IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 12, 2019, accepted June 14, 2019, date of publication June 19, 2019, date of current version July 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923799

A Deep Imbalanced Learning Framework for Transient Stability Assessment of Power System

BENDONG TAN[1], (Member, IEEE), JUN YANG[1], (Member, IEEE), YUFEI TANG[2], (Member, IEEE), SHENGBO JIANG[1], PEIYUAN XIE[3], AND WEN YUAN[3]

[1]School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China
[2]Department of Computer and Electrical Engineering and Computer Science, Institute for Sensing and Embedded Network Systems Engineering, Florida Atlantic University, Boca Raton, FL 33431, USA
[3]State Grid Hunan Electric Power Company Ltd., Changsha 410007, China

Corresponding author: Jun Yang (jyang@whu.edu.cn)

This work was supported in part by the State Grid Hunan Electric Power Company Ltd., and in part by the Hunan Electric Power Company of China State Grid.

ABSTRACT Maintaining transient stability is a core requirement for ensuring safe operation of power systems. Hence, quick and accurate assessment of the transient stability of power systems is particularly critical. As the deployment of wide area measurement systems (WAMS) expands, transient stability assessment (TSA) based on machine learning with data of phasors measurement units (PMUs) also develops rapidly. However, unstable samples of the power system are rarely seen in practice which affects greatly the effectiveness of transient instability recognition. To address this problem, we propose a deep imbalanced learning-based TSA framework. First, an improved denoising autoencoder (DAE) is constructed to map the training set to hidden space for dimension reduction. Then, adaptive synthetic sampling (ADASYN) is further used to synthesize unstable samples in hidden space to balance the proportion of different classes. Third, the synthesized data are decoded into the original space to enhance the training set. Finally, an ensemble cost-sensitive classifier based on a stacked denoising autoencoder (SDAE) is designed to extract different feature patterns, and the SDAEs are merged with a fusion layer to classify the status of the power system. The simulation results of two benchmark power systems indicate that the proposed method can effectively improve the recognition accuracy of unstable cases by combining nonlinear data synthesis with ensemble cost-sensitive learning methods. Compared with other imbalanced learning methods, the proposed framework enjoys superiority both in accuracy and G-mean.

INDEX TERMS Deep imbalanced learning, transient stability of power system, denoising autoencoder (DAE), ensemble cost-sensitive SDAE, feature patterns, G-mean.

I. INTRODUCTION

Transient stability is the capability of a power system to maintain synchronization when subjected to large disturbances [1]. With the interconnection between large power grids, access to high-penetration renewable energy and the construction of power markets, the dynamic characteristics of power systems are becoming increasingly complicated and the risk of transient instability increases correspondingly [2]. Therefore, real-time and accurate assessment of post-disturbance transient stability is crucial for power system security.

TSA can be represented in a mathematical form by a set of high-dimensional differential-algebraic equations which are called time domain simulations (TDS) [3]. The accuracy of TDS can be improved with an increase in model complexity, but the computing time also rises. The transient energy function (TEF) [4], [5] is a model-based method that assesses the stability by analyzing the increased energy in the power system after clearing a fault, and the assessment result is conservative with TSA for the simplification of the model [6]. Due to these properties, TDS and TEF cannot simultaneously meet the need of accuracy and speed for on-line TSA.

With the rapid development of computer and communication technology, wide-area measurement systems (WAMS)

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

are constructed based on many phasor measurement units (PMUs). WAMS uses a high-speed communication network as the platform to transmit the time-stamped data collected by each PMU to the data center station, and synchronous processing and analysis are performed at the central station. With such characteristics of WAMS, many methods are proposed for TSA based on real-time data provided by PMUs. In [7], model-free and model-based maximal Lyapunov exponents are utilized for on-line TSA with accurate assessment based on real-time system variables. In [8], PMU measurement is applied to single machine equivalent and post-fault trajectory analysis [9] for TSA. Although these methods can achieve accurate assessment, their time-consuming characteristics are an obstacle for them as applied to rapid TSA.

