN, the amplitude of the white noise added is 0.2, which is added 50 times, and the maximum number of iterations is 100. The first five high-frequency IMFs and the remainder are extracted for feature waveform reconstruction.

B. SAMPLE CLASSIFICATION

The SVM classification mechanism is used to train the classification model, and the fault voltage waveform is directly used as a feature. The classification criteria are shown in Table 1. A total of 3184 sample waveforms were used in the training process. 80% of the total samples were used to train the model, and 20% were used to test the classification effect. The test results are shown in the confusion matrix in Fig. 12. The "1", "2", "3", and "4" in the figure correspond to each category in the sample classification of Section3. The ordinate represents the actual resistance category corresponding to the sample, and the abscissa represents the predicted resistance category. After calculation, the classification accuracy is 93.216%, and the classification error only occurs in the adjacent class.

C. FAULT LOCATION

The SVM regression mechanism is used to train the classification model, and four regression models are trained according to the different types of training samples. A total of 995 sample waveforms were used in a single model training process. 80% of the total samples was used to train the model, and 20% was used to test the location effect. The location effects of the fault voltage waveform and the active pulse waveform are tested, and the reconstructed characteristic waveform is simultaneously tested as

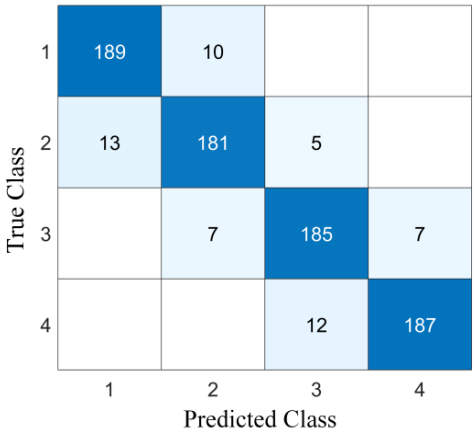

FIGURE 12. Classification result.

a feature location effect. The maximum absolute error and the average absolute error in the regression results of the test set are calculated separately, as shown in Table 3. The calculation formula of the maximum absolute error and the average absolute error is shown in Eq. (10) (11).

$$\text{Maximum error} = \max |pred_i - W_i| \tag{10}$$

$$\text{Average error} = \frac{\frac{1}{N}\sum_{i=0}^{N} |pred_i - W_i|}{L} \tag{11}$$

where W_i is the actual fault location, pre_i is the model location result, L represents the total length of the line, the value is 200km, and N is the number of location samples to be tested.

According to the data in the analysis table, for a fault with extremely low grounding resistance, the fault voltage waveform can be used as a feature to obtain an ideal location effect. Reconstruction of the characteristic waveform and the active pulse cannot further improve the location effect. With the increase in grounding resistance, the two improved methods proposed in this paper significantly improve the location accuracy. However, it must be pointed out that the increase in the location error caused by the increase in the grounding resistance still exists and cannot be completely eliminated.

After analyzing Table 3, it can be seen that using the reconstructed feature waveform of the active pulse can obtain an ideal fault location effect. Since the emission process of the active pulse can be artificially controlled, the number

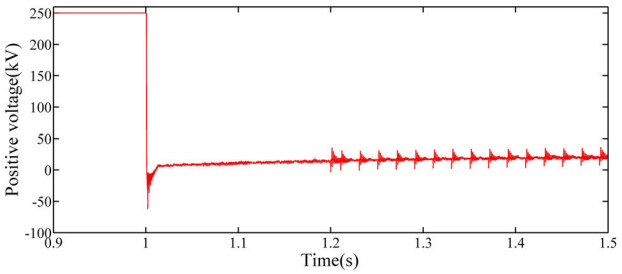

FIGURE 13. Active pulse emission time.

TABLE 4. Location results of multiple pulses.

Class	Single pulse		Three pulses		Five pulses	
	Maximum error/km	Average error/km	Maximum error/km	Average error/km	Maximum error/km	Average error/km
Class 2	0.9532	0.2093	1.0001	0.2315	1.2141	0.2163
Class 3	1.3002	0.4111	1.4345	0.3644	1.3115	0.3469
Class 4	1.2513	0.4672	1.3379	0.4045	1.6736	0.3113

of pulses can also be artificially determined. The control rectifier side sends a plurality of active pulses, and the characteristic waveforms of these active pulses are connected in series into one waveform. Running time is one of the important indicators that reflect the performance of fault location. This article explains the running time of the location method based on the method described in the literature [30]. Fig. 13 shows the overall pulse emission time. The system's fault location running time becomes longer as the number of pulses sent increases. The interval between every two pulses is 0.02 s. Therefore, the running time of the single pulse method is 0.22 s, that of three pulses is 0.26 s, and that of five pulses is 0.3 s. The results of using the new signature waveform for fault location are shown in Table 4. The method has a certain improvement effect on the location of the fault with high grounding resistance, and the improvement of the fault of the grounding resistance below 200 Ω is not obvious.

