A Wind Farm Equivalent Method Based on Multi-View Transfer Clustering and Stack Sparse Auto Encoder

JI HAN, (Student Member, IEEE), SHIHONG MIAO, (Member, IEEE), YAOWANG LI, (Student Member, IEEE), WEICHEN YANG, AND HAORAN YIN
State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
Hubei Electric Power Security and High Efficiency Key Laboratory, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
Corresponding author: Shihong Miao (shmiao@hust.edu.cn)

This work was supported by The National Natural Science Foundation of China under Grant 51777088, and the Science and Technology Projects of State Grid Corporation of China (Key technologies and demonstration application of distributed power clusters regulation).

ABSTRACT Large-scale wind farm (WF) constitutes dozens or even hundreds of wind turbines (WTs), making it complex and even impractical to develop each individual WT in detail when building WF model. Thus, the equivalent model of WF, with a reasonable reduction of the detailed model, is essential to be developed. In this paper, we propose a multi-view transfer clustering and stack sparse auto encoder (SSAE) based WF equivalent method, which can be used in the low voltage ride through (LVRT) analysis of WF. First, to obtain distinguishable deep-level and multi-view representation of wind turbine (WT), stack sparse auto encoder (SSAE) is used to extract features from the time series of several WT physical quantities, and these features are used as the clustering indicator (CI). Then, a multi-view transfer FCM (MVT-FCM) clustering algorithm, which combines transfer learning with multi-view FCM (MV-FCM), is put forward for WT clustering. Two transfer rules are designed in this algorithm, and the clustering center and membership degree in the source domain are transferred to guide the clustering process of target domain samples. Finally, the calculation method of equivalent parameters is presented. To verify the effectiveness of the proposed method, a modified actual system in East Inner Mongolia of China is utilized for case study, and the performance of the proposed model is compared with several state-of-the-art models. Simulation results show that the equivalent errors of the proposed model decrease at least 3% when comparing with other models. Also, the error fluctuations are within 6% under different simulation conditions, which illustrates the well-performed robustness of the proposed model.

INDEX TERMS Wind farm equivalence, multi-view, transfer learning, deep learning, MVT-FCM, SSAE.

NOMENCLATURE

**A. STACK SPARSE AUTO ENCODER (SSAE)**

\[d\] the number of stacked layers

\[s_i\] the number of neurons in the \(i\)th layer (\(i = 1, \ldots, d\))

\[\rho\] sparsity parameter

\[\theta\] weight attenuation parameter

\[\delta\] weight coefficient

\[\text{MAPE}\] minimum mean absolute percentage error

\[y_j\] actual value of the sample

\[\hat{y}_j\] predicted values of the sample

\[R\] data size

**B. MV-FCM AND MVT-FCM**

\[m\] weighted index, and \(m \in \hat{E}(1, \infty)\)

\[N\] number of clustering patterns

\[K\] number of views

\[C\] number of clusters

\[x_{j,k}\] the \(j\)th clustering pattern in the \(k\)th view, \(j = 1, \ldots, N, k = 1, \ldots, K\)

\[v_{i,k}\] the \(i\)th clustering center in the \(k\)th view, \(i = 1, \ldots, C, k = 1, \ldots, K\)

\[d_{ij,k}\] distance between the pattern \(x_{j,k}\) and the clustering center \(v_{i,k}\) in view \(k\).

\[\mu_{ij,k}\] membership degree of \(x_{j,k}\) to \(v_{i,k}\)

\[\alpha_{j,k}\] importance coefficient of \(x_{j,k}\) in rewarding membership degree

\[\beta_{j,k}\] penalty coefficient used to weaken \(\alpha_{j,k}\)
**I. INTRODUCTION**

The installed capacity of wind turbines (WTs) would be 2.1 billion kW in China until 2020 [1]. In a wind farm (WF) system, voltage dip occurs frequently in its output port because of short-circuit faults in the power grid, thereby threatening the safe operation of the WF. The dynamic process of WF under voltage dips is defined as low voltage ride through (LVRT), and the studies of LVRT are essential for developing strategies of preventing WF off-grid and improving stability of power systems. Hence, a dynamic WF model, which can simulate the LVRT process accurately, is quite necessary to establish [2]. However, a WF usually contains dozens or even hundreds of WTs, and it is not feasible to model each WT in detail due to the huge model size and simulation time [3]. Thus, it is critical to develop the WF equivalent model for LVRT analysis on the basis of reasonable reduction [4].

In general, the WF equivalent methods can be categorized as two types: single-machine equivalence and multi-machines equivalence (MME). Studies have shown that MME-based models simulate WF characteristics better and have a border application in practice [5]. Three steps need to be taken in MME.

**Step 1:** Clustering indicator (CI) selection. The operation characteristics are different among the WTs in a WF, so it is important to select the proper CI to assess these differences. Until now, wind speed [3], pitch angle [4], rotor speed [5], and active power [6] have already been fully investigated to be assigned as CIs. These CIs are selected at specific time spot, and with time passes, these CIs might be changeable and the clustering results might not be valid in the whole time span. In [7], the time series of active power is selected as CI, and the clustering result is appropriate for the simulation in the selected time span, increasing the use range of the WF equivalent model. This paper also uses time series-based CIs because of their advantages in use range.

**Step 2:** WT clustering. This step aims to cluster the WTs with the same or similar operation characteristics into the same group, and pushes the WTs operating quite different to different groups. Fuzzy C-means (FCM) clustering [6], K-means clustering [8], hierarchical clustering [9] are the commonly used clustering methods.

**Step 3:** Equivalent parameters calculation. Based on the WT clustering in step 2, the WTs in the WF have been divided into several groups. Each group of WTs is called as one equivalent WT, and the parameters of the equivalent WTs need to be calculated. Equal power loss method [10] and equal voltage loss method [4] are the mostly used for collector network equivalence, where capacity weighted method is commonly used in the parameter estimation of equivalent transformer and equivalent WT [4].

Based on the above three steps, the MME-based WF equivalence is implemented. Although favorable equivalent results are achieved, some limitations are still existed in the present researches: 1) the existed time series-based CIs often focus on single-view, failing to reflect comprehensive aspects of
the WT operation characteristics; 2) the useful information in the time series-based CIs is somewhat hidden and need to be extracted; 3) the clustering results are easily changeable and inaccurate when using traditional clustering methods. The following is the analysis of these limitations.

The time series-based CIs often focus on single-view. The problem is that the single-view CI only reflects the WT operation characteristics in a few aspects, but fails to reflect in more comprehensive aspects. For example, active power time series is used as CI in [7]. Yet, the active power in some WTs might be similar, but their reactive power might perform quite different because different reactive power controls can be adopted in different WTs. Apparently, based on this CI, the final WF equivalent accuracy of active power might be high, but the accuracy of reactive power cannot be ensured. Therefore, the CI selection should focus on more views, by this way, the equivalent accuracy of more state variables in the equivalent WF can be ensured.