To realize an accurate and rapid TSA, many data-driven methods have been proposed, which mostly employ artificial intelligent (AI) technology to build prediction models offline using massive training datasets. At present, previous relevant research in this area can be divided into two categories. In the first category, researchers use machine learning algorithms to construct the mapping relationship between WAMS data and transient stability. Once a fault occurs, the stability status of a power system can be found according to the mapping relationship with real-time measurements [10], [11]. Works in this area focused on the performance of various classification algorithms and application scenarios, such as decision trees (DT) [12], [13], support vector machines (SVM) [14], extreme learning machine [15], and neural networks (NN) [11]–[16] are applied to online TSA. Time-adaptive TSA has also been proposed in recent years, which uses continuous prediction via the mapping relationship between post-fault trajectories and stability status with long short-term memory (LSTM) [17], [18]. In the second category, time series prediction methods are applied to predict the future trajectory of the system directly, and the stability status of the system can be assessed [19].

However, there are serious class-imbalanced problems in the data-driven TSA methods. Due to the robustness of the modern power system, a realistic power grid can remain stable after suffering most disturbances, and becomes unstable only in a few situations. In this case, if the machine learning model is trained directly via using WAMS data, it will dramatically destroy the performance of the classification model. For instance, if there is only one unstable sample in 100 samples, the model needs to regard all the samples as stable so that it can achieve the accuracy of 99%; however, this is obviously unreasonable. The class-imbalanced problem is primarily solved at the algorithm and data level. Related algorithms include model integration and cost-sensitive methods, such as Easy ensemble [20], Ada-cost [21], etc., but the performance is not ideal with extremely class-imbalanced data. Improvements in data can be divided into two categories: under-sampling and over-sampling methods [22]. The under-sampling method can be applied to a few situations because it loses much of the data information of the class. In over-sampling methods, random oversampling (ROS) [23] tends to cause over-fitting problems, which is not beneficial for training a classification model. To overcome the drawback of ROS, data synthesis methods are introduced, such as synthetic minority over-sampling technique (SMOTE) [24] and adaptive synthetic sampling (ADASYN) [25]. However, they are based on linear interpolation so that it is almost impossible to relate the synthesized data with the physical and operational characteristics of a real power system [22]. A nonlinear imbalanced machine for voltage stability assessment was proposed in [22], but it can only synthesize data for one type of time series.

In order to overcome the aforementioned drawbacks to accomplish accurate and rapid TSA considering the serious class-imbalanced problem, a deep imbalanced learning framework for transient stability assessment of power systems is proposed in this paper. The main contributions of this paper are listed as follows:

1) The proposed framework designs a nonlinear data synthesis method and an ensemble cost-sensitive classifier to implement imbalanced learning for TSA, so that the recognition rate of unstable samples and overall accuracy can be effectively improved.

2) In order to adapt to the environment of the power system, the denoising autoencoder (DAE) is improved by adding wide-area noise to the input layer. ADASYN is further introduced to synthesize unstable samples in the hidden space of the DAE to accomplish nonlinear synthesis to handle non-temporal data for imbalanced learning TSA.

3)The ensemble cost-sensitive stacked denoising autoencoder (SDAE) is improved to extract different patterns by employing multi-SDAEs with a dropout layer for TSA, and it can pay more attention to unstable samples by increasing the cost of the unstable class.

The rest of this paper is divided into five parts. Section II briefly introduces the implications of transient stability assessment. Section III presents the models of ADASYN, DAE, SDAE and ensemble cost-sensitive SDAE. Section IV proposes a deep imbalanced learning framework for TSA and section V discusses numerical results on two benchmark power systems. Finally, the conclusions are drawn in section VI.

II. TRANSIENT STABILITY ASSESSMENT

With the extensive installation of phasor measurement units, it is possible to provide online monitoring for power systems. Specifically, modern TSA is designed to predict the stability status of power systems with real-time measurement after clearance of a fault when it is subjected to severe disturbance. Therefore, rapid and accurate assessment is the core requirement of TSA, to leave enough time for emergency control.

Fig. 1 shows the dynamic curves of all generators in the stable case when bus 16 of an IEEE 39-bus system [5] has suffered a three-phase short-circuit fault. And Fig. 2 shows the unstable case. Once the fault is eliminated, the dynamic measurement will be transferred to the control center to

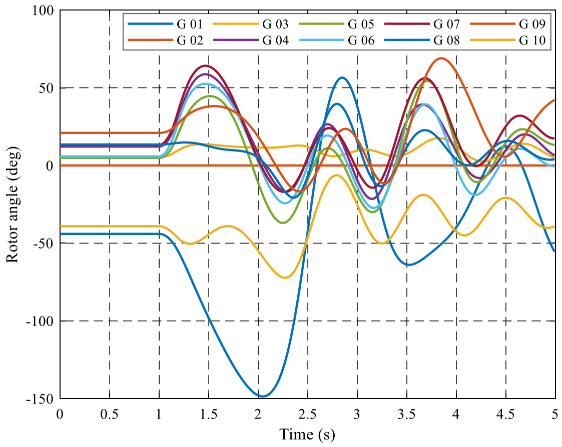


FIGURE 1. Transient stable case.