V. DESIGN OF FAULT LOCATION SYSTEM

According to the research and proposed method above, the complete workflow of the fault location system is shown in Fig. 14. One SVM classification model SVC and six SVM regression models SVR and SVRP ($i = 2, 3, 4$) are used in the system. The training samples of SVC and SVR are fault voltage waveforms, and the training samples of SVRP are active pulse waveforms.

When a pole-to-ground fault occurs, the fault voltage waveform of the fault pole is first obtained, and the fault voltage waveform is input into the support vector machine classifier SVC to determine the impedance class of grounding resistance. After the analysis in Table 1, when the fault

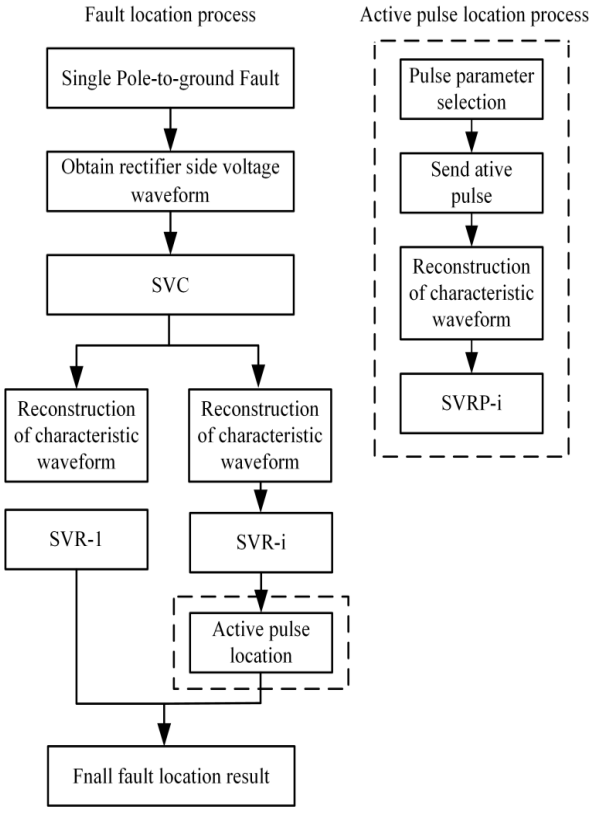

FIGURE 14. The workflow of fault location system.

belongs to Class 1, the location result of SVR1 can be directly output, and no further active pulse-assisted ranging is needed. When the fault is judged to be other categories, the fault voltage waveform is first used for the first fault location to obtain the approximate range of the fault point. Then, the preparation phase of the active pulse emission is entered. According to the description of the active pulse parameter selection in [9], the waveform shape of the pulse is affected by the number of submodules removed and the cutting time on the bridge arm, and the number of pulses emitted is determined by the number of times the submodule is removed.

Through the analysis of Table 2, when the fault belongs to Class 2, a relatively small number of pulses can be used to achieve a relatively stable location result. Considering the effect of active pulses on the system fault steady state and power quality, it is necessary to carefully determine the pulse parameters. When the SVC is judged to be Class 2, it only needs to transmit one pulse for auxiliary location. When the fault ground resistance value is higher, emitting multiple pulses is considered. Since the traveling wave will attenuate with the propagation distance on the transmission line, when the SVR determines that the position of the fault point is closer to the rectifying end, the waveform attenuation is smaller, so it is considered to cut fewer submodules. According to the output result of the SVC, it is determined whether to emit pulses and the number of pulses. The number of submodules to be cut is determined according to the location

result of the SVR. Finally, the final location result can be obtained.

In this system, different methods are used to determine the locations of high-resistance ground faults. Although the location accuracy can be significantly improved, the influence of grounding resistance cannot be completely eliminated. If the grounding resistance exceeds 1000 Ω or the transmission line continues to lengthen, the high-frequency characteristics of the fault waveform will be further attenuated, and the location accuracy will also decrease. If one wants to achieve accurate location results, one needs to increase the cost of equipment, such as increasing the sampling frequency of the recording device, using double-ended ranging and other methods.

VI. CONCLUSION

Aiming at single-pole ground faults with different transition resistances in MMC-HVDC transmission lines, this paper proposes a fault location method based on active pulse and CEEMDAN. Through the simulation experiment, the following conclusions can be drawn:

1) In the proposed method, the active pulse can better reflect the fault location information, and the fault location result is better than the location accuracy of directly using the fault pole-to-ground voltage waveform.