In addition, for the time series-based CIs, the existed researches cluster the time series directly. However, the useful information of the lengthy time series might only exist in some fragments, and the clustering algorithms are generally incapable of digging these useful features and might be affected by the useless fragments. Thus, the useful information extraction from the time series-based CIs, which is beneficial for the processing of clustering algorithms, is quite important.

In the WTs clustering step, the results of the traditional clustering methods can be influenced by the initialization, which is a random process. Thus, the clustering results might be changeable and unstable [11], and it is hard to choose the best one. If several WTs are pushed to the wrong groups, the final equivalent accuracy of WF would be influenced severely. Therefore, a clustering method that can acquire stable and accurate result of WTs clustering is essential to be developed.

To address the above three problems, this paper proposes an equivalent method for large-scale WF based on multi-view transfer clustering and stack sparse auto encoder (SSAE) [12], main contributions of this paper can be summarized as follows.

1) Time series of active power, reactive power, voltage and current, which have multi-view attribute, are used to evaluate the WT operation characteristics. Further, the useful information of these four quantities are extracted using SSAE, which is an effective tool for extracting features by making nonlinear variations to the original signals layer by layer. The multi-view extracted features are used as the CI in this paper. This CI can not only be used to evaluate the WT operation characteristics in more comprehensive aspects, but also convenient for clustering algorithm processing because of more obvious and distinguishable feature information.

2) To make the WTs clustering results more stable and accurate, a new clustering method namely multi-view transfer fuzzy C means (MVT-FCM) algorithm is put forward. In this method, transfer learning [13] mechanism is introduced to multi-view clustering [14]. Specifically, transfer learning migrates the beneficial knowledge of source domain, i.e., the source domain clustering center and membership degree, into the target domain to guide the clustering process. This auxiliary guidance significantly enhances the stability and accuracy of WTs clustering.

The paper is organized as follows. The multi-view CI, which is extracted using SSAE, for clustering WTs is presented in Section 2. In Section 3, MVT-FCM is proposed by applying transfer learning to MV-FCM, and the source domain data determination method is also put forward in this section. In Section 4, the equivalent parameter calculation method after WTs clustering is given. Section 5 provides the case study. Section 6 concludes the paper.

II. FEATURE EXTRACTION OF WT BASED ON SSAE
A. DESCRIPTION OF SSAE

SSAE belongs to the field of deep learning, and its basic unit is sparse self-encoder (SAE). Through combining the sparse penalty and reconstruction errors in the coding layer, SAE can effectively encourage model to learn the data feature in addition to copying the input to output. SAE is generally used in feature extraction, and the feature is used in classification or clustering algorithms. However, SAE cannot obtain deeper feature of data, since it only has one-layer [15]. Therefore, SSAE is proposed by stacking the hidden layer of previous SAE to the input layer of next SAE. The low-level features of the data are combined continuously, and the abstract, easily distinguishable high-level feature representation is finally obtained. The structure of the SSAE is shown in Fig. 1. Due to space limitations, the detailed theory of SSAE can be found in the [15].

![Network structure of stack sparse auto-encoder.](image)

B. FEATURE EXTRACTION FROM WT BASED ON SSAE

The global target of the WF equivalence is guaranteeing WF output characteristics consistent before and after the equivalence. Thus, a proper CI is important. In this paper, the time series of active power, reactive power, voltage and current are used to evaluate the WT operation characteristics. Further, SSAE is used to extract features from these four physical...
TABLE 1. Main factors that influence the output of wind turbine.

| Factor          | Influential mechanism                                                                 | Value range                                                                 |
|-----------------|---------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| Wind speed      | WT generally track the maximum wind power, and wind speed has the greatest impact on active power and current. | From cut-in speed (4.5 m/s) to cut-off speed (25 m/s) with the 0.2 m/s as step size. |
| Voltage dip     | The voltage dip value directly affects the WT voltage and is closely related to the WT reactive output during the short circuit. | From 0.2pu to 0.9pu with the 0.07pu as step size.                         |
| Control mode    | The recovery process after fault under constant voltage control and constant power factor control is different [16]. | Constant voltage control ($V_{ref}=1.1,1.05$); Constant power factor control ($p_{ref}=0.95,0.9,0.8$). |

![Diagram](image_url)

**FIGURE 2.** The determination of stacking number and number of neurons in SSAE model.

quantities. The multi-view extracted features are used as the CI in this paper. To extract these features, the method of acquiring samples used to train SSAE and the SSAE construction method are given as follows.

To acquire a large amount of samples used in the training of SSAE, the important factors that affect active power, reactive power, voltage and current of WT are analyzed, as shown in TABLE 1. In TABLE 1, each factor has multiple values. The value combinations of the three factors are used as the input of WT, and there are $102 \times 11 \times 5 = 5610$ compound modes. The time series of the four physical quantities under these 5610 simulation experiments are used as the samples to train SSAEs.

Based on the acquired samples, the SSAE models are constructed for active power, reactive power, voltage and current, respectively. In the construction of SSAEs, the number of stacked layers and the number of neurons in per layer have an important influence on the final feature extraction [17]. In this paper, the determination method of stacked layer number and neuron number in per layer is shown in Fig. 2, that is, the optimal number of neurons in each layer is determined layer by layer, and the optimal number of stacked layers is determined according to the error after increasing the number of layers [17].

Specifically, some numbers of neurons in the first layer are set, and the training experiments are performed, respectively, the number corresponding to the minimum mean absolute percentage error (MAPE), which is shown in (1), is selected as the number of neurons in the first layer. Then, fix the number of neurons in the first layer, and obtain MAPE in the second layer using the same method. In the first few layers, MAPE would generally decrease because the feature extraction precision is higher, then, MAPE would increase due to over-fitting. Therefore, the number of stacked layers is taken as the optimal number when MAPE starts to rise.

$$I_{MAPE} = \frac{1}{R} \sum_{i=1}^{R} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

### III. WTS CLUSTERING BASED ON MULTI-VIEW TRANSFER LEARNING

#### A. TRADITIONAL MV-FCM

MV-FCM [14] is essentially a multi-view clustering method, which could reward or suppress the corresponding fuzzy membership degrees for the current view, and combines the clustering results in each view into global clustering result. Assuming that the data set can be divided into $C$ clusters, the objective function and constraints of the MV-FCM can be expressed as

$$\min \left\{ \sum_{k=1}^{K} \sum_{j=1}^{N} \sum_{i=1}^{C} [\mu_{ij,k}^m d_{ij,k}^2 + \alpha_{ij,k} \mu_{ij,k}(1 - \mu_{ij,k}^{m-1})] - \beta_{ij,k} \mu_{ij,k}(1 - \mu_{ij,k}^{m-1}) \right\}$$

subject to

$$\sum_{j=1}^{C} \mu_{ij,k} = 1 \quad \mu_{ij,k} \in [0, 1]$$

$$1 \leq i \leq C \quad 1 \leq j \leq N \quad 1 \leq k \leq K$$
Using the Lagrange optimization, the update equations of the clustering center \( v_{i,k} \) and the membership \( \mu_{ij,k} \) can be derived as
\[
\mu_{ij,k} = \frac{1}{C} \sum_{i=1}^{C} \left[ \frac{d_{ij,k}^2 - (\alpha_{ij,k} - \beta_{ij,k})}{d_{ij,k}^2} \right]^{\frac{1}{m-1}} 
\]
\[
v_{i,k} = \sum_{j=1}^{N} \mu_{ij,k}^m \bar{x}_j / \sum_{j=1}^{N} \mu_{ij,k}^m
\]

**B. THE PROPOSED CLUSTERING METHOD MVT-FCM**

We declare firstly, if two variables use the same symbol, the variable with top horizontal line is from source domain, and that without top horizontal line is from target domain, the meanings of the two variables are totally the same.