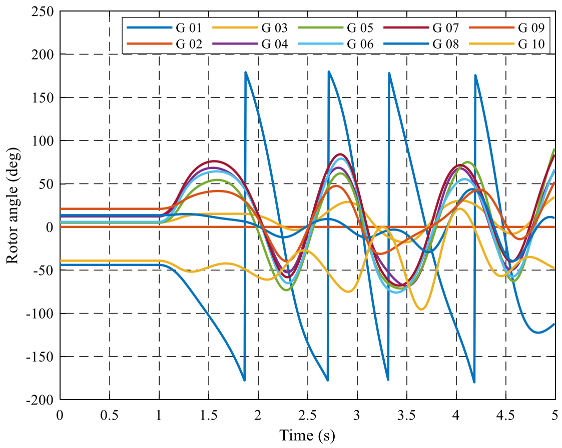


FIGURE 2. Transient unstable case.

predict the stability status for the power system, and then emergency control is conducted. The process of assessment is called TSA. It can be found that the power system loses its stability when some generators lose synchronization. However, with the wide deployment of control devices, it is difficult to cause modern power systems to lose transient stability, which means that there are few unstable cases in the power system database. As a result, data-driven TSA has a huge obstacle in detecting instability patterns.

In order to effectively mine instability patterns from massive power system datasets, this paper considers the factors affecting the mining of instability patterns from two directions. Due to the small amount of unstable data, the data-driven TSA will consider the unstable case to be an abnormal point, so a feasible method is proposed to increase the weight of unstable samples through synthesis. In addition, it can also ensemble models and increase the cost of classification in the unstable case to improve the model's performance in detecting instability, which will help TSA effectively mine unstable patterns. The imbalanced learning framework proposed in this paper is composed of these two aspects, which will be introduced in detail in the following sections.

Algorithm 1 ADASYN

Input: Training data set with m samples $\{x_i, y_i\}$, $i = 1, 2, \cdots, m$, the dimension of x_i is $n \times 1$ and $y_i \in \{1, 0\}$. Suppose the number of minority classes is m_s, the number of majority classes is m_l, and $m_s + m_l = m$.
Output: Synthetic data.

(1) Calculate imbalanced degree of class:

$$d = m_s \div m_l \tag{1}$$

where $d \in (0, 1]$.

(2) Calculate the number of samples that need to be synthesized for the minority class:

$$G = (m_l - m_s) \cdot \beta \tag{2}$$

where $\beta \in (0, 1]$ means G is equal to the difference between the minority class and the majority class when $\beta = 1$, and the minority class is well balanced with the majority class.

(3) Calculate k neighbors with the Euclidean distance for each minority class sample, and Δ is the number of the majority class samples among k neighbors. Then, the ratio r can be calculated as:

$$r = \Delta \div k \tag{3}$$

where $r \in (0, 1]$.

(4) Get the r_i of each minority class sample in (3) and represent the situation for each minority class sample using surrounding majority class samples:

$$\hat{r}_i = r_i / \sum_{i=1}^{m_s} r_i \tag{4}$$

(5) Calculate the number of synthetic samples for each minority class sample:

$$g_i = \hat{r}_i \cdot G \tag{5}$$

(6) Select one minority class sample among k neighbors around each minority class sample to be synthesized. Synthesize data according to the following formula:

$$s_i = x_i + (x_{zi} - x_i) \cdot \lambda \tag{6}$$

where s_i is the synthetic data, x_{zi} is a minority class sample randomly chosen from k neighbors of x_i.

III. ADASYN, DAE AND ENSEMBLE COST-SENSITIVE SDAE MODELS

A. ADASYN PROCESS

He *et al.* [25] proposed a novel adaptive synthetic sampling approach for learning from imbalanced data, termed ADASYN. It can adaptively synthesize minority class samples according to their level of difficulty in learning, which means it generates more data for the minority class that is harder to learn. The ADASYN process can be summarized in Algorithm 1.

