IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 July 2023, accepted 14 July 2023, date of publication 18 July 2023, date of current version 25 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3296533

RESEARCH ARTICLE

Operation State Recognition of Renewable Energy Unit Based on SSAE and Improved KNN Algorithm

LINJUN SHI[1], (Member, IEEE), TAO DAI[1], WENJIE LAO[1], FENG WU[1], (Member, IEEE), KEMAN LIN[1], (Member, IEEE), AND KWANG Y. LEE[2], (Life Fellow, IEEE)

[1]College of Energy and Electrical Engineering, Hohai University, Nanjing, Jiangsu 210098, China
[2]Department of Electrical and Computer Engineering, Baylor University, Waco, TX 76798, USA

Corresponding authors: Linjun Shi (eec@hhu.edu.cn) and Tao Dai (1433630142@qq.com)

This work was supported by the Technology Project of State Grid Corporation under Grant 5100-202140478A-0-5-ZN.

ABSTRACT Quickly recognizing the real-time operating states will be helpful to identify the instantaneous and permanent power loss of the renewable energy station, so as to realize the continuous operation under the influence of the instantaneous disturbances caused by faults. This paper proposes a state recognition method for renewable energy units based on *sparse stacked auto-encoder*(SSAE) feature extraction and improved *k-nearest neighbor* (KNN) algorithm. The characteristics of this method is that the electrical parameters of the unit port are collected directly without relying on the unit's supervisory control and data acquisition (SCADA) system, whose acquisition speed is too slow to meet the recognition accuracy requirement, and that the unit operation states can be recognized quickly and accurately. Firstly, operation states of renewable energy unit are divided, and the framework for the unit's state recognition is proposed. Moreover, improved strategies for state recognition of renewable energy unit are proposed. Finally, the power system analysis software package (PSASP) is used to obtain the electrical parameters of renewable energy units and the improved KNN algorithm is used to recognize operation states after extracting features based on SSAE. By comparing the method proposed with the traditional KNN algorithm, the effect of the proposed method for states recognition is shown to be the best, with an accuracy of 98.16% and computing time of 50ms. The results show the validity of the proposed method.

INDEX TERMS Feature extraction, improved k-nearest neighbor algorithm, renewable energy, sparse stack auto-encoder, state recognition.

I. INTRODUCTION

The renewable energy unit has a fault ride through process under the fault condition of the power grid due to the unit's current carrying capacity. With the increasing penetration of renewable energy and its gradual dominance on the safety and stability characteristics of the power system, the high-power shock brought by power grid fault has an increasingly serious impact on the safety and stability control of the grid. In extreme cases, it may even cause the third line of defense of the grid to act leading to blackouts. Traditional power grid security control is mainly based on the permanent power loss of generating units, which is difficult to adapt to the instantaneous power shock caused by the renewable energy unit fault ride through in a novel power system. Quickly recognizing the real-time operating states of renewable energy units will be beneficial to identify the instantaneous and permanent power loss of the renewable energy station, so as to meet the demand of the power shock perception of renewable energy station during and after the event in the process of power grid security control, and realize the continuous control under the instantaneous power shock of the fault.

Nowadays, most of traditional operation states recognition of renewable energy unit are based on supervisory control and data acquisition (SCADA) system. Kong et al. [1] proposed a condition monitoring method for wind turbines based on SCADA data to accurately recognize the health state of wind turbines. McKinnon et al. [2] recognized the states of wind

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

turbine pitch system based on SCADA data, and monitors the faults of turbine hydraulic system. Furthermore, Dao [3] used the structural fracture detection in SCADA data to recognize the states of wind turbine transmission system. However, the aforementioned studies did not address state recognition for whether the unit is in a fault ride through or off-grid state. A wind turbine gearbox monitoring model based on SCADA data is proposed by Wu et al. [4], which analyzes the unseen SCADA vibration data to detect potential faults. Santolamazza et al. [5] proposed a method based on machine learning technology and used SCADA data to predict abnormal operation states of gearbox and generator. While these two studies solely concentrate on recognizing fault states within specific generator components, the process from data acquisition to states recognition is time-consuming. In addition, Liu et al. [6] used the gearbox oil temperature and generator temperature in SCADA data to warn the fault of wind turbine. However, the speed of SCADA system to collect data is in the second level, which cannot meet the speed requirements of millisecond emergency continuous control of power grid.

Machine learning methods have been employed in the area of fault recognition. For instance, Yu et al. [7] used the k-nearest neighbor (KNN) algorithm to recognize rubbing faults and identify their locations. However, this study pursued the accuracy of recognition only, but neglected the speed of recognition. Duan et al. [8] used sparse stack auto-encoder (SSAE) to represent the target dynamics and then to recognize abnormal dynamics of space targets. Wang et al. [9] proposed a method based on SSAE to diagnose faults of rolling bearings. Zhang et al. [10] built a simulation system based on power system analysis software package (PSASP) and studied the transient power characteristics of photovoltaic generation system with it. However, the proposed methods of above studies have not been applied in the field of state recognition.

In view of the above problems, this paper proposes a state recognition method for renewable energy unit based on SSAE feature extraction and improved KNN algorithm. The characteristic of this method is that it does not need to rely on the SCADA data of the unit itself, and directly collects the electrical parameters of the unit to quickly and accurately judge the operation states of the unit. Firstly, operation states of renewable energy unit are divided, and the process framework of renewable energy unit state recognition is proposed. Secondly, to overcome the problems of low recognition accuracy and speed of traditional KNN algorithm, improved strategies suitable for state recognition of renewable energy unit are proposed. Finally, PSASP is used to obtain the electrical parameters of the renewable energy unit and the improved KNN algorithm is used to recognize operating states after extracting features based on SSAE. Compared with the traditional KNN algorithm, the improved KNN algorithm proposed in this paper has obvious improvements on both speed and accuracy of state recognition.

II. OPERATION STATES RECOGNITION FRAMEWORK OF RENEWABLE ENERGY UNIT

The division of states of the renewable energy unit can refer to the latest wind turbine fault ride through standard in China [11], which stipulates that the unit should have low/high voltage ride through capability, that is, when the grid fault or disturbance causes the voltage to drop/rise, the wind turbine can maintain the ability to operate within the specified voltage change and time. The voltage ride through curve of the wind turbine is shown in Fig. 1, which is modified version of the Chinese national standard. [11]. Considering that operating states of the renewable energy unit should fully reflect the power loss of the unit to meet the stability control requirements, operating states of renewable energy units are divided into: 'normal state', 'low voltage ride through (LVRT) state', 'LVRT recovery state', 'high voltage ride through (HVRT) state' and 'off-grid state'.

