Anti-Jitter and Refined Power System Transient Stability Assessment Based on Long-Short Term Memory Network

BAOQIN LI, JUNYONG WU, LIANGLIANG HAO, MEIYANG SHAO, RUOYU ZHANG, AND WEI ZHAO

1School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China
2State Grid Beijing Tongzhou Electrical Power Supply Company, Beijing 100044, China
Corresponding author: Junyong Wu (wujy@bjtu.edu.cn)

ABSTRACT In order to maintain the stable operation of power systems, quick and accurate transient stability assessment (TSA) after the fault clearance is important. Machine learning methods have been widely used in the transient stability analysis of power systems. However, how to make good use of time series data of PMUs and effectively balance the contradiction between rapidity and accuracy brings new challenges to TSA. To address this problem, we propose an anti-jitter dynamic evaluation method based on long-short term memory (LSTM) network. In this model, the trajectory cluster characteristics of generators power angles after fault clearance are taken as inputs, and an improved LSTM is used to learn the nonlinear mapping relationship between the input characteristics and the transient stability. Meanwhile, by the use of sliding time windows and anti-jitter mechanism, a hierarchical real-time prediction framework is constructed to effectively utilize the time series data of PMUs. The case studies on two systems indicate that the proposed method has superior evaluation accuracy and general performance. In addition, the proposed method can effectively evaluate the stability margin or instability degree of samples, which provides reliable reference information for emergency control.

INDEX TERMS Deep learning, long-short term memory, recurrent neural network, transient stability assessment, artificial intelligence.

I. INTRODUCTION

Transient stability is the capability that synchronous generators transfer to a new steady state or restore to the original steady state when power system is subjected to large disturbance [1]. As modern power grids develop rapidly, the increase in the proportion of new energy access and power electronization of power systems make the dynamic characteristics of power systems more complicated [2]. Therefore, real-time and accurate transient stability assessment of post-disturbance is crucial.

The time domain simulation (TDS) [3] is regarded as the most accurate method of transient stability analysis, and is often used as a standard for verifying other transient stability analysis methods. The method calculates the change of the rotor angle of the generators with time by stepwise integrating the differential-algebraic equations of the system during the fault and after the fault clearance. However, with a heavy computational burden, it becomes more and more difficult for TDS to make the rapid online prediction for large-scale power systems [4]. The direct method [5]–[7] can quickly judge the transient stability of the system, but for large-scale AC-DC hybrid power systems, it is difficult to determine the energy function or to maintain its adaptability in various conditions.

Compared with the traditional methods of transient stability analysis, the machine learning methods do not need the establishment of complex mathematical model. From the perspective of pattern recognition, samples are obtained based on historical running data or TDS, and then offline training is used to obtain the mapping relationship between input features and the output ones. When applied online,
the model can quickly predict the stability of power system based on measured data. A lot of studies have been done on the application of machine learning algorithms in power systems, but most algorithms are limited to shallow learning, such as support vector machine (SVM) [8]–[10], decision tree (DT) [11], [12], random forest (RF) [13], and K nearest neighbor (KNN) [14] etc. Due to their limited ability in data mining, the generalization ability is limited when dealing with complex problems. Accompanied by the rapid development and superior performance in feature extraction, deep learning provides new ideas for power system transient stability assessment.

In recent years, theory, algorithms, and applications related to deep learning have been developed rapidly. Some studies have used deep learning to predict the transient stability of power systems [15]–[18]. In [15], the convolutional neural network integrated with the active learning and fine-tuning techniques is used to improve the performance of the model and reduce the computation time in the scenario of a few unlabeled samples. In [16], based on the data mining and feature extraction capability of deep architecture, the deep belief network and transient stability assessment are combined. Literature [17] uses multi-layer perceptron to evaluate transient stability margins. In [18], the data fusion technique is used to assess transient stability based on ensemble cost-sensitive stacked denoising autoencoder (SDAE).

However, there are still some research gaps in machine learning-based TSA techniques. Firstly, the existing research [15]–[18] tends to use a fixed response time for prediction, which means that it must wait for a period of time before evaluating the stable state of the system. But it is always preferred to have a fast TSA response so that more time can be reserved for remedial control. Secondly, how to reliably determine whether those samples near the stable boundary are stable or not. Most existing studies define the concept of credibility [19], [20], but credibility is only an artificially defined probability value. Thirdly, the existing literatures only regard TSA as a binary problem (stable or not), but it is more important to predict the instability mode of the system.

In order to overcome the aforementioned drawbacks to give reliable output results, in this paper, we propose an anti-jitter refined transient stability assessment of power system model based on long-short term memory network. The main contributions of this paper are as follows:

1) In order to use the spatial and temporal data of power systems effectively, a hierarchical real-time framework for TSA based on LSTM is proposed. This framework is verified to be high prediction accuracy, and can balance the contradiction between rapidity and accuracy.

2) A novel anti-jitter mechanism is presented, through the time windows sliding, most of samples can be identified in the primary layer, and the reliability of the model is greatly improved.

3) In order to provide more refined evaluation results, the regression model is employed to predict the stability margin and instability degree, both of which are of great significance to dispatchers.

The remainder of this paper is organized as follows: Section II briefly introduces the LSTM network employed in this paper. Section III describes the transient stability assessment problem. Section IV proposes an anti-jitter LSTM model for TSA. Section V includes comprehensive case studies and discussions. Finally, we conclude this work in Section VI.

II. THE OVERVIEW OF LONG-SHORT TERM MEMORY

In this section, we will provide a brief introduction to the recurrent neural network (RNN) and LSTM. Such networks are critical to the proposed TSA method.

A. RECURRENT NEURAL NETWORK

RNN is a multi-layer perceptron network used to process time series data. Different from common artificial neural network (ANN), it has additional recurrent connections between the hidden layers, which provide essential memory capabilities. Such connections facilitate the network to keep previous information for later use, and thus capture the interdependencies among the input data.

Fig. 1 shows the structure of RNN. In RNN, each layer of neurons shares the common parameters, but only the input is different. This training method greatly reduces the total amount of parameters and improves the training efficiency of the model [21]. The output is generated by the following formula:

\[ o_t = f(Wx_t + Uh_{t-1} + b) \]  (1)

where \( f(\cdot) \) represents the activation function. The commonly used activation functions include sigmoid, tanh, softmax and so on. The nomenclature is given in Table 1.