B. UNSUPERVISED FEATURE LEARNING AND ENSEMBLE

1) DAE

DAE is a type of autoencoder (AE) [26]. As shown in Fig. 3(a), DAE is a three-layer neural network which is trained to try to copy input to output. By adding wide-area noise to the input, DAE learns to remove noise and to approximate the original input data. As a result, DAE can learn more stable and meaningful features, which constitute a more robust representation of the input data. In general, DAE consists of two parts: the encoder and decoder. The encoder can extract higher order features from input data and the decoder can transform the higher-order features to input data. In other words, the structure of the encoder and decoder are symmetrical.

The function of the encoder can be express as:

$$\boldsymbol{h} = f(\boldsymbol{x}, \boldsymbol{\theta}_e) = \sigma(\boldsymbol{W}_1(\boldsymbol{x} + \boldsymbol{\varepsilon}) + \boldsymbol{b}_1) \tag{7}$$

where $\boldsymbol{\theta}_e = [\boldsymbol{W}_1, \boldsymbol{b}_1]$ is the parameter of the encoder, $\boldsymbol{W}_1$ is the weight matrix, $\boldsymbol{b}_1$ is the bias vector of all neurons in the encoder, $\boldsymbol{\varepsilon}$ is a random error vector composed of the same white noise used in WAMS to adapt the environment of the power system, and $\sigma = 1/(1 + \exp(-x))$ is an active nonlinear transformation function. The hidden layer output $\boldsymbol{h}$ implements compression on the high-dimensional input data by feature extraction, which means the encoder preserves the main features while removing irrelevant information.

Then, the hidden layer output $\boldsymbol{h}$ is decoded to the original high-dimensional space by:

$$\hat{\boldsymbol{x}} = g(\boldsymbol{h}, \boldsymbol{\theta}_d) = \sigma(\boldsymbol{W}_2\boldsymbol{h} + \boldsymbol{b}_2) \tag{8}$$

where $\boldsymbol{\theta}_d = [\boldsymbol{W}_2, \boldsymbol{b}_2]$ is the decoder parameter, $\boldsymbol{W}_2$ is the weight matrix, and $\boldsymbol{b}_2$ is the bias vector of all neurons in decoder. When the norm of the difference between the decoded data and the original data is small, it is believed that the hidden layer output of the encoder can represent the characteristics of the input data. The training process finds the optimal $[\boldsymbol{\theta}_e, \boldsymbol{\theta}_d]$ to minimize the square reconstruction of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ by the gradient descent algorithm:

$$L = \sum_{i=1}^{m} \left\| \boldsymbol{x}_i - \hat{\boldsymbol{x}}_i \right\|^2 \tag{9}$$

where L is the loss of the DAE and m is the number of input data samples. Once trained, the DAE can conduct nonlinear transformation and extract robust features from input data with noise for classification or regression.

2) SDAE

An SDAE [26] consists of multiple DAEs, which stacks each DAE in a deep structure, as shown in Fig. 3. An SDAE is able to overcome the difficulty of determining highly nonlinear and complex patterns by learning hierarchical features. The training process of an SDAE includes multiple unsupervised pre-training steps and a supervised fine-tuning step. In the unsupervised pre-training steps, an effective way to obtain good parameters for an SDAE is by using greedy layer-wise training, which is implemented by training stacked DAEs in an encoder network by training a layer each time before starting the next layer. As shown in Fig3.(b), given a set of training data, the first layer is trained to obtain the hidden representation $\boldsymbol{h}_1$, then the second layer is trained on the hidden representation $\boldsymbol{h}_1$ for higher representation $\boldsymbol{h}_2$, and so on for subsequent layers. In this way, SDAE can extract a robust nonlinear representation from input data in an unsupervised manner.

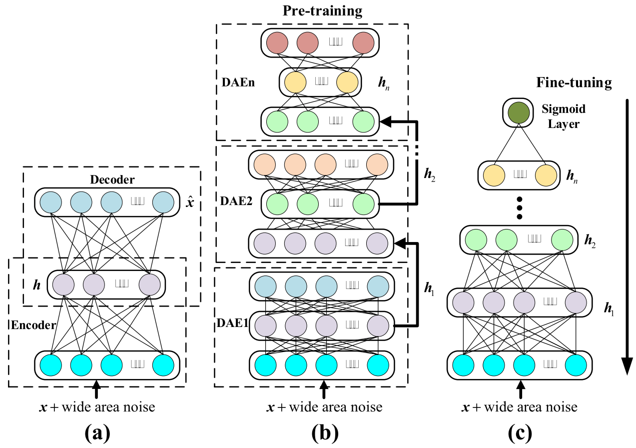


FIGURE 3. (a) DAE structure. (b) SDAE pretraining process. (C) SDAE fine-tuning process.