In recent years, transfer learning has received widely attention. The learning method provides a certain reference for the target domain by referring to the source domain information. In the clustering analysis, transfer learning can effectively deal with the problem of low clustering accuracy and instability [18]. This paper introduces transfer learning into MV-FCM to improve the stability of WTs clustering results.

We firstly transfer the clustering center of source domain to target domain, and the first transfer rule is
\[
T(\bar{v}_{i,k}) = \lambda \sum_{k=1}^{K} \sum_{j=1}^{C} ||v_{i,k} - \bar{v}_{i,k}||^2
\]

In MV-FCM, \( \mu_{ij,k}^m d_{ij,k}^2 \) is used to estimate the deviation between the data and the clustering center, and \( \mu_{ij,k} \) is used as weighting factor. To transfer \( \bar{\mu}_{ij,k} \) to target domain, we define the second transfer rule
\[
T(\bar{\mu}_{ij,k}) = \lambda \sum_{k=1}^{K} \sum_{j=1}^{N} \sum_{i=1}^{C} \bar{\mu}_{ij,k}^m d_{ij,k}^2
\]

Based on (6)-(7), the optimization objective and constraints of MVT-FCM are
\[
\min \left\{ \sum_{k=1}^{K} \sum_{j=1}^{C} \sum_{i=1}^{N} \left[ \gamma \mu_{ij,k}^m d_{ij,k}^2 + (1 - \gamma) \bar{\mu}_{ij,k}^m d_{ij,k}^2 \right] \right. \\
+ \alpha_{ij,k} \mu_{ij,k}(1 - \mu_{ij,k}^{-m}) \\
- \beta_{ij,k} \mu_{ij,k}(1 - \mu_{ij,k}^{-m}) \right\} \\
\text{s.t. } \sum_{j=1}^{N} \mu_{ij,k} = 1; \quad \mu_{ij,k} \in [0, 1]; \\
1 \leq i \leq C; \quad 1 \leq j \leq N; \quad 1 \leq k \leq K
\]

The necessary conditions for minimizing the objective function in (8) yields the following clustering center and membership update equations
\[
\mu_{ij,k} = \frac{1}{C} \sum_{i=1}^{C} \left[ \frac{\gamma d_{ij,k}^2 - (\alpha_{ij,k} - \beta_{ij,k})}{\gamma d_{ij,k}^2 - (\alpha_{ij,k} - \beta_{ij,k})} \right]^{\frac{1}{m-1}} 
\]

**C. DATA SELECTION TRANSFERRED FROM SOURCE DOMAIN**

In this paper, the multi-view features of active power, reactive power, voltage and current in each WT acquired from SSAE reflect the operation characteristics of the WTs in WF.

Thus, these extracted features are regarded as target domain data. The extracted features of the four quantities under the 5610 experiments have abundant data volume, and possess the knowledge to guide the clustering of target domain data.

Therefore, the extracted features in the 5610 experiments are regarded as source domain data.

The premise of MVT-FCM is determined by source domain data, which is used to guide the clustering of target domain data. However, not all source domain data can play a positive role in the clustering of target domain data. If the data distribution of the two domains are different, the introducing of the source domain data would worsen the target domain calculation result. In transfer learning, the similarity test is generally used to determine source domain data that could be transferred to target domain [19].

In the LVRT experiments, wind speed, voltage dip and control method are selected as the main factors affecting the WT output. Therefore, we consider the three factors to judge the similarity of the data in source and target domain, and only transfer the source domain samples that are most similar to target domain. Define the comprehensive similarity coefficient, which is as follows

\[
s_{ij} = \begin{cases} 
\left. \frac{w_1}{W_1} \frac{V_{ref,i} - \bar{V}_{ref,j}}{V_{ref,j}} + w_2 \frac{W_i - \bar{W}_j}{W_j} \right| \\
+ w_3 \frac{V_{dip,i} - \bar{V}_{dip,j}}{V_{dip,j}} \\
(i = 1, \ldots, N_v, j = 1, \ldots, M_v) \\
\left. \frac{w_1}{W_1} \frac{P_{ref,i} - \bar{P}_{ref,j}}{P_{ref,j}} \right| + w_2 \frac{W_i - \bar{W}_j}{W_j} \\
+ w_3 \frac{V_{dip,i} - \bar{V}_{dip,j}}{V_{dip,j}} \\
(i = 1, \ldots, N_p, j = 1, \ldots, M_p) 
\end{cases}
\]

In (13), \( M_v = 102 \times 11 \times 2 = 2244 \), \( M_p = 102 \times 11 \times 3 = 3366 \), representing the number of LVRT experiments that adopt constant voltage control and constant power factor control, respectively. The smaller value of \( s_{ij} \), the more similar
of the operation characteristics between the $i$th WT and the $j$th LVRT experiment. As for the $i$th WT, select the $n_{\text{sim}}$ LVRT experiments that are most similar to it, namely $\text{Ex}_i$. Finally, the multi-view features corresponding to the following LVRT experiments are transferred to target domain.

$$\text{Ex}_\text{tr} = \bigcup_{i=1}^{N} \text{Ex}_i$$  \hspace{1cm} (14)

### D. WTs CLUSTERING AFTER INTRODUCING SOURCE DOMAIN DATA

In this paper, SSAE is used to extract features of active power, reactive power, voltage and current of WT, and these features are input to MVT-FCM to achieve the clustering of WTs. The WTs Clustering process after introducing source domain data can be represented by Fig. 3.

1. According to the wind speed, voltage dip value and control mode of each WT, the active power, reactive power, voltage and current of each WT are obtained through simulation.

2. The active power, reactive power, voltage and current time series of WT are input into the trained SSAE, and the features of four types of physical quantities is obtained, which are used as target domain data for WTs clustering.

3. Select the appropriate $n_{\text{sim}}$, and use (13) and (14) to acquire the LVRT experiments for data transfer. The active power, reactive power, voltage and current in these selected experiments are input into the trained SSAE, and the features are used as source domain data for WTs clustering.

4. For MVT-FCM, set the algorithm parameters, and input the source domain and the target domain data. Finally, the clustering result based on the proposed MVT-FCM algorithm is obtained.

### IV. CALCULATION OF EQUIVALENT PARAMETERS AFTER WTs CLUSTERING

After WTs clustering using the MVT-FCM algorithm, the topology of the equivalent WF is shown in Fig. 4.

The parameters of the equivalent WT, equivalent transformer and equivalent collector network need to be calculated.