B. LONG-SHORT TERM MEMORY NETWORK

When the time interval increases, RNN will suffer from the "vanishing gradient problem" that seriously affects the ability of the model to process temporal data. The problem was not solved until Hochreiter proposed the LSTM [22], which is an improved algorithm based on RNN. The special structure enables LSTM to get rid of long-term dependence
TABLE 1. Nomenclature.

| Notation | Description | Notation | Description |
|----------|-------------|----------|-------------|
| $X$      | Input matrix | $m$      | Number of consecutive sliding time windows' |
| $Y$      | Output matrix| $T$      | Loss weights of stable sample |
| $h$      | Hidden layer matrix | $W_s$ | Loss weights of unstable samples |
| $W$      | Connected to the weights | $W_{ss}$ | Rotor angle of the $i_a$ generator |
| $U$      | Connected to the weights | $\delta_i$ | The moment system gets unstable |
| $b$      | The bias | $t_{cf}$ | Fault clearance time |
| $c$      | Number of categories | $t_{cl}$ | |
| $p$      | Number of labeled samples | $M_s$ | Stability margin |
| $i$      | The $i_{th}$ sample | $M_{as}$ | Instability degree |
| $j$      | The $j_{th}$ category | $MSE$ | The mean square error |
| $\lambda$ | Length of time windows | $N$ | Number of samples |

FIGURE 2. The basic structure diagram of LSTM.

and achieve better learning effectiveness [23]. The structure of LSTM is shown in Fig. 2.

1) Forget gate: Choose to forget some information in the past.

\[ f(t) = \sigma(W_f x_t + U_f h_{t-1} + b_f) \] (2)

2) Input gate: Remember the current information.

\[ i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \] (3)

\[ \tilde{C}_t = \tanh(W_C x_t + U_C h_{t-1} + b_C) \] (4)

\[ C_t = f_t \ast C_{t-1} + i_t \ast \tilde{C}_t \] (5)

3) Output gate: Output the final result.

\[ O_t = \sigma(W_O x_t + U_O h_{t-1} + b_O) \] (6)

\[ h_t = O_t \ast \tanh(C_t) \] (7)

where * is the element-wise product, $x_{t-1}$ and $x_t$ represent inputs of the previous time and the current time respectively. This paper adopts the sigmoid activation function $\sigma(x) = \frac{1}{1 + e^{-x}}$. Due to its simplicity and effectiveness, many LSTM-based applications have been developed. Interested readers can refer to [24] for a comprehensive introduction of this method.

III. TRANSIENT STABILITY ASSESSMENT

A. TRANSIENT STABILITY PREDICTION AND EVALUATION

The essence of transient stability is to find the stable and unstable boundary of the system. From the perspective of function mapping, the key lies in establishing the mapping relationship between the input feature ($X$) of the power system operating state and the evaluation result ($Y$) of the system stability. The transient stability prediction of power systems based on LSTM is used to simulate the process of finding stable boundaries. The input data $X$ can be stored as a high-dimensional array, as shown in Fig. 3.

$Y$, a set of labels corresponding to the input data set $X$, is expressed as equation (8) where $c$ is the number of categories. In this paper, the anti-jitter LSTM prediction model divides the transient stability into two categories: stable and unstable. At this time, $c = 2$.

\[
Y = \begin{bmatrix}
y_1^1 & y_1^2 & \cdots & y_1^c \\
y_2^1 & y_2^2 & \cdots & y_2^c \\
\vdots & \vdots & \ddots & \vdots \\
y_p^1 & y_p^2 & \cdots & y_p^c \\
\end{bmatrix}
\] (8)

Each label corresponds to a vector in space $R^c$, and each sample corresponds to a label. The specific labeling rules are shown in equation (9).

\[
y_j^i = \begin{cases} 
1, & x_i \in j_{th} \text{ category} \\
0, & x_i \notin j_{th} \text{ category}
\end{cases}
\] (9)

B. INPUT FEATURE SELECTION

The dynamic response of the generators power angle under different transient states is shown in Fig. 4. It can be seen that the generators power angle has the ability to directly reflect
During the simulation period. When TSI under the standard load level, and the spatial distribution obtained by the New England 10-machine 39-node system will be obtained [28]. As shown in Fig. 5, it is the data set be formed, and different transient stability prediction models different response time are selected, different data sets will clearly introduced in section III.B. When the features at clearance is introduced in section III.B. When the features at fault clearance, and with reference to other researchers’ experience in feature selection [25]–[27], 27 trajectory cluster features are constructed in this paper. The detailed description and calculation formulae are shown in Table 8 of Appendix.

IV. ANTI-JITTER REFINED TRANSIENT STABILITY PREDICTION MODEL

In this section, we will introduce the anti-jitter and refined evaluation model used in this paper.

A. SLIDING TIME WINDOWS

The “trajectory cluster” of generators power angle after fault clearance is introduced in section III.B. When the features at different response time are selected, different data sets will be formed, and different transient stability prediction models will be obtained [28]. As shown in Fig. 5, it is the data set obtained by the New England 10-machine 39-node system under the standard load level, and the spatial distribution of the characteristics \(f_8\) and \(f_9\) is corresponding to different response time.

\[
\text{TSI} = \frac{360^\circ - |\Delta\delta_{\text{max}}|}{360^\circ + |\Delta\delta_{\text{max}}|}
\]

where \(\Delta\delta_{\text{max}}\) refers to the maximum power angle difference during the simulation period. When \(\text{TSI} > 0\), the sample is stable; otherwise, the sample is unstable.

Therefore, this paper takes the generators power angle as the research object and introduces the concept of “trajectory cluster”. Instead of studying the change of the power angle of a generator, we regard the power angle of all the generators after fault clearance as a whole cluster, and study the overall change law. When the scale of the system increases, the generators’ information increases and the number of input data increases manyfold, which limits the quick and efficient online application. If some of the generators’ information is missing in the actual application, the initial data set will be greatly affected and interfered. The above two reasons fully explain the necessity of introducing the concept of “trajectory cluster”.

Based on the power angle information of generators after fault clearance, and with reference to other researchers’ experience in feature selection [25]–[27], 27 trajectory cluster features are constructed in this paper. The detailed description and calculation formulae are shown in Table 8 of Appendix.

From Fig. 5(a), it can be seen that there are many overlapping regions of the feature space at this time, and the separability of samples is poor. On the contrast, Fig. 5(b) is the feature space formed by the 60th cycle \(f_8\) and \(f_9\) after fault clearance. The comparison shows that using a longer response time, the separability of samples is better, and the stable samples and the unstable samples have a greater distinction.