In the fine-tuning process, as shown in Fig. 3(c), all hidden layers trained during pre-training are stacked and a sigmoid layer is added on the top of the stacked deep learning architecture. To train an SDAE efficiently, the parameters in all layers of the SDAE are connected to corresponding parameters learned in the pre-training phase, then the SDAE is fine-tuned with label information by the back propagation (BP) algorithm [27].

In this paper, an SDAE is used as an excellent feature extractor to learn robust nonlinear representation for classification from noisy input data. To obtain accurate TSA, an ensemble-cost SDAE is proposed in next section while considering imbalanced learning.

3) ENSEMBLE COST-SENSITIVE SDAE

TSA can be regard as an imbalanced learning problem as the power system is too strong to lose its stability. Therefore, if TSA does not consider the imbalance in the power system, it tends to regard the system as stable. Once the unstable situation is judged to be stable, it will cause security risk to the power system. In other words, the cost of stability and instability are different. To mine the pattern of instability effectively, an ensemble cost-sensitive SDAE is designed, as shown in Fig. 4. From left to right, there are input layers with selected features of a power system, a supervised learning process is represented by ensembled DAEs to learn from input layers, and a fusion layer and cost-sigmoid layer are deployed to merge higher order representations in the DAEs.

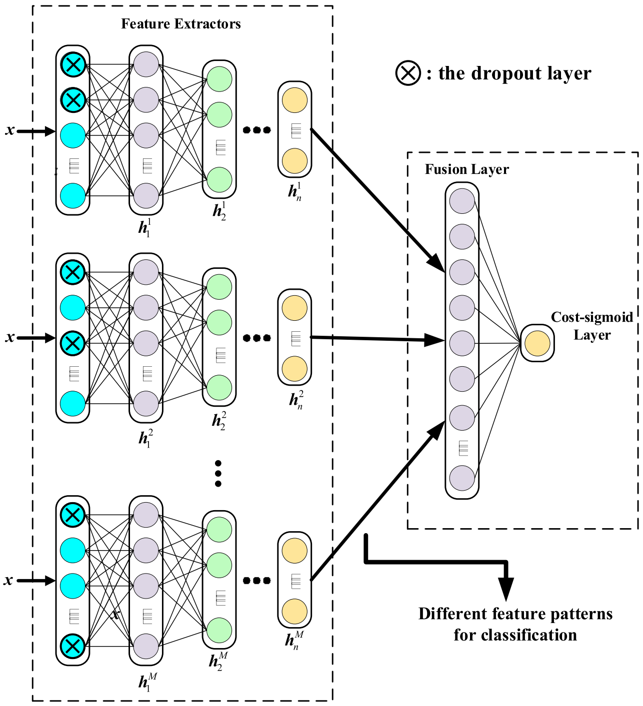


FIGURE 4. Structure of ensemble cost-sensitive SDAE.

At the left of the ensemble-cost SDAE in Fig. 4, M SDAEs are the feature extractors for unlabeled data. A single SDAE can only capture a specific feature of the input data. However, the power system stability pattern is not dominated by just one mode, so a dropout layer is added for every SDAE to obtain different patterns. The features learned by the SDAEs are heterogeneous because the features represent different physical attributes that humans do not understand. Then, the captured features of multiple SDAEs are merged by the fusion layer using successive cost-sigmoid layers, and the loss function of the cost-sigmoid layer, can be written as:

$$L_1 = \sum_{i=1}^{m} y_i \log(t_i) \cdot w_s + (1 - y_i) \log(1 - t_i) \cdot w_l \quad (10)$$

where L_1 is the loss function of the cost-sigmoid layer, y_i is label of i-th sample, t_i is the prediction of the i-th sample, w_s and w_l are weight of unstable and stable samples, respectively. In order to increase the cost of the classification of unstable samples, w_s is set larger than w_l and w_l is always set to 1 as a benchmark so that the ensemble cost-sensitive SDAE can effectively mine the unstable pattern. Finally, the cost-sigmoid layer can fine-tune the features learned by the SDAEs and fuse the layers using labeled data. In short, the different patterns of features in data are extracted via multiple SDAEs and a fusion layer, while the cost-sigmoid layer is regard as a classifier.