As for equivalent WT, the equivalence of wind speed, WT capacity, stator and rotor winding reactance and resistance, shafting inertia time constant, shafting stiffness coefficient, shafting damping coefficient could be found in [20]. For equivalent transformer, the equivalence of capacity and reactance could be found in [20]. As for equivalent collector network, the equivalence of resistance, reactance and capacitance could be found in [10].

Besides, constant power factor control and constant voltage control are adopted in this paper. The control mode in the $i$th equivalent WT follows the following principle:

1. Calculate the proportion of constant power factor control in all WTs, denoted as $p_{\text{pf}}$. Then there would be $N_{\text{pf}} = \lfloor p_{\text{pf}} \times C \rfloor$ equivalent WTs taking constant power factor control, where $C$ is the number of equivalent WTs, and $\lfloor \rfloor$ is the integer symbol following the rounding rule.

2. Calculate the proportion of constant power factor control in each equivalent WT, denoted as $p_{\text{pf},i}$. Take the first $N_{\text{pf}}$ with larger $p_{\text{pf},i}$, then the corresponding equivalent WTs would take constant power factor control, and other equivalent WTs would take constant voltage control.
V. SIMULATION RESULTS

A. THE CONSTRUCTIONS OF SSAEs

To acquire the extracted features as the multi-view CI, the SSAEs need to be well-trained. Using the wind speeds, voltage dips and control modes shown in TABLE 1, a total of 5,610 LVRT experiments are performed to acquire the active power, reactive power, voltage and current time series, which are used as the training samples of SSAEs. In these LVRT experiments, three-phase short-circuit fault occurs at the output port of the WT, and the fault occurs at 90s for 0.15s. Among these time series, Fig. 5 shows the active power curves of the WT at various experimental conditions.

In the training of SSAEs, the time series of active power, reactive power, voltage and current from 89.9s to 90.5s are selected as the input of the SSAEs, respectively. Set $\rho = 0.01$, $\theta = 0.0001$, $\delta = 3$, and Fig. 6 shows $I_{MAPE}$ under different number of stacked layers and neurons. The stacked layers $d$, and the number of neurons in per layer $s_i (i = 1, \ldots, d)$ can be acquired using the method in Fig. 2. Taking active power as an example, from Fig. 6 (a), $I_{MAPE}$ performs a downward trend when the stacked layer number increases to four. However, $I_{MAPE}$ increases when the stacked layer number is five. Therefore, the number of stacked layers of active power SSAE is four. At the same time, considering that the data dimension input to the MVT-FCM should be consistent in each view, the neurons number in the last layer of the SSAE of four type physical quantities should be the same. Finally, the number of stacked layers of SSAEs of active power, reactive power, voltage and current and the number of neurons in per layer are shown in TABLE 2.

B. THE TEST WF AND ITS EQUIVALENT RESULT

In this paper, the simulation is carried out based on an actual system in East Inner Mongolia of China and a certain degree of expansion. The layout of the test case is shown in Fig. 7.
FIGURE 7. The layout of the modified wind farm consisting of 61 wind turbines.

In this system, the WF consists of 61 WTs, and all of the WTs connect to collector lines by 0.69/35kV box transformers and access to external system by 35/220kV transformer. Main parameters of the system are shown in TABLE 3.

Constant voltage control and constant power factor control are randomly employed to the WTs in the WF at the ratio of 1:2. Set three-phase short-circuit fault at PCC with voltage dip value 0.2pu. When fault occurs, the wind speed range is set between 10m/s and 20m/s, and the wind speed of each WT is randomly set in this interval. Specifically, the wind speed, voltage dip and control mode of each WT in the WF are shown in Fig. 8, and the time series of active power, reactive power, voltage and current in each WT are simulated.

The time series of these four quantities are then input to the well-trained SSAEs, respectively, and the extracted features are shown in Fig. 9. In Fig. 9 (a)-(d), each curve represents the extracted features of a WT in one view, and the dimension of the features is 500, in other words, each curve is a 500-dimensional vector. The multi-view extracted features in Fig. 9 are selected as the target domain samples, which would be used as the target domain data in MVT-FCM for WTs clustering.

Select \( n_{sim} = 20 \), and the LVRT experiments that are most similar to WTs in WF are obtained using equations (13) and (14), totally \( 20 \times 61 = 1220 \) WT experiments. The multi-view extracted features corresponding to these 1220 experiments are selected as the source domain samples, which would be used as the source domain data in MVT-FCM for WTs clustering.

Using the above target domain and source domain samples, the WTs in the WF can be clustered based on MVT-FCM. According to the transfer rules shown in (6) and (7), not the source domain samples are transferred to target domain directly, but the clustering center \( \bar{v}_{i,k} \) and membership degree \( \bar{\mu}_{ij,k} \) are transferred. Thus, the \( \bar{v}_{i,k} \) and \( \bar{\mu}_{ij,k} \) are calculated based on the source domain samples using the traditional MV-FCM. The parameters setting of MV-FCM is shown in TABLE 4, then the \( \bar{v}_{i,k} \) and \( \bar{\mu}_{ij,k} \) can be computed using (4) and (5), iteratively.

After acquiring \( \bar{v}_{i,k} \) and \( \bar{\mu}_{ij,k} \) of source domain, these two items are transferred and used to guide the clustering of the target domain samples. The parameters setting of MVT-FCM is also shown in TABLE 4. Based on the target domain samples and the transferred items \( \bar{v}_{i,k} \) and \( \bar{\mu}_{ij,k} \), the \( v_{i,k} \) and \( \mu_{ij,k} \) are computed using (10) and (11), iteratively. Finally, the \( \hat{\mu}_{ij,k} \) is calculated using (12), and the result is shown in Fig. 10. In Fig. 10, the darker of the color, the larger of the value. The largest value in each column is circled, and for each row, all the circled WTs are clustered to a group. Through statistics, the WTs clustering results are obtained, and are shown in TABLE 5. After WTs clustering, the equivalent parameters are calculated. As for the control mode, \( \text{Weq1} \) takes constant...
TABLE 3. Parameters of component in the power systems.

| Component             | Value                                                                 |
|-----------------------|----------------------------------------------------------------------|
| Wind turbine          | Nominal power, L-L voltage 1.5MW and 575V;                           |
|                       | Stator inductance and resistance 0.18pu and 0.023pu;                |
|                       | Rotor inductance and resistance 0.16pu and 0.016pu;                 |
|                       | Magnetizing inductance 2.9pu;                                       |
|                       | Grid-side coupling inductor 0.3pu and 0.03pu;                       |
|                       | DC bus capacitor 0.01F;                                             |
| Box transformer (35kV/690V) | Winding 1 parameters: L-L voltage, inductance and resistance 35e3V, 0.025pu and 0.025/30pu; |
|                       | Winding 2 parameters: L-L voltage, inductance and resistance 690V, 0.025pu and 0.025/30pu; |
|                       | Magnetization inductance and resistance 10e4pu and 500pu.           |
| Main transformer (220kV/35kV) | Winding 1 parameters: L-L voltage, inductance and resistance 220e3V, 0.08pu and 0.08/30pu; |
|                       | Winding 2 parameters: L-L voltage, inductance and resistance 35e3V, 0.08pu and 0.08/30pu; |
|                       | Magnetization inductance and resistance 500pu and 500pu.            |
| Collector network     | Positive-sequence resistances 3.459Ohms;                           |
|                       | Positive-sequence inductances 3.15e-2H;                             |
|                       | Positive-sequence capacitances 3.399e-7F.                           |

FIGURE 9. Extracted target domain features of WTs. (a) Active power features; (b) Reactive power features; (c) Voltage features; (d) Current features; (e) Reactive power features of a certain WT.