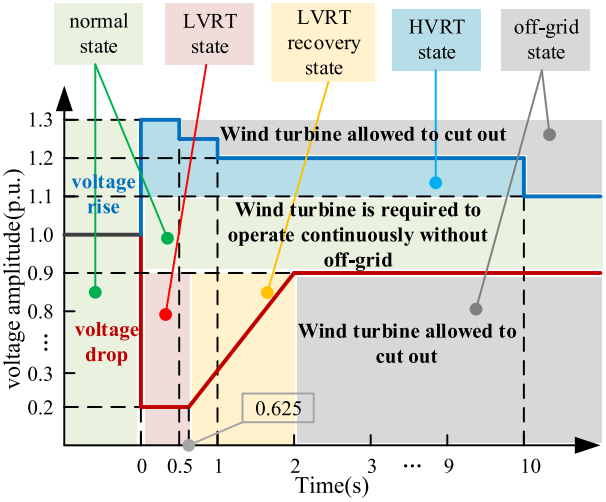

FIGURE 1. Voltage ride through curve of the wind turbine.

The feature extraction and state recognition of renewable energy units can be divided into the following steps, as shown in Fig. 2. The 4 steps in Fig. 2 are explained as follows:

Step 1: Collect electrical parameters of renewable energy unit. The electrical parameters are collected at a millisecond speed through a source control terminal device mounted at the port of renewable energy unit.

Step 2: Data pre-processing and sample set generation. The electrical parameters collected are stored, and the local outlier factor (LOF) algorithm is applied to exclude any bad data in the electrical parameters collected, followed by normalization of the remaining valid data as sample set.

Step 3: Extract features of the renewable energy unit operation states based on SSAE. The sample set generated from step 2 is input into SSAE, and then the activation function and the number of SSAE layers are preset. The extracted features from the sample set are output and the network parameters are recorded after SSAE training.

Step 4: Operation states recognition of renewable energy unit. The extracted features from step 3 are input into the improved KNN classifier for training. The five operating

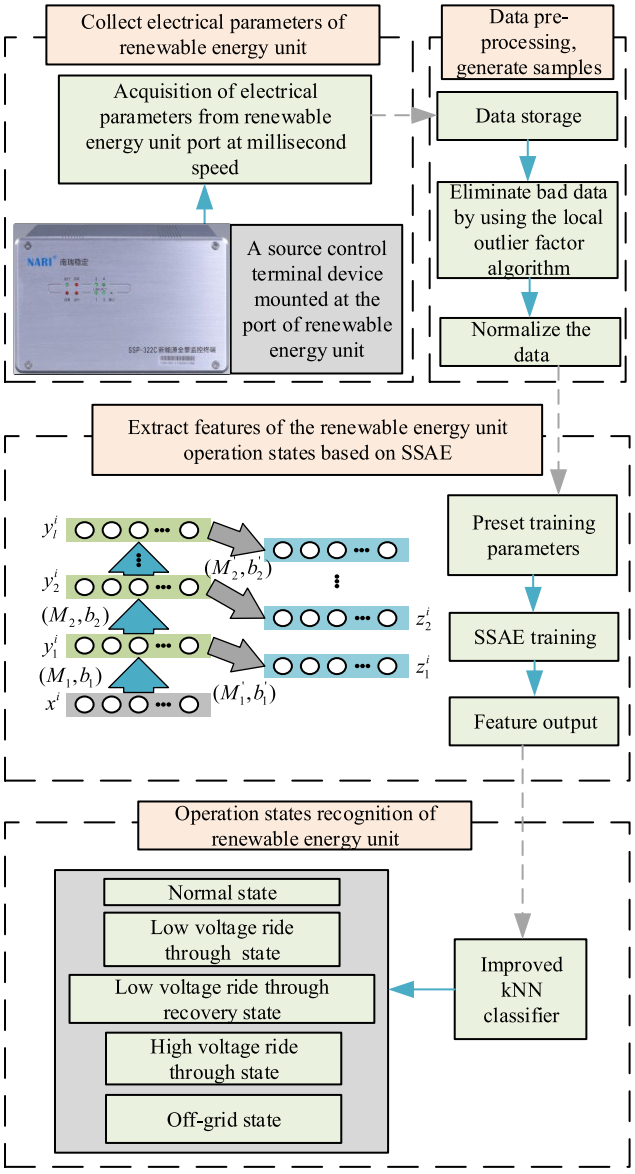

FIGURE 2. Flowchart of renewable energy unit states recognition.

states of the renewable energy unit are recognized after running the improved KNN classifier, and then calculate the recognition speed and accuracy.

III. STACK AUTO-ENCODER

In recent years, deep learning has developed rapidly in various industries and achieved remarkable results, which also provides a new direction for the state recognition of renewable energy units. Different from the traditional methods with complex modeling and threshold judgment, deep learning extracts more useful features from a large number of training data by building a hidden layer neural network, and finally completes the classification task with high speed and high accuracy. Moreover, after collecting the electrical parameters corresponding to various operating states of the renewable energy unit, extracting the effective features that characterize operating states of the unit is the key to recognize.

Stack auto-encoder (SAE) is a deep learning algorithm with flexible structure and high efficiency. It has the powerful ability to extract data abstract features through layer-by-layer unsupervised greedy training, and avoids relying on manual experience to select features. With the prosperity of the era of big data, SAE is widely used in image recognition classification, data analysis visualization, audio classification and other fields of life.

A. AUTO-ENCODER

Auto-encoder (AE) is the basic unit of SAE, which is composed of three layers of neural network: input layer, hidden layer, and output layer [12]. The structure of AE is shown in Fig. 3. The AE has two functions: encoding and decoding [13]. It converts the input information into low-dimensional features and outputs them to the decoder through the encoder, and then reconstructs the features into the original input information through the decoder [14]. When the reconstruction error between the input and output data is small enough, the output of the hidden layer is the best dimensionality reduction data structure, thus successfully extracting the features of the input data.