In order to balance the speed and accuracy while effectively utilize the time series data of power systems, this paper proposes an anti-jitter and refined prediction model based on the “sliding time windows”. The length of time windows is set as \(\Delta t\). According to the definition of the trajectory cluster features, the calculation of the “acceleration” related features must include at least three sampling points in the original trajectory, so the time window should have at least three cycles. Set the length of time windows as \(\lambda\), that is \(\Delta t = \lambda\). The principle of minimum information and calculation efficiency should be taken into consideration when taking specific values of \(\lambda\). If the sliding time windows’ sequence label is set to \(T\), when \(T = 1\), the input time series set of the prediction model is \(S_1\{t = 1, t = 2, \ldots, t = \lambda\}\); when \(T = 2\), \(S_2\{t = 2, t = 3, \ldots, t = \lambda + 1\}\); when \(T = 3\), \(S_3\{t = 3, t = 4, \ldots, t = \lambda + 2\}\). When the input time series set of the prediction model is \(\{S_1, S_2, S_3\}\), the total time length is \(\lambda + 2\) cycles.

B. HIERARCHICAL REAL-TIME AND ANTI-JITTER PREDICTION MODEL

Existing researches usually select a characteristic data set with fixed response time to construct a transient stability prediction model. This paper proposes a hierarchical real-time and anti-jitter prediction model. The anti-jitter model is shown in Fig. 6.

In Fig. 6, the LSTM is trained separately with data from different windows. The length of the window is \(\lambda\), and the step size of sliding is 1 (each window contains data of \(\lambda\) cycles, one cycle per slide). In transient stability prediction, the samples that are near the stable boundary are most difficult to identify. For those samples, the characteristics of stability or instability are not obvious, which leads to different prediction results of the consecutive sliding windows. For example, when using \(S_1\), the sample is predicted as stable,
FIGURE 6. The anti-jitter model for transient stability prediction.

FIGURE 7. Hierarchical real-time prediction framework based on LSTM.

but when sliding to $S_2$, the result is unstable and vice versa. So we define this as a jitter phenomenon. In order to reliably discriminate the true label of the sample and identify those samples at the stable boundary quickly and accurately. When prediction results of $m$ consecutive windows are consistent, the final category (stable/unstable) is the output. Otherwise, the window will continue sliding forward until the final prediction results are consistent.

As shown in Fig. 7, different LSTM models that form different levels of hierarchical real-time prediction framework can be obtained by using data from different time windows. Once the prediction results are same for $m$ times in succession, the model will output the prediction result stable or not of the system, and then will evaluate the stability margin and instability degree of no-jitter samples. For the sample with jitter, the classifier temporarily classifies it as an uncertain sample, and makes no judgment until the next layer. With such a hierarchical real-time prediction method, the prediction result can be given in an earlier time after the fault clearance, thereby ensuring the rapidity; further prediction of the stability margin and the instability degree ensures the fineness of the result; the jitter sample is temporarily classified as an uncertain sample, and accurate judgment is performed after more information is obtained at a later layer, thereby ensuring the accuracy of the result.

C. EVALUATION PROCESS

In Fig. 8, the evaluation process for transient stability prediction is shown. The specific steps of the anti-jitter LSTM offline training are as follows:

1) The data set is obtained by TDS including different load levels, different network structures and different fault sets, and then the “trajectory cluster” features of generators power angle after fault clearance are calculated.

2) The data set is randomly divided into a training set and a test set according to the set ratio. The training set of different time windows is used to train the LSTM models with different response time to form different levels of TSA. The test set is used to test the performance of the hierarchical real-time and anti-jitter model. In addition, in order to solve the imbalance problem of the number of unstable samples and the number of stable samples, the loss function is improved in this paper, and the weighted cross entropy loss function is used [29]. The specific weighting process is shown in equation (11). The Adam optimization algorithm is used to fine tune the weight.
and offset of the LSTM from top to bottom until the number of iterations reaches the set value.

\[ L = - \frac{1}{N} \sum_{i=1}^{N} \left[ W_{sy} \log(\hat{y}_i^1) + W_{usy} \log(\hat{y}_i^0) \right] \] (11)

where \((y_i^1, y_i^0)\) is the actual label of the sample, for example the label of stable sample is \((1, 0)\), the label of unstable sample is \((0, 1)\), \(\hat{y}_i^1 / \hat{y}_i^0\) is the probability that the ith sample predicted by the LSTM model as stable/unstable and \(\hat{y}_i^1 + \hat{y}_i^0 = 1\). Generally \(W_s = W_{us} = 1\), the importance of stable samples and unstable samples is not distinguished. This paper sets \(W_s < W_{us}\) and the specific value is related to the proportion of samples category.

3) In the online application, the trained LSTM model is invoked through parameter transfer. Once the data of PMUs is obtained, the hierarchical real-time and anti-jitter model will be carried out.

**D. TRANSIENT STABILITY ASSESSMENT**

1) EVALUATION INDICES

In practical applications, the cost of misidentification and false alarm is very different. Additionally, the number of unstable samples used for model training is much less than the number of stable samples. Therefore, the transient stability evaluation is a typical unbalanced classification problem. This paper evaluates the classification effect of the algorithm by constructing the confusion matrix [30]. The specific representation of the confusion matrix is shown in Table 2.

| Confusion matrix | Stable(Predicted) | Unstable(Predicted) |
|------------------|-------------------|---------------------|
| Stable(Actual)   | \(T_s\)           | \(F_{us}\)          |
| Unstable(Actual) | \(F_s\)           | \(T_{us}\)          |

Compared with ACC, the \(G_{\text{mean}}\) value can objectively reflect the prediction performance of the model for unstable samples. Suppose that in the test set, when the ratio of the number of stable samples to the number of unstable samples...
is 3:1 and when the model judges all samples as “stable”, ACC is 75% but G_mean is 0. Therefore, ACC and G_mean should be considered in the evaluation of the performance of the model at the same time.