For an ensemble cost-sensitive SDAE, the M is larger, and the performance of the model will improve and the training time will increase correspondingly. To speed up learning with M as large as possible, the batch stochastic gradient descent (SGD) method with momentum, which is called Adam [28], can be introduced to update the weight of the model. In this way, an ensemble cost-sensitive SDAE can extract robust nonlinear features from data and perform excellent imbalanced learning to realize balanced TSA.

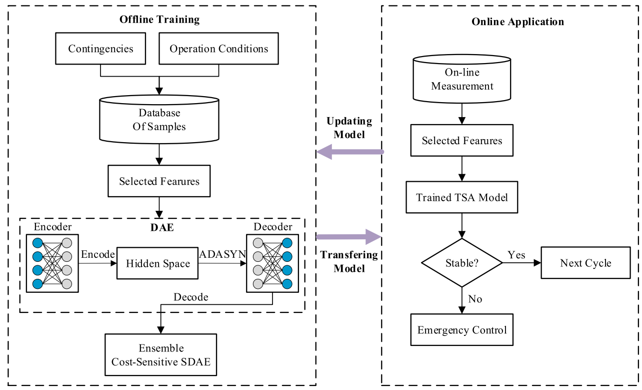


FIGURE 5. Structure of the deep imbalanced learning framework.

IV. DEEP IMBALANCED LEARNING FRAMEQORK FOR TSA

Until now, many methods have been proposed for online TSA, and they can be summarized into two modes: mode A and mode B. Mode A generates a large number of transient samples offline to train the model. When the real-rime data arrives, the trained TSA model can be deployed immediately, and the model is updated periodically as many new transient samples are collected. Mode B trains the TSA model with a massive number of samples produced by time domain simulation offline as well, but it can update itself online as it conducts TSA for a power system. However, the fast update speed of the model in mode B sacrifices generalization, so mode A is used in this paper, which is also called ''offline training, online application''.

The deep imbalanced learning framework for online TSA proposed in this paper can be summarized in four steps: 1) feature construction for TSA; 2) index for model performance evaluation; 3) offline training; and 4) online application. The deep imbalanced learning framework is shown in Fig. 5.

A. FEATURE CONSTRUCTION FOR TSA

A simple WAMS consists of several PMUs and a phase data concentrator (PDC). Typically, a PMU is installed at a substation and the data sampled by all PMUs is sent to a PDC at a location where the data are aggregated and analyzed. As a result, the WAMS can report data at the frequency in which the power system works for situational awareness to provide complex analysis, control and protection in real-time. In order to effectively utilize the real-time information provided by a WAMS, many studies have conducted much meaningful work [29]–[31]. This research generally transforms measurements into statistics, such as maximum value of voltage, and then uses these statistics as feature inputs of the TSA model. However, information loss is inevitable in such artificial feature engineering. The model proposed in this paper is an end-to-end analysis process, which avoids

TABLE 1. Confusion matrix.

	Actually stable	Actually unstable
Predictively stable	f_{11}	f_{12}
Predictively unstable	f_{21}	f_{22}

information loss in the powerful feature extraction ability without the feature conversion process. Some elements that can reflect system dynamics are chosen to construct the input features for TSA; they include generator active power, generator reactive power, bus voltage magnitude, bus voltage angle, branch active power, branch reactive power, load active power and load reactive power.

B. INDEX FOR MODEL PERFORMANCE EVALUATION

Since TSA has the defect of ignoring unstable samples, it is unreasonable to use only the accuracy of all samples as the evaluation index for the model. Assume that stable samples account for 99% of the dataset, then the TSA model can achieve an accuracy of 99% by judging all samples to be stable. In order to effectively evaluate the performance of the model under the class imbalance, this paper utilizes the confusion matrix [32] shown in TABLE 1.

In a confusion matrix, f_{11} represents the number of actual stable samples that are predicted to be stable and so on for f_{12}, f_{21} and f_{22}. The index to evaluate the performance of the model is defined as follows:

$$TSR = \frac{f_{11}}{f_{11} + f_{21}} \times 100\% \tag{11}$$

$$TUR = \frac{f_{22}}{f_{12} + f_{22}} \times 100\% \tag{12}$$

$$G - mean = \sqrt{TSR \cdot TUR} \times 100\% \tag{13}$$

$$ACC = \frac{f_{11} + f_{22}}{f_{11} + f_{12} + f_{21} + f_{22}} \times 100\% \tag{14}$$

TSR represents the proportion of correct results predicted to be stable from all stable samples, TUR represents the proportion of correct results predicted to be unstable for all unstable samples, G-mean is the geometric mean of TSR and TUR, which can effectively evaluate the performance of the imbalanced data, and ACC represents the overall accuracy.