TABLE 4. Parameters setting of MV-FCM and MVT-FCM.

| Method      | Parameters                                                                 |
|-------------|-----------------------------------------------------------------------------|
| MV-FCM      | $C=4$, $m=2$, $\alpha=0.5$ and $\beta=0.3$                                |
| MVT-FCM     | $C=4$, $m=2$, $\lambda=1$, $\beta=0.8$, $\alpha=0.5$ and $\beta=0.3$     |

C. SUPERIORITY OF THE PROPOSED WF EQUIVALENT METHOD

In this paper, the features of time series are extracted using SSAE. In addition, MVT-FCM is proposed for increasing the voltage control. Weq2-Weq4 take constant power factor control. Other parameters can be calculated using the methods in [10], [20], and this paper would not list them.

TABLE 5. Clustering result of wind turbines.

| Equivalent WT | WT number |
|---------------|-----------|
| W_1           | 6, 7, 18, 31, 32, 38, 47, 57, 58, 59, 60 |
| W_2           | 8, 9, 10, 11, 19, 20, 21, 22, 39, 40, 41, 48, 49, 50, 61 |
| W_3           | 1, 3, 12, 13, 15, 16, 23, 24, 28, 30, 33, 34, 35, 36, 42, 43, 44, 51, 52, 53, 56 |
| W_4           | 2, 4, 5, 14, 17, 25, 26, 27, 29, 37, 45, 46, 54, 55 |
FIGURE 10. The final membership degree of the WTs in each cluster.

TABLE 6. The proposed model and Other MME models for comparison.

| Model       | Clustering indicator                          | Clustering method |
|-------------|-----------------------------------------------|-------------------|
| M0 (Proposed model) | The extracted features of 4 physical quantities using SSAE | MVT-FCM |
| M1          | The time series of 4 physical quantities without feature extraction | MVT-FCM |
| M2          | The extracted features of 4 physical quantities using SSAE | MV-FCM |
| M3          | The time series of 4 physical quantities without feature extraction | MV-FCM |
| M4          | The extracted features of active power using SSAE | FCM |
| M5          | The time series of active power without feature extraction | FCM |

Note: The 4 physical quantities denote the time series of active power, reactive power, voltage and current.

M1 is used to illustrate the superiority of feature extraction, M2-M3 are used to illustrate the advantages of the application of transfer learning in clustering process, and M4-M5 are used to illustrate the benefits of multi-view consideration in CI selection.

To verify the accuracy of the proposed WF equivalent model in statistical level, root mean square error ($e_{RMSE}$), mean absolute error ($e_{MAE}$), and mean absolute percentage error ($e_{MAPE}$) are introduced in the paper, which are widely used to evaluate the error of time series. The expressions of $e_{RMSE}$, $e_{MAE}$ and $e_{MAPE}$ are shown in (15). The $e_{RMSE}$, $e_{MAE}$ and $e_{MAPE}$ of M0-M5 are shown in TABLE 7. In TABLE 7, by comparing the value of $e_{RMSE}$, $e_{MAE}$ and $e_{MAPE}$ in M0-M5, M0 performs better in the accuracy aspect than the other five models. The reasons are listed as follows.

First, as for M0 and M1, with the help of SSAE in M0, the clustering algorithm can identify the hard-to-find data distribution characteristics. Yet, there is no feature extraction in M1, and the data features might only reflect in some fragments of lengthy time series, thereby influencing the clustering accuracy.

With regard to M0 and M2/M3, although feature extraction steps are applied to both M0 and M2, there is no introducing of transfer learning in the clustering process of M2. Thus, only the target domain samples are clustered in M2 without the reference of source domain knowledge, and the samples without obvious distribution characteristics might be pushed to the wrong groups. For M3, there is also no feature extraction step except for no transfer learning, so its equivalent accuracy is even worse than M2.

$$
e_{RMSE} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} (X(l) - \hat{X}(l))^2}$$

$$e_{MAE} = \frac{1}{L} \sum_{l=1}^{L} |X(l) - \hat{X}(l)|$$

$$e_{MAPE} = \frac{1}{L} \sum_{l=1}^{L} \left| \frac{X(l) - \hat{X}(l)}{X(l)} \right|$$

TABLE 7. The simulation errors ($e_{RMSE}$, $e_{MAE}$ and $e_{MAPE}$) in different models.

| Models       | Physical quantities | M0 | M1 | M2 | M3 | M4 | M5 |
|--------------|---------------------|----|----|----|----|----|----|
|              | Error types         |    |    |    |    |    |    |
|              | $e_{RMSE}$          |    |    |    |    |    |    |
| M0 (Proposed model) | Active power       | $7.24 \times 10^2$ | $1.17 \times 10^2$ | $1.33 \times 10^2$ |
|              | Reactive power      | $4.73 \times 10^2$ | $2.71 \times 10^2$ | $3.04 \times 10^2$ |
|              | Voltage             | $6.77 \times 10^3$ | $1.54 \times 10^3$ | $1.89 \times 10^3$ |
|              | Current             | $4.35 \times 10^3$ | $3.57 \times 10^3$ | $4.00 \times 10^3$ |
| M1           | Active power        | $9.04 \times 10^2$ | $1.44 \times 10^2$ | $1.67 \times 10^2$ |
|              | Reactive power      | $5.47 \times 10^2$ | $2.92 \times 10^2$ | $3.60 \times 10^2$ |
|              | Voltage             | $8.74 \times 10^3$ | $1.96 \times 10^3$ | $2.37 \times 10^3$ |
|              | Current             | $5.56 \times 10^3$ | $4.43 \times 10^3$ | $5.18 \times 10^3$ |
| M2           | Active power        | $1.45 \times 10^3$ | $2.47 \times 10^3$ | $2.76 \times 10^3$ |
|              | Reactive power      | $2.23 \times 10^3$ | $2.39 \times 10^3$ | $2.67 \times 10^3$ |
|              | Voltage             | $1.65 \times 10^3$ | $4.36 \times 10^3$ | $5.34 \times 10^3$ |
|              | Current             | $8.91 \times 10^3$ | $6.58 \times 10^3$ | $7.34 \times 10^3$ |
| M3           | Active power        | $1.76 \times 10^3$ | $2.95 \times 10^3$ | $3.29 \times 10^3$ |
|              | Reactive power      | $2.35 \times 10^3$ | $2.78 \times 10^3$ | $3.28 \times 10^3$ |
|              | Voltage             | $2.57 \times 10^3$ | $5.52 \times 10^3$ | $6.27 \times 10^3$ |
|              | Current             | $1.14 \times 10^3$ | $7.53 \times 10^3$ | $9.25 \times 10^3$ |
| M4           | Active power        | $1.45 \times 10^3$ | $3.18 \times 10^3$ | $3.77 \times 10^3$ |
|              | Reactive power      | $3.05 \times 10^3$ | $3.28 \times 10^3$ | $3.68 \times 10^3$ |
|              | Voltage             | $2.22 \times 10^3$ | $5.91 \times 10^3$ | $6.72 \times 10^3$ |
|              | Current             | $1.17 \times 10^3$ | $8.96 \times 10^3$ | $1.00 \times 10^3$ |
| M5           | Active power        | $1.86 \times 10^3$ | $4.09 \times 10^3$ | $4.73 \times 10^3$ |
|              | Reactive power      | $3.78 \times 10^3$ | $3.61 \times 10^3$ | $4.49 \times 10^3$ |
|              | Voltage             | $2.76 \times 10^3$ | $7.31 \times 10^3$ | $8.26 \times 10^3$ |
|              | Current             | $1.47 \times 10^3$ | $1.11 \times 10^3$ | $1.24 \times 10^3$ |
With M0 and M4/M5, M4 only use single-view extracted features as the CI, which is corresponding to active power, so the accuracy of active power is quite high, but the accuracy of other physical quantities cannot be ensured. This illustrates the necessity of the multi-view consideration in CI selection. With M5, the time series of active power is directly used as the CI, which features are hard-to-find for the clustering algorithm. Thus, its accuracy is not comparable to other models.