2) Combining the phase-to-mode conversion with the reconstructed characteristic waveform of CEEMDAN can realize adaptive extraction of the characteristics of the fault point location, thereby improving the model's ability to withstand transition resistance. Through simulation verification, the system can accurately locate single-pole ground fault points within 1000Ω under the condition of low transition resistance of 20kHz.

REFERENCES

[1] X. Shi, Z. Wang, B. Liu, Y. Liu, L. M. Tolbert, and F. Wang, "Characteristic investigation and control of a modular multilevel converter-based HVDC system under single-line-to-ground fault conditions," *IEEE Trans. Power Electron.*, vol. 30, no. 1, pp. 408–421, Jan. 2015.

[2] N. Flourentzou, V. G. Agelidis, and G. D. Demetriades, "VSC-based HVDC power transmission systems: An overview," *IEEE Trans. Power Electron.*, vol. 24, no. 3, pp. 592–602, Mar. 2009.

[3] M. Zhang, C. Yang, Z. Liu, and J. Chang, "Research on the fault location method of double-terminal MMC-HVDC transmission line without the influence of wave velocity," *J. Phys., Conf. Ser.*, vol. 1626, Oct. 2020, Art. no. 012051.

[4] S. Yang, W. Xiang, X. Lu, W. Zuo, and J. Wen, "An adaptive reclosing strategy for MMC-HVDC systems with hybrid DC circuit breakers," *IEEE Trans. Power Del.*, vol. 35, no. 3, pp. 1111–1123, Jun. 2020.

[5] X. Chen, C. Zhao, and C. Cao, "Research on the fault characteristics of HVDC based on modular multilevel converter," in *Proc. IEEE Electr. Power Energy Conf.*, Oct. 2011, pp. 91–96.

[6] X. Huang, L. Qi, and J. Pan, "A new protection scheme for MMC-based MVdc distribution systems with complete converter fault current handling capability," *IEEE Trans. Ind. Appl.*, vol. 55, no. 5, pp. 4515–4523, Sep. 2019.

[7] N. Johannesson and S. Norrga, "Estimation of travelling wave arrival time in longitudinal differential protections for multi-terminal HVDC systems," *J. Eng.*, vol. 2018, no. 15, pp. 1007–1011, Oct. 2018.

[8] Q. Huai, L. Qin, K. Liu, H. Ding, X. Liao, and T. Tan, "Combined line fault location method for MMC–HVDC transmission systems," *IEEE Access*, vol. 8, pp. 170794–170806, 2020.

[9] J. Ding, L. Li, Y. Zheng, C. Zhao, H. Chen, and X. Wang, "Distributed travelling-wave-based fault location without time synchronisation and wave velocity error," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 8, pp. 2085–2093, Jun. 2017.

[10] O. M. K. K. Nanayakkara, A. D. Rajapakse, and R. Wachal, "Location of DC line faults in conventional HVDC systems with segments of cables and overhead lines using terminal measurements," *IEEE Trans. Power Del.*, vol. 27, no. 1, pp. 279–288, Jan. 2012.

[11] Z. He, K. Liao, X. Li, S. Lin, J. Yang, and R. Mai, "Natural frequency-based line fault location in HVDC lines," in *IEEE Trans. Power Del.*, vol. 29, no. 2, pp. 851–859, Apr. 2014.

[12] L. Yuansheng, W. Gang, and L. Haifeng, "Time-domain fault-location method on HVDC transmission lines under unsynchronized two-end measurement and uncertain line parameters," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1031–1038, Jun. 2015.

[13] D. Jiandong, Z. Zhao, and L. Jing, "Combined fault location algorithm for HVDC transmission lines based on natural frequency," in *Proc. 5th Int. Conf. Electric Utility Deregulation Restructuring Power Technol. (DRPT)*, Changsha, China, Nov. 2015, pp. 1720–1725.

[14] T. Bi, S. Wang, and K. Jia, "Single pole-to-ground fault location method for MMC-HVDC system using active pulse," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 2, pp. 272–278, Jan. 2018.

[15] M. Farshad and J. Sadeh, "Accurate single-phase fault-location method for transmission lines based on K-Nearest neighbor algorithm using one-end voltage," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 2360–2367, Oct. 2012.

[16] M.-F. Guo, X.-D. Zeng, D.-Y. Chen, and N.-C. Yang, "Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1291–1300, Feb. 2018.