C. OFFLINE TRAINING

1) TRAIN THE IMBALANCED LEARNING MODEL OFFLINE

With access to a high proportion of renewable energy, modern power systems operate under a variety of conditions, so as many typical operatiing conditions as possible are considered. In order to find the boundary of instability, this paper only considers the most serious contingency, which is the three-phase short circuit. Datasets include thousands of samples that can be generated with different operation conditions and contingencies.

In the procedure as Figure 5 shown, the datasets with selected features are divided into training and testing data. DAE converts training data into hidden space, and the hidden space data are synthesized by ADASYN to balance the class amounts, which means the ratio of numbers of stable samples and unstable samples is set to 1:1 by synthetically increasing the number of unstable samples. Then, the synthetic data are converted back to the original space of the training data. Finally, the decoded data are used as input for the ensemble cost-sensitive SDAE, and it is trained using the method introduced in section III. Testing data are utilized to evaluate the trained model with the indices introduced in section IV for good performance in subsequent online application.

2) MODEL UPDATING MECHANISM

As the working conditions change dynamically, the model requires incremental learning to maintain adaptability. A model updating mechanism is designed to implement incremental learning as follows:

(1) When the power system collects records of some contingencies without a topology change, it retrains the model with the parameter initialization of the existing model, and repeats this step periodically.

(2) When the system finds that the topology has changed, the model needs to regenerate the datasets and select critical features for retraining.

D. ONLINE APPLICATION

Once a fault is detected, the online application program is triggered. After a cycle of a fault clearing, real-time measurements are acquired in observation window of a cycle, and the critical features are selected from them. Then, we input these critical features to the ensemble cost-sensitive model, and the model will perform TSA for the power system. If the power system is assessed as unstable, emergency control will be performed immediately; otherwise, the trained TSA continues monitoring the stability status of the power system during the next cycle.

V. CASE STUDY

In this section, analysis is implemented on an IEEE 39-bus system and an South Carolina 500-Bus System [33] in order to test the effectiveness of the proposed method. Keras [34] offers consistent and simple APIs for deep learning, and their tools are used to construct the DAE and ensemble cost-sensitive SDAE; the framework proposed in this paper is built in Python. All programs are compiled on a laptop with Intel Core i5-7300HQ CPU, 8GB RAM and a 1050 2D/3D graphics card with 2GB memory.

A. CASE OF THE IEEE 39-BUS SYSTEM

The IEEE 39-bus system has been widely used to test TSA algorithms. A line diagram of an IEEE 39-bus system is shown in Fig. 6. The system is composed of 39 buses, 10 generators, 19 loads and 34 transmission lines. The generator on the bus-39 bus is the equivalent of the external large power system, and its rotor angle is used as a reference for other generators. The required time domain simulation is

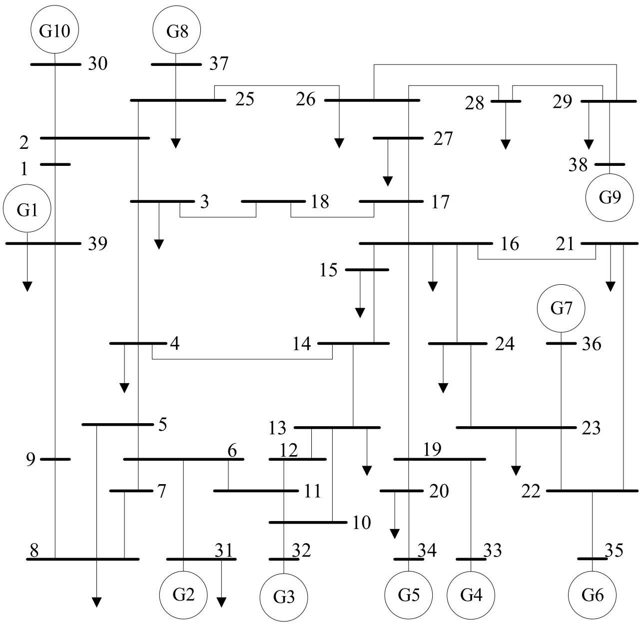


FIGURE 6. Line diagram of an IEEE 39-bus power system.

performed with DIgSILENT PowerFactory [35], which is a power system simulation tool.