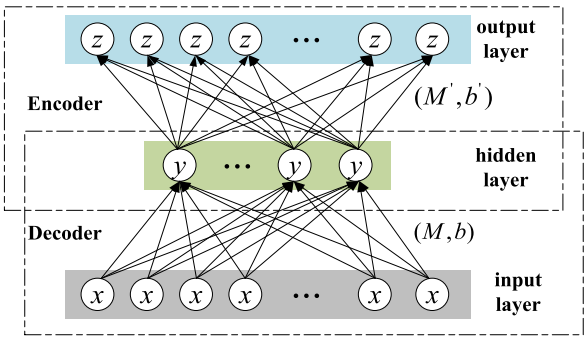

FIGURE 3. Structure of AE.

For a training sample set $X = \{x^1, x^2, \ldots, x^N\}$ with a total number of N samples, x^i is the i^{th} sample in the training sample set, and the dimension of each sample is m. Let the output vector of the hidden layer after the encoder is $Y = \{y^1, y^2, \ldots, y^N\}$, y^i is the i^{th} vector of the hidden layer output, and the dimension of each vector is n, then there is an encoder model:

$$Y = f_{(M,b)}(X) = s_f(MX + b) \tag{1}$$

where M represents the weight matrix between the input layer and the hidden layer; b represents the coding bias vector; s_f represents the encoding activation function.

Sigmoid function is a mathematical function with a beautiful S-shaped curve, which has a wide range of applications in logistic regression and artificial neural networks. The sigmoid function maps the domain of definition between 0 and 1. It is continuous and smooth and monotonically differentiable in the domain of definition, and these properties make the output of the sigmoid function interpreted as the probability of the classification result, which is suitable for classification

problems. In this paper, sigmoid is used as the activation function, which is sensitive to small changes in the middle and relatively insensitive in both ends, and is suitable for the AE. The sigmoid function is defined as $sigmoid(a) = 1/(1+e^{-a})$, where a is the input vector of each neuron.

The decoder uses the output vector of the hidden layer to reconstruct the original data. Let Z be the output vector set after decoding, $Z = \{z^1, z^2, \ldots, z^N\}$, where z^i is the output vector corresponding to the i^{th} sample, and the dimension of each vector is also m. The decoder model is:

$$Z = g_{(M',b')}(Y) = s_g(M'Y + b') \tag{2}$$

where M' represents the weight matrix between the output layer and the hidden layer; b' represents the decoding bias vector; s_g represents the decoding activation function.

The goal of AE is to extract the features of the original data by minimizing the reconstruction error $L(x^i, z^i)$ between the input and output vectors. The reconstruction error can be expressed as:

$$L(x^i, z^i) = \frac{1}{N}\sum_{i=1}^{N}\left\|x^i - z^i\right\|^2 \tag{3}$$

The network weights and biases can be continuously updated by minimizing the reconstruction error through the gradient descent algorithm. The update criteria can be defined as:

$$\begin{cases} M^{(k+1)} = M^{(k)} - \varepsilon\dfrac{\partial L(x,z)}{\partial M} \\ b^{(k+1)} = b^{(k)} - \varepsilon\dfrac{\partial L(x,z)}{\partial b} \end{cases} \tag{4}$$

where ε is the initial learning rate.

B. SPARSE STACK AUTO-ENCODER

The SAE is composed of multiple AE stacks, and the input of the upper AE is provided by the lower AE [15]. By training each AE model layer by layer, the entire network can be trained to extract more compact and effective abstract features.

The SAE model structure is shown in Fig. 4.

In the first layer of the model, the weight and bias of AE are randomly initialized, and the input from the training sample set is mapped to the hidden layer output vector using the encoder. Next, the decoder reconstructs the hidden layer output vector while minimizing the reconstruction error (3) by continuously updating the weight and bias. This process completes the training of the first AE. Furthermore, the output vector of the hidden layer is used as the input of the next level AE, and the next level AE is trained according to the above method, and so on, until all AE training is completed. At this time, the output vector of the last AE hidden layer is the final feature vector.

When the number of hidden layer nodes is greater than the number of input layer nodes, the AE will not be able to successfully extract the features of the input data, and the sparse auto-encoder can solve this problem to ensure that the encoder achieves feature dimension reduction extraction.

The sparse auto-encoder adds sparse regularization constraints on the basis of each AE to control the number of hidden layer node activation [16], that is, adds sparse regularization penalty terms to the reconstruction error of AE. For the sigmoid function, when the node output approaches 1, it means that the node is 'activated', while when the node output approaches 0, it means that the node is 'suppressed'. The sparse auto-encoder can 'suppress' most of the hidden layer nodes to avoid extracting redundant features.

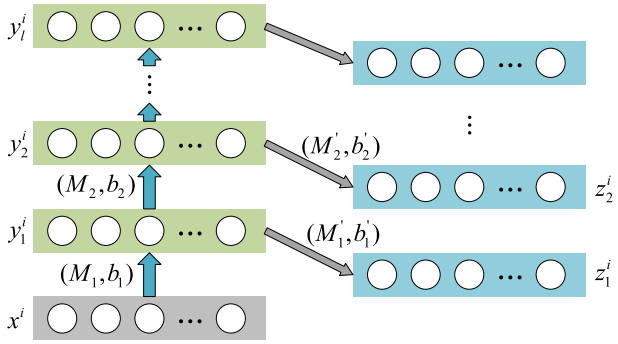

FIGURE 4. Model structure of SAE.

The sparse auto-encoder uses a regularization term to sparsely represent the hidden layer, and its loss function L can be expressed as:

$$L = L(x^i, z^i) + \lambda\omega_{L2} + \beta\omega_s \tag{5}$$

where ω_{L2} is the $L2$ regular term; ω_s is a sparse regular term; λ and β are the corresponding coefficients.

The sparse regularization constraint is added to each AE in the SAE to form SSAE, which can ideally extract the dimensionality reduction features of the input data and avoid feature redundancy.

IV. IMPROVED KNN ALGORITHM RECOGNIZER

A. TRADITIONAL KNN ALGORITHM RECOGNITION

The KNN algorithm was proposed by Cover and Hart in 1967. It is a simple and effective classification method and is widely used in pattern recognition, text classification and other fields [17].

It is assumed that the operating state of the new energy unit at a certain moment can be expressed by an n-dimensional vector:

$$\mathbf{X} = [x_1, x_2, x_3, \cdots, x_n] \tag{6}$$

where $\mathbf{X}$ is the operating state of the unit; x_i is the i^{th} feature and there are n features.

The key of KNN algorithm is to determine the proximity between the classification objects. Nowadays, there are many methods to measure the proximity, such as Euclidean distance, Hamming distance and so on.