D. PERFORMANCE OF THE PROPOSED CLUSTERING ALGORITHM

The superiority of the presented WF equivalent method is discussed qualitatively in the above Section V-C. In particular, the proposed clustering algorithm MVT-FCM plays a very helpful role in increasing the equivalent accuracy because of the introducing of transfer learning. Thus, in this part, the advantages of MVT-FCM are analyzed in a quantitative manner by comparing with other state-of-the-art clustering algorithms. Except for the used MV-FCM and FCM in M2-M5, several novel multi-view clustering algorithms, including multi-view fuzzy clustering (MV-FC) [21], multi-view fuzzy k-Means (MV-FKM) [22], and multi-view expectation maximization (MV-EM) [23], are also used to compare with MVT-FCM. For MV-FC, a local inertia term is defined by the inertia of the fuzzy clusters in a certain view, and penalized by a disagreement with the other views. MV-FKM makes use of a penalty term in its objective function, which aims at reducing the disagreement between organizations on different views. With MV-EM, the objective function combines the local log-likelihoods from any views and a disagreement term.

The test datasets include two parts, i.e., multi-view extracted features of source and target domains acquired in Section V-B, and the multi-view time series of source and target domains acquired in Section V-B. Yet, not all the samples in the dataset are required to be fed to some algorithms because of their principles. More specifically, transfer learning is introduced into MVT-FCM, thus, the multi-view samples in both domains are fed to it to acquire the clustering result. For MV-FCM, MV-KM, MV-FKM and MV-EM, there is no transfer learning, so only the multi-view samples in target domain are fed to these algorithms. FCM is a single-view clustering algorithm without transfer learning, so the single-view samples (we use active power) in the target domain are fed to it.

Three popular validity indices are adopted in this paper, i.e., DVI, SI, and DBI [24], for verifying clustering performance of these algorithms. Larger values of DVI and SI indicate better clustering performance. In contrast, smaller values of DBI are preferred. Besides, smaller values of the standard deviation mean the stability of the clustering is stronger. The experimental results are reported in terms of the means and the standard deviations, and are calculated after 100 repeated runs. The clustering performances in different clustering algorithms are shown in TABLE 8.

| Algorithms | Test dataset | Value | DVI | SI | DBI |
|------------|--------------|-------|-----|----|-----|
| MVT-FCM    | Extracted features | Mean | 0.9132 | 0.9554 | 0.5239 |
|            | Time series   | Mean | 0.9048 | 0.9443 | 0.5946 |
| MV-FCM     | Extracted features | Mean | 0.7781 | 0.9066 | 0.8830 |
|            | Time series   | Mean | 0.8489 | 0.9263 | 0.8082 |
| FCM        | Extracted features | Mean | 0.7256 | 0.8566 | 0.6614 |
|            | Time series   | Mean | 0.7015 | 0.8984 | 0.8461 |
| MV-FKM     | Extracted features | Mean | 0.8127 | 0.8973 | 0.9600 |
|            | Time series   | Mean | 0.7115 | 0.8184 | 0.7458 |
| MV-EM      | Extracted features | Mean | 0.8959 | 0.8400 | 0.8801 |
|            | Time series   | Mean | 0.8829 | 0.9612 | 0.9284 |

In TABLE 8, the mean values of DVI and SI in MVT-FCM are apparently larger than those of other clustering algorithms, and the DBI values of MVT-FCM are smaller. This indicates that better performances are acquired by MVT-FCM. The reason is that MVT-FCM introduces the knowledge from source domain using transfer learning, and the transferred knowledge remedies the disadvantages of unobvious data distribution in the target samples.

In addition, the standard deviation values of DVI, SI, and DBI in MVT-FCM are 0, which illustrates that the same clustering results are acquired in the 100 repeated runs. Conversely, the standard deviation values in other algorithms are larger than 0, which means different clustering results are acquired. This is also because the algorithms without the introducing of transfer learning are sensitive to the initialization. Thus, the applying of transfer learning mechanism is quite helpful to increase the stability of clustering results.

Comparing the clustering performances in the two test datasets, we can see the performances are better when using the extracted features as the clustered samples. Take the clustering performances of MVT-FCM as example, the DVI and SI are larger and the DBI is smaller when clustering extracted features. The reason lies in that data features of time series have been fully mined using SSAEs, these features are easily recognizable for clustering algorithms. For the time series, the features might only reflect in some fragments, and it is difficult for the clustering algorithms to mine these features. This indicates the positive effect of feature extraction in WTs clustering.

The proposed MVT-FCM introduces transfer learning in its clustering process, thereby increasing the algorithm complexity when comparing with the clustering algorithms in TABLE 8. Thus, the complexity of the proposed MVT-FCM
is analyzed. Generally, the algorithm complexity can be divided into two aspects, i.e., time complexity and space complexity. The computer run time of the algorithm can be used to denote the time complexity, and the maximum memory usage can denote the space complexity. It should be noted that clustering center matrix and membership degree matrix in the source domain are transferred to target domain, and the two matrices are computed using the traditional MV-FCM. Therefore, only the calculation process using the target domain samples and the acquired two matrices are taken into account to assess the complexity of MVT-FCM.