2) STABILITY MARGIN AND INSTABILITY DEGREE
In order to further make a more refined assessment of power system transient stability, this paper calculates the stability margin and instability degree. Since the critical clearing time (CCT) requires time domain simulation to be repeatedly tested, the calculation takes a lot of time. Therefore, based on [33], this paper constructs a degree of disturbance based on the envelope integral of rotor angle locus cluster.

\[ S = \int_{0}^{t_s} (\max_{i} \delta_i - \min_{i} \delta_i) dt \]  

(16)

where \( t_s \) is the simulation time. The maximum and minimum normalization is taken as the stability margin index:

\[ M_s = \frac{S_{\text{max}} - S}{S_{\text{max}} - S_{\text{min}}} \]  

(17)

When the system is unstable, it is more important to predict the instability mode of the system. \( T \) refers to the time from fault clearance to instability of the system. It is used to classify the instability degree of the system:

\[ T = t_{\text{as}} - t_{cl} \]  

(18)

The instability degree \( M_{\text{as}} \) is defined as follows:

\[ M_{\text{as}} = \frac{T_{\text{max}} - T}{T_{\text{max}} - T_{\text{min}}} \]  

(19)

The mean square error (MSE) is defined as follows. It is used to measure the accuracy of the transient stability regression model.

\[ \text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i) \]  

(20)

where \( \hat{y}_i \) is the predicted value of the regression model, and \( y_i \) is the real stability margin and instability degree of the sample.

V. CASE STUDIES
In order to verify the effectiveness of the proposed method, analysis is implemented on the New England 10-machine 39-node system and Central China Power Grid. The LSTM network is built in the TensorFlow environment. The programming language is Python and the PC is configured as: Intel(R) Core(TM) i7-8700 CPU/8.00GB RAM.

A. DATA GENERATION OF IEEE 39-BUS SYSTEM
The IEEE 10-machine 39-node system has been widely used in power system transient stability analysis. The system consists of 10 generators, 39 busbars and 46 lines, representing a 345KV power network in New England, USA. Considering 10 different load levels of 75%, 80%, ..., 120%, the output of generators is adjusted accordingly to ensure the convergence of the power flow calculation. There are 46 lines in total, 12 lines contain transformers, so fault is set on the remaining 34 lines and it is located at 0%, 10%, ..., 90% of each line respectively. The fault type is set to the most severe three-phase short circuit of the power system, and the fault clearance time of the first cycle, the second cycle, ..., the eleventh cycle after fault is taken into account respectively. Power System Tool (PST) 3.0 [34] is used to generate 34 × 11 × 10 × 10 = 37400 samples, including 22,712 stable samples and 14,688 unstable samples. The samples are randomly divided into the training set and the test set at a 4:1 ratio. The sample collection is classified as shown in Table 3.

| Samples Collection | Amount | Stable | Unstable |
|--------------------|--------|-------|---------|
| Total Samples      | 37400  | 22712 | 14688   |
| Training Set       | 29920  | 18102 | 11818   |
| Testing Set        | 7480   | 4586  | 2894    |

B. PARAMETERS SELECTION AND MODEL ANALYSIS
1) PARAMETERS SELECTION
LSTM is a kind of deep neural networks. The selection of hyperparameters’ values plays a crucial role in the evaluation of model performance [23]. This paper uses the Adam algorithm to train the LSTM model. The learning rate is set to 0.001, the number of hidden layers is set to 4, and the maximum training iterations is 400. In addition, the number of hidden layer neurons and the batch size have a great influence on the LSTM model. Hyperparameters are selected by the use of grid search and 5-fold cross-validation. The search range set for each layer of neurons is \{200, 100, 50\}, \{100, 80, 50\}, \{50, 30, 15\}, \{30, 20, 15\} respectively, and the search size set of the batch size is \{400, 300, 200, 100, 50\}. The hyperparameter combination with the smallest loss is selected. Finally, the number of per layer neurons is 100, 50, 30, and 15 respectively, and the batch size is 100.

Secondly, the parameters that have a great influence on the performance of the proposed method the time windows’ length \( \lambda \) and the number \( m \) of consecutive output category labels. The grid search method is used to optimize the two parameters. Fig. 9 shows the contours of various evaluation indices. It can be seen that when the window length \( \lambda \) is fixed, as the number of consecutive outputs \( m \) increases, various indices show an upward trend; similarly, when the number \( m \) is fixed, as the window length \( \lambda \) increases, various indices increase. And it is observed that the monotonicity of the contour of TSR and TUR is opposite, because TSR and TUR are two mutually constrained indices. Considering the principle of minimum information and calculation efficiency, in this paper, we set \( \lambda = 6 \), \( m = 4 \). Each sliding time window contains 6 cycles of data, and 4 consecutive windows output the result of this prediction.

| TABLE 3. The classification of sample sets. |
|-------------------------------------------|
| Samples Collection | Amount | Stable | Unstable |
|-------------------|--------|-------|---------|
| Total Samples      | 37400  | 22712 | 14688   |
| Training Set       | 29920  | 18102 | 11818   |
| Testing Set        | 7480   | 4586  | 2894    |
TABLE 4. The results of hierarchical real-time transient stability prediction ($\lambda = 6$, $M = 4$).

| Sliding Windows | State (Actual) | Predict | Evaluation Indices | Uncertain Samples | Uncertain Rate |
|-----------------|----------------|---------|-------------------|------------------|---------------|
| $S_1S_2S_3S_4$ | Stable, Unstable | 4470, 12 | ACC 99.73% G_mean 99.72% TSR 99.73% TUR 99.72% | 170 | 2.27% |
| $S_5S_6S_7S_8$ | Stable, Unstable | 34, 10 | ACC 97.78% G_mean 97.14% TSR 97.14% TUR 100% | 125 | 1.67% |

FIGURE 9. Lines of various evaluation indices. (a) Contour of ACC. (b) Contour of G_mean. (c) Contour of TSR. (d) Contour of TUR.

2) ANTI-JITTER MECHANISM EFFECTIVENESS

The evaluation performance of the model with and without anti-jitter mechanism is compared as shown in Fig. 10. It can be seen that when the anti-jitter mechanism is introduced, the evaluation performance is better than that without anti-jitter mechanism, and the ACC of anti-jitter LSTM is 1.23% higher than that of LSTM without anti-jitter mechanism, G_mean is 1.16% higher, TSR is 0.91% higher and TUR is 1.42% higher. The reason why anti-jitter mechanism can improve the assessing performance is that the model will not output its prediction results immediately. As time goes on, when the stable or unstable characteristics are more obvious, the label of whether the sample is stable or not is output. By introducing the anti-jitter mechanism, the reliability of the model is greatly improved, and the contradiction between the evaluation speed and the prediction accuracy can be effectively balanced. As a result, a better evaluation result is obtained.