The Euclidean distance $D(\mathbf{X}, \mathbf{Y})$ between two points $\mathbf{X}$ and $\mathbf{Y}$ in the sample space is defined as [18]:

$$D(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{7}$$

First, a sample set is formed with data whose states are known. Once a new data with unknown state is entered, the Euclidean distance between the new data vector and each vector in the sample set is calculated. Subsequently, the first k vectors in the sample data set that are most similar to the new data are selected, and usually k is an odd number less than 20. Finally, algorithm selects the state with the most occurrences in these k vectors as the classification of new data. The classification principle of KNN algorithm is shown in Fig. 5, where the query is assigned to Class 1 in the case of $k = 5$ according to the KNN algorithm.

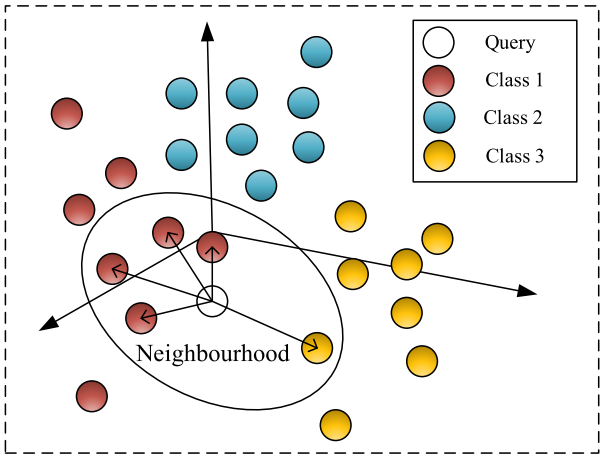

FIGURE 5. Classification diagram of KNN algorithm.

B. IMPROVED STRATEGIES OF KNN ALGORITHM RECOGNITION

The structure of the traditional KNN algorithm is relatively simple. When the input feature data dimension is too high and the sample set data is large, the calculation amount will increase significantly, which will slow down the output of the recognition result. In addition, when the sample training set changes, the k value of the traditional KNN remains fixed, which is unable to adapt to the sample and output the most accurate result. When the sample data distribution is uneven, the recognition accuracy of traditional KNN algorithm will decrease. Considering that the dimension of the electrical parameters from the renewable energy unit port is high, and the distribution of the electrical parameters in each state is uneven, it is difficult to obtain great effect of state recognition by using the traditional KNN algorithm. Therefore, this paper improves the traditional KNN algorithm by adding improved strategies and inputting the features extracted by SSAE into the KNN classifier, so that the algorithm is able to get high accuracy and speed for state recognition of renewable energy units.

1) BAD DATA PROCESSING

When collecting electrical parameters from the generator port, bad data affecting the accuracy of recognition are usually collected due to sensor failure, human record deviation and other reasons. In view of this situation, this paper introduces the anomaly detection LOF algorithm to process the input electrical parameter samples, eliminate the bad data in the sample set, and then improve the accuracy of the algorithm recognition.

In 2000, Breunig et al. proposed the idea of using LOF to detect local outliers in samples, which determines outliers by comparing the ratio of the local density of a point in the sample to the local density of its adjacent points [19].

In the sample set, the distance between the objects m and n is defined as $d(m, n)$. The k-distance of an object n is represented by $d_k(n)$ when there are at least k objects x such that $d(x, n) \leq d(m, n)$, and at most k-1 objects x with $d(x, n) < d(m, n)$. The set of objects with a distance less than or equal to $d_k(n)$ from the object n is called the k^{th} distance field of n, denoted by $N_k(n)$. Obviously, the value of $d_k(n)$ is smaller in the area with high object density, while the value of $d_k(n)$ is larger in the area with low object density. The reachable distance $d_{reach}(p, q)$ of object p relative to object q in the sample set is defined as [20]:

$$d_{reach}(p, q) = \begin{cases} d_k(q)(d(p, q) \leq d_k(q)) \\ d(p, q)(d(p, q) > d_k(q)) \end{cases} \tag{8}$$

In Fig. 6, $k = 6$, the reachable distance between object p_1 and object q is $d(p_1, q)$, and the reachable distance between object p_2 and object q is $d_k(q)$.

The local reachable density $lrd_k(p)$ of object p is defined as the derivative of the average of all reachable distances in its k_{th} distance neighborhood, namely:

$$lrd_k(p) = |N_k(p)| / \sum_{q \in N_k(p)} d_{reach}(q, p) \tag{9}$$

where $N_k(p)$ is the k^{th} distance neighborhood of object p. The neighbor data of the object p becomes more concentrated as $lrd_k(p)$ increases, which increases the possibility that the object p and its neighbors are similar data.

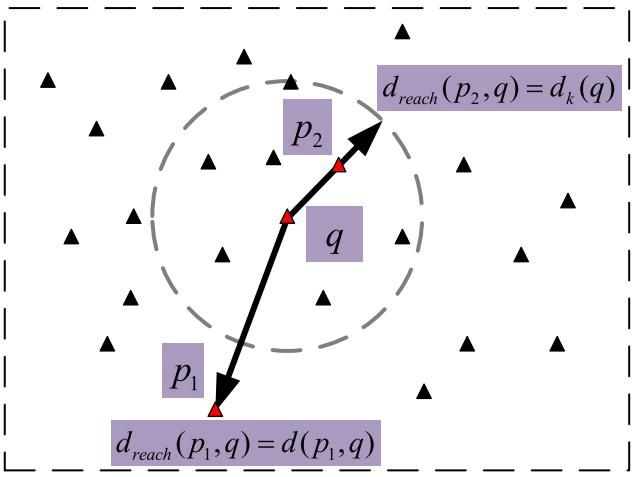

FIGURE 6. Schematic diagram of reachable distance.

Define the $LOF_k(p)$ to characterize the possibility that the object p is an outlier [21]:

$$LOF_k(p) = (\sum_{q \in N_k(p)} lrd_k(q)/lrd_k(p))/ |N_k(p)| \tag{10}$$

If $LOF_k(p)$ is close to 1, it means that the density of p is close to that of the points in its k-distance neighborhood, so p may belong to the same class as the point in its k-distance neighborhood. The smaller $LOF_k(p)$ is less than 1, indicating that the density of p is higher than that of other points in its neighborhood, and p is more likely to belong to the same class as the points in its neighborhood. The more $LOF_k(p)$ is greater than 1, indicating that the density of p is less than the density of other points in its neighborhood, and the more likely p is an outlier.