The above two test datasets are also used to assess the algorithm complexity. The computer configuration is Intel(R) Core(TM) i5-8250, CPU@1.60 GHz, 8.00 GB of RAM. For each dataset, the algorithm repeated runs for 100 times, and the mean results are calculated. TABLE 9 shows the complexity assessment results of the proposed method and several compared methods. From TABLE 9, we can see the proposed clustering algorithm is slightly more time-consuming than other multi-view clustering algorithms. For example, the run time of MVT-FCM increases about 36% when comparing with MV-FCM. This is because the introducing of the knowledge from the source domain increase the complexity of the objective function of MVT-FCM. Yet, this increasing in run time is quite acceptable because the better clustering accuracies and stabilities are acquired. In the aspect of space complexity, the maximum memory usage of MVT-FCM has not increased too much. This is because only the source domain knowledge (the source domain clustering center and membership degree), but not the source domain samples are transferred to the target domain. Thus, the data amount that are required to be processed for MVT-FCM increases slightly. This also indicates that the proposed MVT-FCM has the ability to process the WTs clustering of the WF with larger scale.

TABLE 9. Complexity assessment results of several clustering methods.

| Algorithms | Test dataset | Complexity |
|------------|--------------|------------|
| MVT-FCM    | Extracted features | Time complexity (run time/s) 0.113 |
|            |               | Space complexity (maximum memory usage/%) 12.74 |
| MV-FCM     | Time series   | 1.44       |
|            |               | 27.48      |
| MV-FC      | Extracted features | 0.083 |
|            | Time series   | 1.29       |
|            |               | 25.36      |
| FCM        | Extracted features | 0.045 |
|            | Time series   | 0.34       |
|            |               | 5.88       |
| MV-FC      | Extracted features | 0.079 |
|            | Time series   | 1.04       |
|            |               | 20.58      |
| MV-PFM     | Extracted features | 0.104 |
|            | Time series   | 1.34       |
|            |               | 23.69      |
| MV-UM      | Extracted features | 0.080 |
|            | Time series   | 1.28       |
|            |               | 24.72      |

E. COMPARISON WITH OTHER WF EQUIVALENT METHODS

Besides comparing the proposed model M0 with our designed contrast models M1-M5, the equivalent results are also compared with several novel WF equivalent approaches in other literatures [7], [25], [26]. In [7], the researchers proposed a WF equivalent model taking the WF output consistency at a selected time span as objective. In [25], a dynamic voltage equivalent method for WF is proposed, and the randomness and fuzziness of WTs are used to evaluate the similarity among WTs based on cloud model. In [26], the WF equivalent model is established in the view of energy conservation.

To facilitate the following analysis, the used CI and clustering algorithms of the models in [7], [25], [26] are listed in TABLE 10. We also use the WF system shown in Fig. 5 as the test system. In addition, the experimental conditions in Section V-B are used. The clustering algorithm in each compared model repeated runs 100 times, and we random select one result from these runs for each model.

TABLE 10. The used CI and clustering algorithm of the models in [7], [25], [26].

| Literature | Used CI | Used clustering algorithm |
|------------|---------|--------------------------|
| [7]        | Time series of active power | Geometric template matching and quality threshold clustering |
| [25]       | Three cloud model characteristic quantities of dynamic voltage | k-means clustering |
| [26]       | Rotating speed of WT when fault is cleared | k-means clustering |

The $e_{\text{RMSE}}$, $e_{\text{MAE}}$ and $e_{\text{MAPE}}$ values of the compared models are calculated, and are shown in TABLE 11. In addition, the DVI, SI, and DBI values of these compared models are shown in TABLE 12. It should be noted that the error results and the clustering performances of the proposed model are shown in TABLE 7 and TABLE 8. By comparing the values of $e_{\text{RMSE}}$, $e_{\text{MAE}}$ and $e_{\text{MAPE}}$ between the proposed model and the compared models, the proposed model performs the same good or even better in the accuracy aspect than the other three compared models. Meanwhile, the values of DVI, SI and DBI also show an advantage comparing with other models. The analysis are listed as follows.

TABLE 11. The simulation errors ($e_{\text{RMSE}}$, $e_{\text{MAE}}$, $e_{\text{MAPE}}$) in different models.

| Models | Physical quantities | Error types |
|--------|---------------------|-------------|
|        |                     | $e_{\text{RMSE}}$ | $e_{\text{MAE}}$ | $e_{\text{MAPE}}$ |
| [7]    | Active power        | $7.57 \times 10^{-2}$ | $1.08 \times 10^{-2}$ | $2.67 \times 10^{-2}$ |
|        | Reactive power      | $1.86 \times 10^{-2}$ | $9.24 \times 10^{-2}$ | $8.93 \times 10^{-2}$ |
|        | Voltage             | $2.17 \times 10^{-2}$ | $7.83 \times 10^{-2}$ | $8.22 \times 10^{-2}$ |
|        | Current             | $1.86 \times 10^{-2}$ | $2.06 \times 10^{-2}$ | $2.60 \times 10^{-2}$ |
| [25]   | Active power        | $1.57 \times 10^{-1}$ | $6.32 \times 10^{-2}$ | $9.45 \times 10^{-2}$ |
|        | Reactive power      | $2.14 \times 10^{-3}$ | $9.91 \times 10^{-4}$ | $2.11 \times 10^{-3}$ |
|        | Voltage             | $1.02 \times 10^{-3}$ | $4.38 \times 10^{-4}$ | $3.20 \times 10^{-3}$ |
|        | Current             | $1.82 \times 10^{-2}$ | $3.26 \times 10^{-2}$ | $4.40 \times 10^{-2}$ |
| [26]   | Active power        | $8.88 \times 10^{-2}$ | $2.05 \times 10^{-2}$ | $4.54 \times 10^{-2}$ |
|        | Reactive power      | $1.75 \times 10^{-2}$ | $8.22 \times 10^{-2}$ | $7.81 \times 10^{-2}$ |
|        | Voltage             | $3.05 \times 10^{-2}$ | $7.54 \times 10^{-2}$ | $8.81 \times 10^{-2}$ |
|        | Current             | $2.17 \times 10^{-2}$ | $3.05 \times 10^{-2}$ | $3.64 \times 10^{-2}$ |

The compared models all use single-view CI for WTs clustering, thus, their accuracies are higher in the aspect
corresponding to their CIs. Specifically, the model in [7] use active power as the CI, so its active power accuracy is higher than other physical quantities. Similarly, for the model in [25], the voltage accuracy is higher, this is because the dynamic voltage is selected as the CI in this model. The model in [26] clusters the WTs using their rotating speed, which is a physical quantity relevant to active power, so this model performs better in the equivalence of active power. Conversely, the proposed model uses multi-view CI, thus, its equivalent accuracies are ensured in more aspects.

From TABLE 12, we can see the standard deviation values of DVI, SI and DBI are greater than zero, which means the clustering algorithms has acquired different results. This is also because that no transfer learning is introduced in these clustering algorithms, and the clustering results are easy to be influenced by the random initialization. This again verifies the applying of transfer learning in MVT-FCM is quite helpful to increase the stability of WTs clustering results.