FIGURE 10. Comparison between anti-jitter mechanism model and without anti-jitter mechanism model.

The results of hierarchical real-time transient stability prediction are shown in Table 4. In the first layer prediction, 7310 non-jitter samples output the prediction results accounting for 97.73% of all the test samples, which means most samples in the first layer can be judged quickly and accurately. ACC, G_mean, TSR, TUR indices have risen over 99.5%. It can be seen that as time goes by and the sliding of window, more and more samples are gradually recognized, and the number of uncertain samples is gradually decreased. In the prediction of the fourth layer, all the samples are recognized, and the window slides to $S_7$ which is the 12th cycle after the fault clearance ($t = 0.2$ s). The result reflects the rapidity of the proposed method.

C. PERFORMANCE OF THE ANTI_JITTER LSTM

1) ANTI_JITTER LSTM MODEL VS OTHER KINDS OF CLASSIFIERS

In view of the current methods used in power system transient stability prediction, SVM, DT, RF, and KNN in shallow learning, and LSTM, artificial neural network (ANN), and multilayer perceptron (MLP) in deep learning are compared to show that the proposed method is not only superior to conventional shallow learning, but also better than partial deep learning models. For the above classifiers, the optimal parameters are selected respectively. The structure of the ANN is the same as that of LSTM, the number of layers is set to 6, and the number of neurons in each layer is [162,100,50,30,15,2] respectively. The structure of MLP is...
selected as [162,500,500,2]. SVM adopts radial basis function as kernel function, grid search method and five-fold cross-validation are used to optimize related parameters. DT adopts C5.0 algorithm. RF adopts 100 trees to ensemble. After the parameter of KNN is optimized, K is set as 30. The test models are built on the open source machine learning library scikit-learn. Taking data set of $S_1 - S_4$ windows, the above methods are used for test. The test results are shown in Table 5.

LSTM and ANN, MLP are deep learning methods, and the latter four are shallow learning methods. The proposed method shows superior prediction performance. The ACC of anti-jitter LSTM model is 4.22% higher than that of ANN, and the G_mean is 4.32% higher. The structures of the ANN and LSTM settings are identical, but the evaluation indices are different, which fully demonstrates that the LSTM model effectively utilizes the time series data to achieve a better evaluation result.

The ACC of LSTM is 3.02%, 2.39%, 1.48%, and 6% higher than SVM, DT, RF, and KNN respectively, and the G_mean is 3.26%, 2.52%, 1.66%, and 7.46% higher respectively. This shows that the deep structure can better mine the nonlinear mapping between the input data and the transient stability prediction results, so as to obtain better evaluation results than the shallow learning.

2) PERFORMANCE OF MODELS WITH NOISE
In practical online applications, the test sets are all from real-time PMUs data. PMUs devices have certain variation factors for the measurement characteristics of dynamic data, therefore errors may be involved. In this paper, the measurement errors in practical application are simulated by adding Gaussian white noise whose instantaneous value is Gaussian distribution and power spectral density is uniform distribution to the test set. The noise structure [15] is as follows:

$$test_x' = test_x \times (1 + \theta), \theta \sim N(0, \sigma)$$  \hspace{1cm} \text{(21)}$$

$test_x$ is the test set without noise, and $test_x'$ is the test set with Gaussian white noise. $\theta$ obeys the Gaussian distribution with mean value of 0 and variance of $\sigma$.

The trained model is tested with a new test set $test_x'$. The value range of $\sigma$ is [0.01, 0.09]. The prediction results of different classifiers are shown in Fig. 11. It can be seen from Fig. 11 that the proposed anti-jitter LSTM model has higher ACC than ANN and MLP when different noises are added. In the shallow learning models, as the noise intensity increases, the ACC and G_mean of KNN, DT, and RF fluctuate greatly. Especially the prediction performance of DT and RF declines sharply, which shows the poor robustness of DT and RF. The prediction performance of SVM is not seriously affected by noise, but its ACC and G_mean are lower than those of the anti-jitter LSTM model. When the noise $\sigma$ is 0.09, the ACC of SVM is 4.49% lower than that of LSTM, and the G_mean is 5.22% lower. Therefore, the proposed method still has good robustness in the presence of noise.

3) MODEL EVALUATION PERFORMANCE WHEN PMU IS INCOMPLETE MEASUREMENTS
Due to the high cost of PMUs, in the actual large power grid, the coverage rate of PMUs cannot reach 100%. Sometimes only some important generators and hub substations are equipped with PMUs, which brings great challenges to online application. In this paper, according to the two PMU configuration schemes in reference [35], the power angle information of the nodes that the PMU cannot detect is

| Models       | ACC    | TSR    | TUR    | G-mean |
|--------------|--------|--------|--------|--------|
| Proposed method | 99.73% | 99.73% | 99.72% | 99.72% |
| ANN          | 95.51% | 95.88% | 94.91% | 95.40% |
| MLP          | 95.48% | 96.02% | 94.66% | 95.34% |
| SVM          | 96.71% | 97.56% | 95.36% | 96.46% |
| DT           | 97.34% | 97.80% | 96.61% | 97.20% |
| RF           | 98.25% | 98.87% | 97.27% | 98.06% |
| KNN          | 93.73% | 98.13% | 86.75% | 92.26% |
The ACC and G\_mean are degraded compared with the case of the full nodes, but they are both above 99%. The experimental results show that when the PMU configuration follows the principle of complete observability of the system, the proposed method still has good performance, which further proves the effectiveness of trajectory cluster.

4) RAPIDITY ANALYSIS

In order to analyze the rapidity of the proposed method, the histogram of instability occurrence time distribution in the test set is shown in Fig. 12. Fig. 12 (a) shows the histogram of instability occurrence time distribution of all unstable samples in the test set, Fig. 12 (b) shows the histogram of instability occurrence time distribution of unstable samples detected by anti-jitter LSTM in the first layer prediction, Fig. 12 (c) shows the histogram of instability occurrence time distribution of unstable samples undetected by anti-jitter LSTM in the prediction of the first layer, and Fig. 12 (d) shows the histogram of instability occurrence time distribution of unstable samples in jitter samples after the first layer prediction.

The average instability occurrence time of all unstable samples in the test set is 1.242s, while the anti-jitter LSTM only needs 0.066ms to predict whether a sample is stable or not. The response time in this paper is 0.15s (4 windows, including 9 cycles, 60HZ), which shows that the average preemptive time for dispatcher to take emergency control measures is 1.092s before system is really unstable.