The LOF algorithm can effectively eliminate the bad data in the sample set, and is equally accurate in the case of uneven distribution of electrical parameters in different states of renewable energy unit, which can effectively improve the accuracy of state recognition.

2) SAMPLE SET COMPRESSION

Since the KNN algorithm needs to calculate and sort the distance between the sample points to be tested and each point in the training set, and the feature of the SSAE input KNN classifier is a large sample set, which will greatly increase the computational complexity of the KNN algorithm, resulting in slower recognition speed. In order to solve this problem, this paper proposes a method of compressing sample set to reduce the number of sample point distance calculations of KNN algorithm.

Because operating states of renewable energy unit have been divided into five categories, and each category has multiple samples, this paper compresses the samples of SSAE input KNN classifier based on the idea of clustering using representatives (CURE) [22]. The specific steps are as follows:

Step 1: The input samples are divided into five sample sets according to five states, named $S_n(n = 1, \ldots, 5)$.

Step 2: Remove duplicate data in sample S_n.

Step 3: Calculate the Euclidean distance between each two points in the sample set S_n, and use the average value to merge two points whose Euclidean distance is relatively small. Repeat the above process until the Euclidean distance from the nearest two points in the sample set is greater than the set value ε, and then form a new sample set S_n'.

After the sample sets are compressed, the number of sample sets input to the KNN classifier is lessen, which reduces the number of calculations of the Euclidean distance, ensures the accuracy of the recognition results, and improves the recognition efficiency of the algorithm.

3) K VALUE ADAPTIVE

The selection of k value in KNN algorithm will directly affect the recognition results. If the k value is too small, the recognition results may be affected by some outliers. If the k value is too large, the recognition result may be affected by too many samples and uneven distribution [23]. The traditional KNN algorithm is difficult to guarantee the recognition accuracy when the training samples change because k takes a fixed value. Aiming at the above problems, this paper adopts the k value adaptive strategy, selects the best k value according to the training samples.

The optimal k value is determined by the test recognition accuracy of the training samples, and the recognition accuracy P is defined as:

$$P = 1 - Y/Z \tag{11}$$

where Y is the number of output error recognition results in the test, and Z is the total number of tests.

When the data of the sample training set changes, the value of k will also be changed, and then the training samples will be used to calculate the recognition accuracy under each k value. The k value with the highest accuracy is returned as the best k value under the training sample.

V. EXAMPLE ANALYSIS

A. ELECTRICAL PARAMETERS ACQUISITION OF RENEWABLE ENERGY UNIT

During the actual operation of renewable energy units, most of them are in 'normal state', and the electrical output parameters are stable within a certain range. It is not easy to obtain the electrical output parameters of actual renewable energy unit under other operating states such as 'LVRT state' or 'HVRT state'. Therefore, this paper uses PSASP to simulate various operating states of renewable energy unit and obtain the electrical parameters. The electrical parameters waveform of renewable energy unit under different operating states is shown in Figs. 7-9, where U, I, P, Q, I_{p}, I_{q}, and f represent voltage, current, active power, reactive power, active current, reactive current, and frequency, respectively. Among them, the waveform before 0.2 s is the electrical parameters in 'normal state', and different faults are set in PSASP at 0.2 s, so that the unit can enter the LVRT mode, HVRT mode, or off-grid state. Figure 7 shows the variation of unit electrical parameters under 'LVRT state' and 'LVRT recovery state'. Figures 8 and 9 show the variation of unit electrical parameters in 'off-grid state' and 'HVRT state', respectively. In Fig. 7, U, P and I_{p} fall at 0.2 s due to a short circuit fault, then start to recover after the fault ends; I increases steeply at the moment of failure, then decays and finally recovers after the fault ends; I_{q} increases during the fault to provide reactive support to the system; Q increases steeply at the moment of the fault, then decays, and increases in reverse after the fault ends and then returns to normal; f increases slightly during the fault. Figures 8 and 9 are in different states due to different faults and the unit electrical parameters change differently.

Considering the influence of different types of renewable energy unit and different types of faults on electrical parameters, this paper sets doubly-fed wind turbine, direct-drive wind turbine, and photovoltaic power generation as the renewable energy unit, respectively, and then sets single-line-to-ground short-circuit, line-to-line short-circuit, three-phase-to-ground short-circuit faults and load shedding disturbance faults. The renewable energy unit will enter different operating states in the face of different faults. The sampling

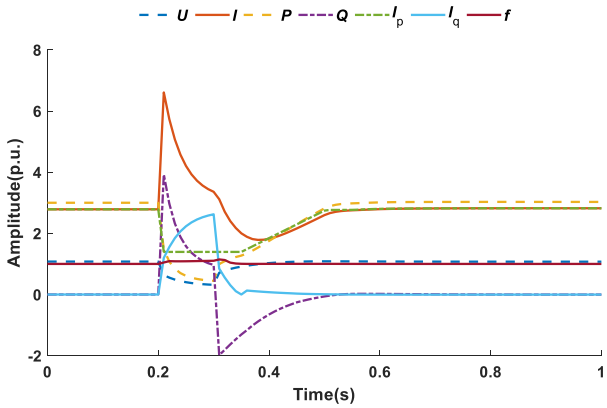

FIGURE 7. Electrical parameters waveform in 'LVRT state' and 'LVRT recovery state'.

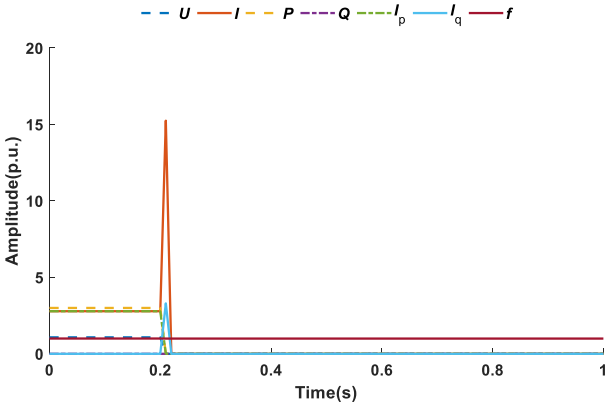

FIGURE 8. Electrical parameters waveform in 'off-grid state'.