### F. THE ROBUSTNESS ANALYSIS

In this paper, the robustness refers to the equivalent accuracies of the proposed model under different LVRT scenarios. To evaluate the robustness of the proposed model, the representation abilities of dynamic responses under different scenarios are tested. There are 21 scenarios with different wind speeds, voltage dips and control modes for the WTs in the WF, and these scenarios are shown in Fig. 11. In Fig. 11, the x-axis denotes the WT number, and the y-axis denotes the scenario number. Take the 1st scenario as example, the voltage dip value in this scenario is 0.2pu; WT1-WT41 take constant power factor control, and the control modes for WT42-WT61 are constant voltage control; besides, the wind speeds for the WTs are the 1st row of this figure.

The $e_{\text{RMSE}}$, $e_{\text{MAE}}$ and $e_{\text{MAPE}}$ in different LVRT scenarios are shown in Fig. 12. Take the Fig. 12 (a) as example, the expectation and median $e_{\text{RMSE}}$ are $7.18 \times 10^{-2}$ and $7.11 \times 10^{-2}$, respectively. Sort the error sequence from small to large and take the median $e_{\text{RMSE}}$ as the center, 50% errors around this center are in the range of $[6.95 \times 10^{-2}, 7.41 \times 10^{-2}]$, and 82% errors around this center are in $[6.86 \times 10^{-2}, 7.52 \times 10^{-2}]$. The minimum error is $6.81 \times 10^{-2}$, where the maximum error is $7.62 \times 10^{-2}$. Using the same method to analyze these errors, we can find the errors are quite small and the fluctuation of the errors is less than 6% of their expectation. In other words, the errors are quite stable when LVRT scenarios changes. Thus, the robustness against various LVRT circumstances of the proposed model is quite high. The model is suitable for dynamic response analysis of WF, and has high practical value.

### VI. CONCLUSION

A MVT-FCM and SSAE based WF equivalent method is proposed in this paper. Based on the simulation study of the actual system, several conclusions can be drawn:

1. Using the SSAE-based extracted features as the multi-view CI and the proposed MVT-FCM for WTs clustering, the equivalent errors of the proposed model decrease at least 3% when comparing with other state-of-the-art models.
Based on MVT-FCM, the knowledge of source domain is transferred to guide the clustering process of target domain samples. The introducing of transfer learning is quite helpful to increase the clustering stability of MVT-FCM. The tests show that the clustering results would be only and would not be influenced by initialization.

The equivalent error fluctuations are within 6% under different LVRT scenarios. Thus, the proposed WF equivalent model is applicable under different LVRT circumstances, and the robustness of the proposed model performs excellent in the dynamic response analysis of WF.

REFERENCES
[1] The 13th Five-Year Plan for Power Development, National Development and Reform Commission, Beijing, China, 2016, pp. 10–13.
[2] Y. Zhou, L. Zhao, and W.-J. Lee, “Robustness analysis of dynamic equivalent model of DFIG wind farm for stability study,” IEEE Trans. Ind. Appl., vol. 54, no. 6, pp. 5682–5690, Nov. 2018.
[3] L. M. Fernández, C. A. García, J. R. Saenz, and F. Jurado, “Equivalent models of wind farms by using aggregated wind turbines and equivalent winds,” Energy Convers. Manage., vol. 50, no. 3, pp. 691–704, Mar. 2009.
[4] R. Fang, R. Shang, M. Wu, C. Peng, and X. Guo, “Application of gray relational analysis to k-means clustering for dynamic equivalent modeling of wind farm,” Int. J. Hydrogen Energy, vol. 42, no. 31, pp. 20154–20163, Aug. 2017.
[5] J. Zou, C. Peng, H. Xu, and Y. Yan, “A fuzzy clustering algorithm-based dynamic equivalent modeling method for wind farm with DFIG,” IEEE Trans. Energy Convers., vol. 30, no. 4, pp. 1329–1337, Dec. 2015.
[6] P. Chao, W. Li, X. Jin, J. Qi, and X. Li, “An active power response based practical equivalent method for DFIG wind farms,” Proc. CSEE, vol. 38, no. 6, pp. 1639–1646, Mar. 2018.
[7] Z. Xing, L. Longyuan, H. Xiaobo, X. Wan, and X. Zhou, “Wind farm dynamic equivalence based on clustering by output time series data of wind turbine generators,” Power Syst. Technol., vol. 39, no. 10, pp. 2787–2793, Oct. 2015.
[8] L. Lin, X. Pan, L. Zhang, and S. Zhao, “The K-means clustering algorithm for wind farm based on immune-outlier data and immune-sensitive initial center,” Proc. CSEE, vol. 36, no. 20, pp. 5461–5468, Oct. 2016.
[9] P. Wang, Z. Zhang, Q. Huang, N. Wang, X. Zhang, and W.-J. Lee, “Improved wind farm aggregated modeling method for large-scale power system stability studies,” IEEE Trans. Power Syst., vol. 33, no. 6, pp. 6332–6342, Nov. 2018.
[10] W. Li, P. Chao, X. Liang, J. Ma, D. Xu, and X. Jin, “A practical equivalence method for DFIG wind farms,” IEEE Trans. Sustain. Energy, vol. 9, no. 2, pp. 610–620, Apr. 2018.
[11] Z. Deng, Y. Jiang, F.-L. Chung, H. Ishibuchi, K.-S. Choi, and S. Wang, “Transfer prototype-based fuzzy clustering,” IEEE Trans. Fuzzy Syst., vol. 24, no. 5, pp. 1210–1232, Oct. 2016.
[12] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, “Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images,” IEEE Trans. Med. Imag., vol. 35, no. 1, pp. 119–130, Jan. 2016.
[13] S. Jialin Pan and Q. Yang, “A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[14] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, “Collaborative fuzzy clustering from multiple weighted views,” IEEE Trans. Cybern., vol. 45, no. 4, pp. 688–701, Apr. 2015.
[15] L. Hongkun, H. Baitian, D. Yuebang, and Y. Rui, “Wear status recognition for milling cutter based on compressed sensing and noise stacking sparse auto-encoder,” J. Mech. Eng., vol. 55, no. 14, p. 1, Jul. 2019.
[16] Y. Xu and D. Liu, “Equivalence of wind farms with DFIG based on two-step clustering method,” Power Syst. Protection Control, vol. 45, no. 6, pp. 108–114, Mar. 2017.
[17] D. M. S. Arsa, M. A. Ma’sum, M. F. Rachmadi, and W. Jatmiko, “Optimization of stacked unsupervised extreme learning machine to improve classifier performance,” in Proc. Int. Workshop Big Data Inf. Secur. (IWBIS), Sep. 2017, pp. 63–68.
WEICHEN YANG was born in Henan, China, in 1995. He received the B.S. degree in electrical engineering from North China Electric Power University, Baoding, China, in 2017. He is currently pursuing the Ph.D. degree with the Huazhong University of Science and Technology (HUST).

His research interests include high-voltage direct current (HVDC) technology and renewable energy technology.

HAORAN YIN was born in Hebei, China, in 1997. He received the B.S. degree in electrical engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2019, where he is currently pursuing the M.S. degree.

His research interests include application of machine learning in power systems and fault diagnosis of power systems.

* * *