In addition, in the first layer of jitter samples, the instability occurrence time of all unstable samples is after 0.5s (30 cycles), but the hierarchical real-time anti-jitter model can judge the stability of all samples in 0.2s (12 cycles), which fully meets the requirements of time. Because of its fast prediction performance, the proposed method can leave enough time for the dispatcher to decide the next control strategy. It has great advantage in time, which is of great significance for online application.

D. STABILITY MARGIN/INSTABILITY DEGREE PREDICTION

When the system is stable / unstable, it is more important to predict the stability margin / instability degree of the system. In order to provide more refined evaluation results, this section will analyze the regression prediction model of transient stability. The prediction object is the stable samples and unstable samples without jitter. The stable samples and unstable samples in the training set are used to train the stability margin evaluation model and instability degree evaluation model respectively. The calculation method of stability margin and instability degree is shown in formula (16) - (19).

The stability margin prediction model is used to predict the stable samples detected by hierarchical real-time anti-jitter model. In order to facilitate comparison, the stability margin is arranged in an ascending order. The true value and predicted value are shown in Fig. 13 (a). It can be seen that the LSTM model can accurately fit the stability margin of the samples without jitter, providing a better reference value for transient stability assessment. The mean square error of the regression prediction of this model is 0.0006. The observation of chart reveals that for the samples with large stability margin, the fitting effect between the predicted value and the true value is better, while for the samples with small stability margin, the fitting error is slightly greater. The main reason is that these samples with low stability margin are closer to the boundary of transient stability prediction, and these samples in the training data set are relatively few, which leads to the prediction error of the model.

The instability degree prediction model of LSTM is used to make regression prediction for the unstable samples. Similarly, the M_\text{in} is arranged in an ascending order. The regression prediction results are shown in Fig. 13 (b). The MSE of regression prediction is 0.002.
E. LARGER POWER SYSTEM CASE

In order to verify the effectiveness of the proposed method for large power grid, the data of Central China power grid is used for the test. There are 690 generators, 8492 busbars, 4474 lines, 19 conventional DC lines and 6022 transformers in the whole network. PSASP developed by China Electric Power Research Institute is selected as the transient stability calculation program. Take 5% as the step, set 10 load levels in the range of 75% ∼ 120%, and adjust the generators output accordingly; randomly select 4 lines, take 10% as the step, and set 9 fault positions in the range of 10% ∼ 90%; the fault types include single-phase short circuit fault, two-phase short circuit fault and three-phase short circuit fault. A total of 14040 samples are generated by recording the power angle information of generators above 1000MVA. There are 10446 stable samples and 3594 unstable samples. The training set and test set are divided at the ratio of 2:1. Among them, 9360 samples that are recorded as the training set are used to learn the nonlinear mapping relationship, and the remaining 4680 samples are used as the testing set. The process of parameters optimization is the same as that of the 39-node system.

1) VISUAL ANALYSIS

The t-distributed logistic neighbor embedding (t-SNE) [36] is used to map the original “trajectory cluster” features of the test samples and the features extracted from each layer of LSTM to the two-dimensional plane. t-SNE is a non-linear dimensionality reduction method which does not change the relative position of sample points. It is the best dimensionality reduction method so far. The visualization results are shown in Fig. 14.

It can be seen from Fig. 14 (a) that in the original feature space of the trajectory cluster, stable samples and unstable samples are mixed together, so it is difficult to distinguish whether they are stable or not; while Fig. 14 (b) - (e) LSTM gradually distinguishes stable samples and unstable samples by extracting features layer by layer, which intuitively shows the contribution of each layer of LSTM to distinguishing stable and unstable samples. It embodies the essence of layer by layer optimization of the LSTM model.
2) TSA RESULTS OF ANTI-JITTER LSTM MODEL
In order to further demonstrate the effectiveness of anti-jitter mechanism, comparison of experiments with and without anti-jitter mechanism in Central China power grid is show in Fig. 15. The ACC of anti-jitter LSTM is 1.36% higher than that of LSTM without anti-jitter mechanism, G_mean is 1.29% higher, TSR is 1.58% higher and TUR is 1.29% higher. The experimental result shows that all of the indices are improved by introducing the anti-jitter mechanism.

The transient stability prediction results of the hierarchical real-time anti-jitter model mentioned above are shown in Table 7. All the evaluation indexes are kept above 99% and the number of unstable samples that are predicted as stable is 0. The ratio of uncertain samples in the second layer is less than 1%. When the window slides to S7, all jitter samples are fully identified, and the time is only 0.24s (the 12th cycle, 50Hz) after fault clearance. Because of the rapidity and accuracy of the prediction, the anti-jitter LSTM model is still applicable to large-scale power grid.

VI. CONCLUSION
In order to make good use of time series data of PMUs and effectively balance the contradiction between rapidity and accuracy, a hierarchical real-time anti-jitter framework for TSA based on LSTM is presented. Comprehensive case studies of the proposed framework on 10-machine 39-node benchmark power system and the Central China Power Grid both demonstrate superior performance of TSA. The main conclusions can be drawn as follows:

1) The proposed hierarchical real-time anti-jitter model can identify the stability of power system quickly, so it has significant application value in the online prediction. For the samples far away from the stable boundary, fast judgment can be realized in the primary layer. For the samples near the stable boundary, as the window slides, reliable evaluation results can be obtained in the next layers. The anti-jitter mechanism makes the model have higher accuracy and G_mean.

2) The transient stability characteristics can be represented more effectively and completely by LSTM, therefore the proposed method has higher assessment accuracy than other artificial intelligence tools. It still has high accuracy and G_mean under incomplete PMU measurements and noise environment, which has strong robustness and is more suitable for the actual system.

3) The proposed method can not only predict binary information about stable state of samples, but also predict stability margin or instability degree accurately for all samples, which is instructive for emergency control.

In practice, how to use transfer learning and incremental learning to quickly and adaptively update the trained model online when the operation mode or topology of the system changes greatly will be considered in future works. In addition, if a stable case is false alarmed, then there will be little impact on the power system. In contrast, an unstable case that is misidentified will result in a disaster when no measure has been taken to prevent the collapse. So how to avoid misidentification with the least false alarmed cost should also be considered in future works.