[17] S. Lan, M.-J. Chen, and D.-Y. Chen, "A novel HVDC double-terminal non-synchronous fault location method based on convolutional neural network," *IEEE Trans. Power Del.*, vol. 34, no. 3, pp. 848–857, Jun. 2019.

[18] H. Livani and C. Y. Evrenosoglu, "A fault classification and localization method for three-terminal circuits using machine learning," *IEEE Trans. Power Del.*, vol. 28, no. 4, pp. 2282–2290, Oct. 2013.

[19] S. Ankar, U. Sahu, and A. Yadav, "Wavelet-ANN based fault location scheme for bipolar CSC-based HVDC transmission system," in *Proc. 1st Int. Conf. Power, Control Comput. Technol. (ICPCT)*, Jan. 2020, pp. 85–90.

[20] J. M. Johnson and A. Yadav, "Complete protection scheme for fault detection, classification and location estimation in HVDC transmission lines using support vector machines," *IET Sci., Meas. Technol.*, vol. 11, no. 3, pp. 279–287, May 2017.

[21] M. Zhang and H. Wang, "Fault location for MMC–MTDC transmission lines based on least squares-support vector regression," *J. Eng.*, vol. 2019, no. 16, pp. 2125–2130, Mar. 2019.

[22] H. Livani and C. Y. Evrenosoglu, "A machine learning and wavelet-based fault location method for hybrid transmission lines," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 51–59, Jan. 2014.

[23] W. Gao, R.-J. Wai, S.-P. Qiao, and M.-F. Guo, "Mechanical faults diagnosis of high-voltage circuit breaker via hybrid features and integrated extreme learning machine," *IEEE Access*, vol. 7, pp. 60091–60103, 2019.

[24] X. Wang, J. Gao, X. Wei, Z. Zeng, Y. Wei, and M. Kheshti, "Single line to ground fault detection in a non-effectively grounded distribution network," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 3173–3186, Dec. 2018.

[25] A. M. Andrew, "An introduction to support vector machines and other kernel-based learning methods," *Kybernetes*, vol. 30, no. 1, pp. 103–115, 2000.

[26] C.-F. Juang, R.-B. Huang, and W.-Y. Cheng, "An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 4, pp. 686–699, Aug. 2010.

[27] Y. Wu, Z. Tang, Y. Xu, Y. Guo, and B. Zhang, "Support vector regression for measuring electromagnetic parameters of magnetic thin-film materials," *IEEE Trans. Magn.*, vol. 43, no. 12, pp. 4071–4075, Dec. 2007.

[28] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, 2004.

[29] X. Shi, Z. Wang, B. Liu, Y. Liu, L. M. Tolbert, and F. Wang, "Characteristic investigation and control of a modular multilevel converter-based HVDC system under single-line-to-ground fault conditions," *IEEE Trans. Power Electron.*, vol. 30, no. 1, pp. 408–421, Jan. 2015.

[30] D. Ma, X. Hu, H. Zhang, Q. Sun, and X. Xie, "A hierarchical event detection method based on spectral theory of multidimensional matrix for power system," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Aug. 9, 2019, doi: 10.1109/TSMC.2019.2931316.

JIAN-YU WU was born in Fuzhou, China, in 1984. He received the B.Sc. and M.Sc. degrees from Fuzhou University, Fuzhou, in 2010 and 2015, respectively. He works with Fuzhou Minjiang Park Office, Fuzhou. His research interests include power systems and automation, control engineering, and labor intelligence.

SHENG LAN (Member, IEEE) was born in Heilongjiang, China, in 1971. He received the B.Sc. degree from Northeast Electric Power University, Jilin, China, in 1992, and the M.S. and Ph.D. degrees from the Harbin University of Science and Technology, Harbin, China, in 2004 and 2009, respectively. He works with the College of Electrical Engineering and Automation, Fuzhou University, as a Postgraduate Tutor, where he studies high-voltage technology and power systems. His research interests include high-voltage transmission systems, insulation aging, and plasma discharge.

SI-JIE XIAO was born in Shandong, China, in 1992. He received the B.Sc. degree from the Institute of Electrical Engineering and Automation, Shandong University of Technology, China, in 2015, and the M.S. degree from the College of Electrical Engineering and Automation, Fuzhou University, China, in 2020. His research interests include high-voltage transmission systems, smart grid, artificial intelligence, and digital signal processing.

YONG-BIN YUAN was born in Heilongjiang, China, in 1974. She received the Ph.D. degree from the College of Computer Science and Technology, Harbin Engineering University, China. She is currently an Associate Professor with the College of Electrical Engineering and Automation, Fuzhou University, China. Her current research interests include data mining and privacy preservation.