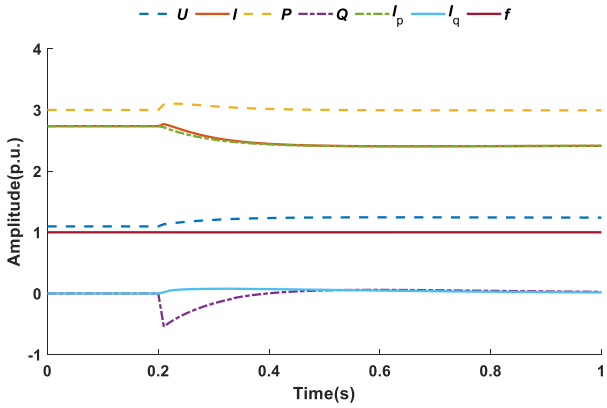

FIGURE 9. Electrical parameters waveform in 'HVRT state'.

interval is set to 1ms, and a total of 6,000 electrical parameter samples under each operating state are collected through multiple simulations.

B. ELIMINATE ERROR DATA

Considering that error data will be collected due to sensor failure and other reasons in practical applications, this paper uses a random number generation function to generate error

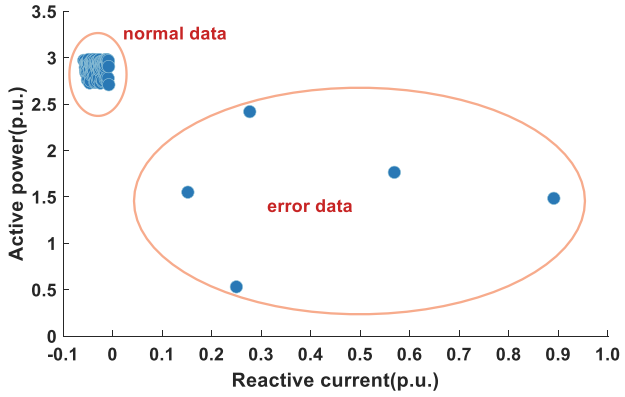

FIGURE 10. Distribution of error data.

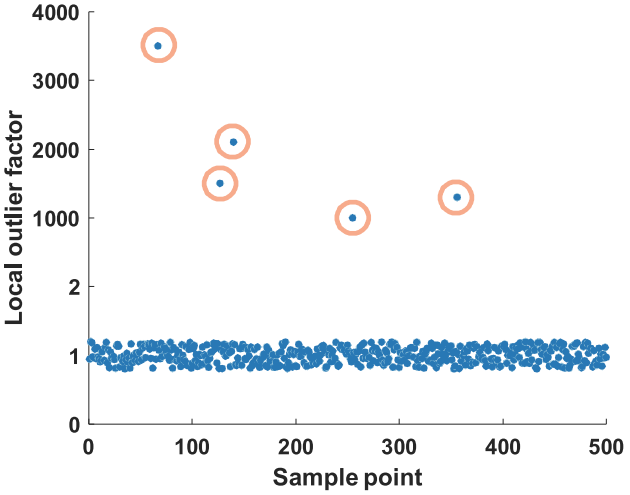

FIGURE 11. Distribution of LOF.

data and insert it into the 500 'normal state' sample sets irregularly. The error data distribution is shown in Fig. 10. On this basis, the bad data processing strategy is used to calculate the LOF of the sample points which is shown in Fig. 11. The LOF of five sample points in Fig. 11 is much larger than 1, indicating that it is the error data that should be eliminated, which is consistent with the generated error data.

C. FEATURE EXTRACTION OF OPERATION STATES BASED ON SSAE

In this paper, the neuron activation function of SSAE is set to 'sigmoid' function, and the training round of each AE is set to 200 times. Since the number of input layer nodes of SSAE is the same as the dimension of electrical parameters samples, and the number of input/output layer nodes is set accordingly, so only the number of hidden layer nodes and layers of SSAE need to be determined. The types of electrical parameters collected for states recognition of renewable energy unit are shown in Table 1. A total of 6,000 samples were collected to form a 6000 $\times$ 7-dimensional matrix as the input of SSAE.

Since the input samples to SSAE is 7-dimensional data, the number of hidden layer nodes is considered to be within 100, and the number of hidden layers is considered to be an

integer of 1 to 5. After obtaining the structural parameters of the first layer AE, the number of nodes in the hidden layer of the next layer AE is superimposed until the performance is optimal. Finally, the optimal structural parameters of SSAE are obtained.

TABLE 1. The type of 7 electrical parameters collected.

Label	The type of electrical parameters	Label	The type of electrical parameters
U	Voltage	P	Active power
I	Current	Q	Reactive power
I_q	Reactive current	I_p	Active current
f	Frequency	-	-

Feature extraction tasks often use principal component analysis (PCA) to complete [24], but the PCA is based on linear transformation to achieve feature extraction, and SSAE extracts abstract data features by nonlinear mapping of input data through multiple hidden layers, which improves the performance of nonlinear classification tasks. This paper compares the ability of PCA and SSAE to extract the features of electrical parameters. First, PCA is used to extract the features of electrical parameters, and the two principal components with the highest contribution rate are projected into the rectangular plane, as shown in Fig. 12. Then, a hidden layer with two nodes is inserted after the last hidden layer of the SSAE model. After the electrical parameters are inputted into the SSAE model, the two-dimensional abstract features extracted from the hidden layer are projected in the rectangular plane, as shown in Fig. 13.

It can be seen from Fig. 12 that PCA has a weak ability to distinguish the states of renewable energy unit, and the samples of each type of the states overlap in a large area on the two-dimensional plane. As can be seen in Fig. 13, the two-dimensional features extracted from the electrical parameters by SSAE are distinguished clearly for various states of renewable energy unit. Samples in different states form a cluster, and there is almost no overlap between clusters. In summary, SSAE has a strong ability to extract abstract features through nonlinear mapping, and its feature extraction performance is better than the PCA based on linear mapping.

The SSAE is used to extract 2-, 3- and 4-dimensional features, and then inputted them into the KNN classifier. The recognition accuracy and time consumption are calculated. The method described in Section IV-B.2 is used to compress the features of the output from SSAE into a new sample training set. After compression, the number of samples is reduced from 6,000 to 4,800, which decreases the amount of calculation and improves the speed of the algorithm.