APPENDIX
A. BASIC CLUSTER FEATURES
The trajectory cluster is represented by \( \{x_{ij}\}_{m \times n} \), where \( m \) represents the number of trajectories and \( n \) represents the number of cycles of the sample.

Arithmetic mean: \( C_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij} \)

Degree of dispersion: \( d_j = \frac{1}{m} \left( \sum_{i=1}^{m} |x_{ij} - C_j|^2 \right) \)

Upper envelope: \( \{\max(x_{ij})\}_{j = 1, 2, \cdots, n} \)

Lower envelope: \( \{\min(x_{ij})\}_{j = 1, 2, \cdots, n} \)

Mid-range: \( m_{ej} = \frac{1}{2}(\max(x_{ij}) + \min(x_{ij}))_{j = 1, 2, \cdots, n} \)

B. GRADIENT AND CURVATURE FEATURES
Gradient: \( r_j = \frac{1}{h} [x_{j+1} - x_j], j = 1, 2, \cdots, n - 1 \)
where \( h \) is the time sampling interval.

Curvature: \( C_j = \frac{1}{h^2} \left( \sum_{i=1}^{m} x_{i,j-1} - 2x_{ij} + x_{i,j+1} \right)^2 \cdot j = 2, 3, \cdots, n - 1 \)

C. ACCELERATION FEATURES
\( a_{ij} = \frac{1}{h} [r_{c, j+1} - r_{c, j}], j = 1, 2, \cdots, n - 2 \)
where \( r_c \) is the gradient of the trajectory cluster. See Table 8.

### TABLE 7. TSA results of a larger power system.

| Sliding Windows | State (Actual) | Predict | Evaluation Indices | Uncertain Samples | Uncertain Rate |
|-----------------|---------------|---------|---------------------|------------------|---------------|
| S1,S2,S3,S4     | Stable        | 3452    | 6                   | ACC = 99.87%     | TSR = 99.91%  |
|                 | Unstable      | 0       | 1170                | G_mean = 99.83%  | TUR = 100%    |
|                 |               |         |                     |                  | 52            | 1.11%         |
| S1,S2,S3        | Stable        | 4       | 0                   | 100%             | 100%          |
|                 | Unstable      | 0       | 4                   | 100%             | 100%          |
|                 |               |         |                     |                  | 44            | 0.94%         |
| S1,S2,S3,S4     | Stable        | 30      | 0                   | 100%             | 100%          |
|                 | Unstable      | 0       | 11                  | 100%             | 100%          |
|                 |               |         |                     |                  | 3             | 0.06%         |
| S2,S3,S4        | Stable        | 2       | 0                   | 100%             | 100%          |
|                 | Unstable      | 0       | 1                   | 100%             | 100%          |
|                 |               |         |                     |                  | 0             | 0%            |
TABLE 8. Features definition of trajectory cluster.

| No. | Description                      | No. | Description                        | No. | Description                       |
|-----|----------------------------------|-----|------------------------------------|-----|-----------------------------------|
| F1  | Arithmetic mean                  | F11 | Gradient of dispersion             | F22 | Arithmetic mean variation         |
| F2  | Degree of dispersion             | F12 | Gradient of the upper envelope     | F23 | Acceleration                      |
| F3  | Upper envelope                   | F13 | Gradient of the lower envelope     | F24 | Acceleration                      |
| F4  | Lower envelope                   | F14 | Gradient of the arithmetic mean    | F25 | Acceleration                      |
| F5  | Mid-range                        | F15 | Gradient of the envelope height    | F26 | Acceleration                      |
| F6  | Difference between the upper envelope and arithmetic mean | F16 | Trajectory curvature               | F27 | Envelope height variation         |
| F7  | Difference between the lower envelope and arithmetic mean | F17 | Curve of the arithmetic mean       |     |                                   |
| F8  | Envelope height                  | F18 | Curve of the upper envelope        |     |                                   |
| F9  | Difference between the arithmetic mean and mid-range | F19 | Curve of the lower envelope        |     |                                   |

REFERENCES

[1] X. He, L. Chu, R. C. Qu, Q. Ai, and Z. Ling, “A novel data-driven situation awareness approach for future Grids—Using large random matrices for big data modeling,” IEEE Access, vol. 6, pp. 13855–13865, 2018.

[2] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, “Energy big data: A survey,” IEEE Access, vol. 4, pp. 3844–3861, 2016.

[3] S. Zadkhast, J. Jatskevich, and E. Vaahedi, “A multi-decomposition approach for accelerated time-domain simulation of transient stability problems,” IEEE Trans. Power Syst., vol. 30, no. 5, pp. 2301–2311, Sep. 2015.

[4] S. Zhang, Y. Wang, M. Liu, and Z. Bao, “Data-based line trip fault prediction in power systems using LSTM networks and SVM,” IEEE Access, vol. 6, pp. 7675–7686, 2018.

[5] H.-D. Chang, C.-C. Chu, and G. Cauley, “Direct stability analysis of electric power systems using energy functions: Theory, applications, and perspective,” Proc. IEEE, vol. 83, no. 11, pp. 1497–1529, Nov. 1995.

[6] Y. Xue, T. Van Cutsem, and M. Ribbens-Pavella, “A simple direct method for fast transient stability assessment of large power systems,” IEEE Trans. Power Syst., vol. 3, no. 2, pp. 400–412, May 1988.

[7] I. A. Hiskens and D. J. Hill, “Energy functions, transient stability and voltage behaviour in power systems with nonlinear loads,” IEEE Trans. Power Syst., vol. 4, no. 4, pp. 1525–1533, Nov. 1989.

[8] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, “Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements,” IEEE Trans. Power Syst., vol. 26, no. 3, pp. 1474–1483, Aug. 2011.

[9] D. You, K. Wang, L. Ye, J. Wu, and R. Huang, “Transient stability assessment of power system using support vector machine with generator combinatorial trajectories inputs,” Int. J. Electr. Power Energy Syst., vol. 44, no. 1, pp. 318–325, Jan. 2013.

[10] Y. Zhou, J. Wu, L. Ji, Z. Yu, K. Lin, and L. Hao, “Transient stability preventive control of power systems using chaotic particle swarm optimization combined with two-stage support vector machine,” Electric Power Syst. Res., vol. 155, pp. 111–120, Feb. 2018.

[11] T. Guo and J. V. Milanovic, “Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability,” IEEE Trans. Power Syst., vol. 29, no. 1, pp. 377–385, Jan. 2014.