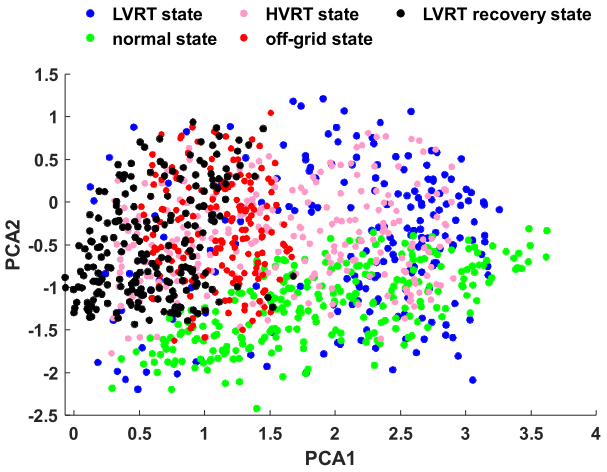

FIGURE 12. Extract features used PCA.

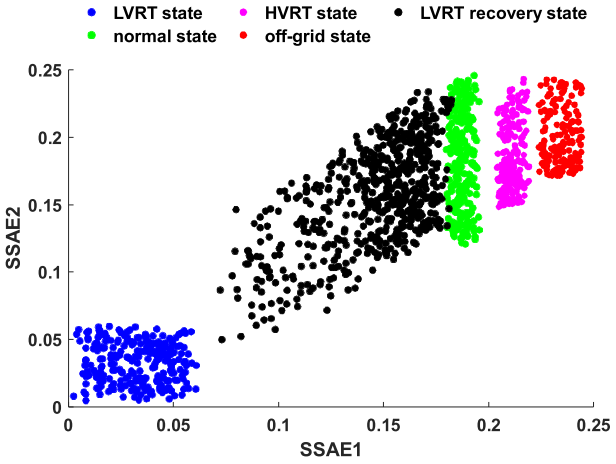

FIGURE 13. Extract features used SSAE.

D. RECOGNITION RESULTS AND ANALYSIS

To verify the speed and accuracy of the SSAE-based and improved KNN algorithm proposed in this paper, the recognition accuracy and time consumption of the traditional KNN algorithm and the proposed method are compared, as shown in Table 2. In the Table, S1 represents the traditional KNN algorithm, S2 represents the improved KNN algorithm without using SSAE to extract features, and S3 represents the improved KNN algorithm using SSAE to extract features.

It can be seen from Table 2 that the traditional KNN algorithm takes the longest time to recognize states of renewable energy unit and has the lowest accuracy. The improved methods of 'bad data processing', 'sample set compression' and 'k value adaptive' increase the recognition accuracy of KNN algorithm by about 10% and reduce the recognition time consuming by 180 ms. Using SSAE to extract features further improves the recognition accuracy of KNN algorithm. Moreover, the 2-dimensional SSAE feature recognition takes the shortest time, but the recognition accuracy is not as good as the 3- and 4-dimensional features. The accuracy of

TABLE 2. Recognition results.

Method	Dimension of sample training set	Sample size	Adaptive k value	Recognition accuracy	Time (s)
S1	7	6000	-	78.52%	1.530
S2	7	4800	9	88.25%	1.350
S3	2	4800	5	94.35%	0.020
S3	3	4800	5	98.16%	0.050
S3	4	4800	7	98.23%	0.100

the 4-dimensional SSAE feature recognition is only 0.07% higher than that of the 3-dimensional, but its time consumption is twice of that of the 3-dimensional feature. In summary, the effect of extracting the 3-dimensional SSAE features for states recognition is the best, with an accuracy of 98.16% and time consuming of 50 ms, which meets the requirements of speed and accuracy. The results provide a control basis for the stability control devices of the electrical system which is able to judge the power plant operation and the occurrence of faults based on the unit operation states recognized by using the improved KNN and SSAE algorithms, so as to make corresponding actions to ensure the safe and stable operation of the power system.

VI. CONCLUSION

In this paper, operation states of renewable energy unit are divided into 'normal state', 'LVRT state', 'LVRT recovery state', 'HVRT state' and 'off-grid state'. A method based on SSAE feature extraction and improved KNN algorithm is proposed to recognize operation states of renewable energy unit. The recognition accuracy and time consumption of traditional KNN algorithm, improved KNN algorithm with and without SSAE feature extraction proposed in this paper are compared. The results show that the traditional KNN algorithm takes the longest time to recognize. The improved methods of 'bad data processing', 'sample set compression' and 'k value adaptive' increase the recognition accuracy of KNN algorithm and reduce the recognition time consumption. Using SSAE to extract features further improves the recognition accuracy of KNN algorithm. In addition, the 2-dimensional SSAE feature recognition takes the shortest time, but the recognition accuracy is not as good as the 3- and 4-dimensional features. The accuracy of the 4-dimensional SSAE feature recognition is slightly higher than that of 3-dimensional, but its time consumption is twice that of 3-dimensional SSAE feature.

In summary, the recognition accuracy and speed of the proposed method are significantly improved compared with the traditional algorithm. Using the SSAE to extract the 3-dimensional features of electrical parameters and the improved KNN algorithm proposed in this paper to recognize operating states of the renewable energy unit has achieved the best results. The recognition accuracy and the recognition time consumption can meet the states perception requirements of renewable energy unit in the stability control process and reduce the shock of faults from the grid.

REFERENCES

[1] Z. Kong, B. Tang, L. Deng, W. Liu, and Y. Han, "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units," *Renew. Energy*, vol. 146, pp. 760–768, Feb. 2020.

[2] C. McKinnon, J. Carroll, A. McDonald, S. Koukoura, and C. Plumley, "Investigation of isolation forest for wind turbine pitch system condition monitoring using SCADA data," *Energies*, vol. 14, no. 20, p. 6601, Oct. 2021.

[3] P. B. Dao, "Condition monitoring and fault diagnosis of wind turbines based on structural break detection in SCADA data," *Renew. Energy*, vol. 185, pp. 641–654, Feb. 2022.

[4] X. Wu, H. Wang, G. Jiang, P. Xie, and X. Li, "Monitoring wind turbine gearbox with echo state network modeling and dynamic threshold using SCADA vibration data," *Energies*, vol. 12, no. 6, p. 982, Mar. 2019.

[5] A. Santolamazza, D. Dadi, and V. Introna, "A data-mining approach for wind turbine fault detection based on SCADA data analysis using artificial neural networks," *Energies*, vol. 14, no. 7, p. 1845, Mar. 2021.

[6] Y. Liu, Z. Wu, and X. Wang, "Research on fault diagnosis of wind turbine based on SCADA data," *IEEE Access*, vol. 8, pp. 185557–185569, 2020.