[12] M. R. Aghamohammadi and M. Abedi, “DT based intelligent predictor for out of step condition of generator by using PMU data,” Int. J. Electr. Power Energy Syst., vol. 99, pp. 95–106, Jul. 2018.

[13] I. Kanwa, S. R. Samantaray, and G. Joos, “Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices,” IEEE Trans. Smart Grid, vol. 1, no. 2, pp. 144–158, Sep. 2010.

[14] L. Wang, F. Lyu, Y. Su, and J. Yue, “Kernel entropy-based classification approach for superbucket converter circuit fault diagnosis,” IEEE Access, vol. 6, pp. 45504–45514, 2018.

[15] Y. Zhou, Q. Guo, H. Sun, Z. Yu, J. Wu, and L. Hao, “A novel data-driven approach for transient stability prediction of power systems considering the operational variability,” Int. J. Electr. Power Energy Syst., vol. 107, pp. 379–394, May 2019.

[16] Q. Zhu, J. Dang, and J. Chen, “A method for power system transient stability assessment Based on deep belief networks,” (in Chinese), Proc. CSEE, vol. 38, no. 03, pp. 735–743, 2018.

[17] A. Karami, “Power system transient stability margin estimation using neural networks,” Int. J. Electr. Power Energy Syst., vol. 33, no. 4, pp. 983–991, May 2011.

[18] B. Tan, J. Yang, Y. Tang, S. Jiang, P. Xie, and W. Yuan, “A deep imbalanced learning framework for transient stability assessment of power system,” IEEE Access, vol. 7, pp. 81759–81769, 2019.

[19] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam, J. Gu, and V. O. K. Li, “Intelligent time-adaptive transient stability assessment system,” IEEE Trans. Power Syst., vol. 33, no. 1, pp. 1049–1058, Jan. 2018.

[20] Y. Xu, Z. Y. Dong, J. H. Zhao, P. Zhang, and K. P. Wong, “A reliable intelligent system for real-time dynamic security assessment of power systems,” IEEE Trans. Power Syst., vol. 27, no. 3, pp. 1253–1263, Aug. 2012.

[21] K. Narendra and K. Parthasarathy, “Identification and control of dynamical systems using neural networks,” IEEE Trans. Neural Netw., vol. 1, no. 1, pp. 4–27, Mar. 1990.

[22] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[23] M. A. Hannan, N. N. Islam, A. Mohamed, M. S. Hossain Lipu, P. J. Ker, M. M. Rashid, and H. Shareef, “Artificial intelligent based damping controller optimization for the multi-machine power system: A review,” IEEE Access, vol. 6, pp. 39574–39594, 2018.

[24] Y. Le, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
[25] L. Ji, J. Wu, Y. Zhou, and L. Hao, “Using trajectory clusters to define the most relevant features for transient stability prediction based on machine learning method,” *Energies*, vol. 9, no. 11, pp. 898–918, Nov. 2016.

[26] X. Li, Z. Zheng, L. Wu, R. Li, J. Huang, X. Hu, and P. Guo, “A stratified method for large-scale power system transient stability assessment based on maximum relevance minimum redundancy arithmetic,” *IEEE Access*, vol. 7, pp. 61414–61432, 2019.

[27] M. Chen, Q. Liu, S. Chen, Y. Liu, C.-H. Zhang, and R. Liu, “XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system,” *IEEE Access*, vol. 7, pp. 13149–13158, 2019.

[28] Y. Zhou, J. Wu, Z. Yu, L. Ji, and L. Hao, “A hierarchical method for transient stability prediction of power systems using the confidence of a SVM-based ensemble classifier,” *Energies*, vol. 9, no. 10, pp. 778–798, Sep. 2016.

[29] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.

[30] L. B. Lusted, “Logical analysis in roentgen diagnosis,” *Radiology*, vol. 74, no. 2, pp. 178–193, Feb. 1960.

[31] S. Wang, J. Yu, and W. Zhang, “Transient stability assessment using individual machine equal area criterion part II: Stability margin,” *IEEE Access*, vol. 6, pp. 38693–38705, 2018.

[32] I. Kamwa, S. R. Samantaray, and G. Joós, “On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors,” *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 152–161, Mar. 2012.

[33] R. Zhang, J. Wu, Y. Xu, B. Li, and M. Shao, “A hierarchical self-adaptive method for post-disturbance transient stability assessment of power systems using an integrated CNN-based ensemble classifier,” *Energies*, vol. 12, no. 17, pp. 3217–3236, Aug. 2019.

[34] J. H. Chow and K. W. Cheung, “A toolbox for power system dynamics and control engineering education and research,” *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1559–1564, Nov. 1992.

[35] Y. Zhang, Y. Xu, and Z. Y. Dong, “Robust ensemble data analytics for incomplete PMU measurements-based power system stability assessment,” *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1124–1126, Jan. 2018.

[36] L. van der Maaten, “Accelerating T-SNE using tree-based algorithms,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Oct. 2014.

---

**BAOQIN LI** received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2018, where she is currently pursuing the Ph.D. degree. Her research interests include transient stability assessment of power systems, data mining, and deep learning.

**JUNYONG WU** received the B.S., M.S., and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 1987, 1989, and 1993, respectively.

From 1998 to 2004, he was a Postdoctoral Researcher with The University of Tokyo, Tokyo, Japan. Since 2004, he has been a Professor, a Ph.D. Supervisor, the Department Head, and the Vice-Dean of the School of Electrical Engineering, Beijing Jiaotong University, Beijing, China. His research interests include power system analysis and automation, smart grid, electric machine and its systems, the reliability and risk assessment of electrical equipment, and the big data and AI technology and its applications in power systems.

**LIANGLIANG HAO** received the Ph.D. degree from Tsinghua University, Beijing, China, in 2012.

His research interests include fault analysis and protection in power systems, dc transmission control and protection, and excitation control.

**MEIYANG SHAO** received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2017, where she is currently pursuing the M.S. degree. Her research interests include transient stability assessment of power systems, data mining, and deep learning.

**RUOYU ZHANG** received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2014, where she is currently pursuing the Ph.D. degree.

She is also a Visiting Student Researcher with Nanyang Technological University, Singapore. Her research interests include power transient stability assessment, machine learning, and deep learning.

**WEI ZHAO** received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2018. She works at State Grid Beijing Tongzhou Electrical Power Supply Company, China, specializing in distribution automation and relay protection office.