[7] M. Yu, W. Chen, and Y. Lu, "Fault diagnosis and location identification of rotor–stator rub-impact based on Hjorth parameters," *Eng. Failure Anal.*, vol. 138, Aug. 2022, Art. no. 106299.

[8] J. Duan, Y. Ma, L. Zhang, and P. Xie, "Abnormal dynamic recognition of space targets from ISAR image sequences with SSAE-LSTM network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5102916.

[9] J. Wang, C. Zhan, D. Yu, Q. Zhao, and Z. Xie, "Rolling bearing fault diagnosis method based on SSAE and softmax classifier with improved K-fold cross-validation," *Meas. Sci. Technol.*, vol. 33, no. 10, Oct. 2022, Art. no. 105110.

[10] Y. Zhang, N. Ji, W. Qi, Z. Gao, J. Hao, and P. Luo, "Research on low voltage ride through characteristics of distributed photovoltaic system," *Energy Rep.*, vol. 9, pp. 887–893, Mar. 2023.

[11] *Wind Turbines—Test Procedure of Voltage Fault Ride Through Capability*, Chinese Standard GB/T 36995-2018, 2018.

[12] A. Li, F. Zhang, S. Li, T. Chen, P. Su, and H. Wang, "Efficiently generating sentence-level textual adversarial examples with Seq2seq stacked auto-encoder," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119170.

[13] S. P. Mishra, V. Krishna Rayi, P. K. Dash, and R. Bisoi, "Multi-objective auto-encoder deep learning-based stack switching scheme for improved battery life using error prediction of wind-battery storage microgrid," *Int. J. Energy Res.*, vol. 45, no. 14, pp. 20331–20355, Nov. 2021.

[14] Y.-D. Lin, Z.-Q. Liu, R.-H. Hwang, V.-L. Nguyen, P.-C. Lin, and Y.-C. Lai, "Machine learning with variational autoencoder for imbalanced datasets in intrusion detection," *IEEE Access*, vol. 10, pp. 15247–15260, 2022.

[15] M. Gnouma, R. Ejbali, and M. Zaied, "A two-stream abnormal detection using a cascade of extreme learning machines and stacked auto encoder," *Multimedia Tools Appl.*, vol. 2023, pp. 1–28, Mar. 2023.

[16] Y. Zhu, X. Wu, J. Qiang, X. Hu, Y. Zhang, and P. Li, "Representation learning with deep sparse auto-encoder for multi-task learning," *Pattern Recognit.*, vol. 129, Sep. 2022, Art. no. 108742.

[17] J. Kim, S. Jang, E. Park, and S. Choi, "Text classification using capsules," *Neurocomputing*, vol. 376, pp. 214–221, Feb. 2020.

[18] R. Yang and J. Kan, "Euclidean distance-based adaptive collaborative representation with Tikhonov regularization for hyperspectral image classification," *Multimedia Tools Appl.*, vol. 82, no. 4, pp. 5823–5838, Feb. 2023.

[19] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," *ACM SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, Jun. 2000.

[20] V. P. Kumaravel, M. Buiatti, E. Parise, and E. Farella, "Adaptable and robust EEG bad channel detection using local outlier factor (LOF)," *Sensors*, vol. 22, no. 19, p. 7314, Sep. 2022.

[21] J. Ma, Z. Teng, Q. Tang, W. Qiu, Y. Yang, and J. Duan, "Measurement error prediction of power metering equipment using improved local outlier factor and kernel support vector regression," *IEEE Trans. Ind. Electron.*, vol. 69, no. 9, pp. 9575–9585, Sep. 2022.

[22] S. Guha, R. Rastogi, and K. Shim, "Cure: An efficient clustering algorithm for large databases," *Inf. Syst.*, vol. 26, no. 1, pp. 35–58, Mar. 2001.

[23] Z. Pan, Y. Wang, and Y. Pan, "A new locally adaptive k-nearest neighbor algorithm based on discrimination class," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106185.
[24] L. Zhu, S. Zhang, Q. Ma, H. Zhao, S. Chen, and D. Wei, "Classification of UAV-to-ground targets based on enhanced micro-Doppler features extracted via PCA and compressed sensing," *IEEE Sensors J.*, vol. 20, no. 23, pp. 14360–14368, Dec. 2020.

LINJUN SHI (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Hohai University, Nanjing, China, in 1999 and 2003, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 2010.

He was a Visiting Scholar with the Department of Electrical and Computer Engineering, Baylor University, Waco, TX, USA, in 2014. He is currently an Associate Professor with the College of Energy and Electrical Engineering, Hohai University. His research interests include power system analysis and control, new energy, and energy storages applications to power systems.

TAO DAI received the B.S. degree in electrical engineering from Hohai University, Nanjing, China, in 2021, where he is currently pursuing the M.S. degree. His main research interest includes state recognition technology of renewable energy units and its application in power systems.

WENJIE LAO received the B.S. degree in electrical engineering from Hohai University, Nanjing, China, in 2021, where she is currently pursuing the M.S. degree. Her main research interests include pumped storage technology, energy storage technology, and their application in power systems.

FENG WU (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Hohai University, Nanjing, China, in 1998 and 2002, respectively, and the Ph.D. degree from the University of Birmingham, U.K., in 2009.

He is currently a Professor of electrical engineering with the College of Energy and Electrical Engineering, Hohai University. His research interests include modeling and control of renewable power generation systems. He was a recipient of the National Outstanding Youth Science Foundation of China.

KEMAN LIN (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 2010 and 2016, respectively.

She is currently an Associate Professor of electrical engineering with the College of Energy and Electrical Engineering, Hohai University. Her research interests include modeling and control of renewable power generation systems.

KWANG Y. LEE (Life Fellow, IEEE) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, the M.S. degree in electrical engineering from North Dakota State University, Fargo, ND, USA, and the Ph.D. degree in system science from Michigan State University, East Lansing, MI, USA.

He has been a Faculty Member with Michigan State University, Oregon State University, the University of Houston, The Pennsylvania State University, and Baylor University, where he is currently a Professor and the Chair of electrical and computer engineering and the Director of the Power and Energy Systems Laboratory. His research interests include power systems control, operation and planning, and intelligent systems applications to power plant and power systems control. He was elected as a fellow of IEEE for his contributions to the development and implementation of intelligent system techniques for power plants and power systems control, in January 2001.