B. DATABASE GENERATION

The datasets are constructed via time domain simulation, the generator is a six-order model, and the load uses the constant impedance model. Considering a load level ranging from 75% to 125% in a step size of 5%, with the generator output level adjusted accordingly. In other words, the load level and generator output level change at the same ratio. Then, power flow is calculated, and if the power flow has converged, the operation condition is preserved.

The contingencies considered are the three-phase grounding fault that occurs on all buses and on all lines, where the fault is located at 20%, 40%, 60%, and 80% of their length. Further, fault duration includes 0.1 s, 0.2 s and 0.3 s, and the time domain simulation duration is set to 10 s. Finally, 5775 samples are generated to include 3890 stable samples and 1885 unstable samples, and the dimension of the dataset variable is $10 + 10 + 39 + 39 + 46 + 46 + 19 + 19 = 228$. In order to simulate the situation where the number of unstable samples in the actual power system is scarce, some unstable samples are discarded randomly so that the number of unstable samples accounts for approximately 5% of the total number of samples.

Power system transient stability assessment can be divided into two states based on the transient stability index (TSI), by calculating the maximum rotor angle separation of any two generators. TSI η can be defined as [32]:

$$\eta = \frac{360 - \Delta\delta_{\max}}{360 + \Delta\delta_{\max}} \tag{15}$$

where $\Delta\delta_{\max}$ is the maximum rotor angle separation of any two generators at the end of post-fault time domain simulation. If $\eta > 0$, the sample is label as "1" for the stable case, otherwise it is labeled as "0" for the unstable case.

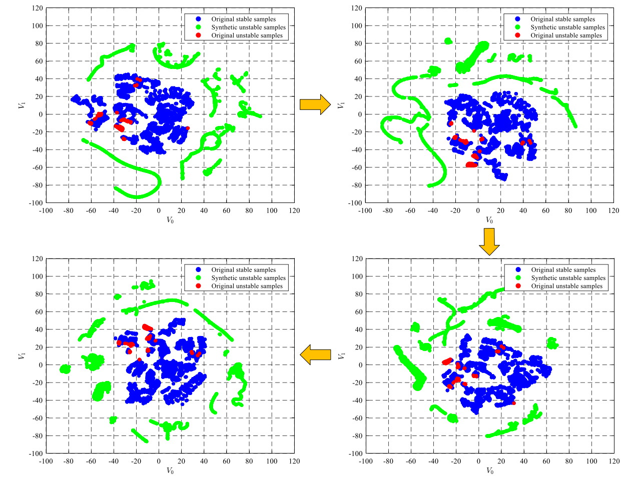


FIGURE 7. Dynamic training process of SYNDAE.

To prevent the model from overfitting and effectively evaluating its performance, datasets are randomly divided into 3:1 as training data and testing data, respectively. The proposed method is trained in several minutes so that it is efficient to deploy it in real system. And detailed discussion for the proposed methods is introduced in the following sections.

C. SPATIAL DISTRIBUTION ANALYSIS OF SYNTHETIC DATA

The experiment is implemented to explain the physical meaning of the nonlinear data synthesis method proposed in this paper, and the methods including DAE and ADASYN are hereafter referred to as SYNDAE.

To obtain excellent representation of the input, the dimension of the encoder is set to 100 because the dimension of the encoder is generally smaller than the input, and the overall training epoch is 40. Fig. 7 shows the training process of SYNDAE which presents spatial changes of synthetic data dynamically. According to the direction pointed by the arrow, the data synthesized by the 1-th, 11-th, 21-th and 36-th training generations are compressed into two-dimensions V_0 and V_1 by TSNE [36] respectively, where the blue points are the original stable samples, the red points are the original unstable samples and the green points are the synthetic unstable samples. From Fig. 7 it can be observed that the data synthesized by SYNDAE becomes more and more nonlinear as the training progresses, which indicates that it captures the nonlinear unstable patterns and effectively synthesizes it. As SYNDAE encodes the original data to hidden space where mined patterns of transients related to physical characteristics of transient processes of the power system, then these patterns are to be balanced with synthesis of ADASYN. Therefore, it can enhance the transient characteristics of the data when unstable patterns are decoded to balance the original data.

Different data synthesis methods, including SMOTE, ADASYN, ROS and SYNDAE, respectively balance the original data and compress these balanced sets into a

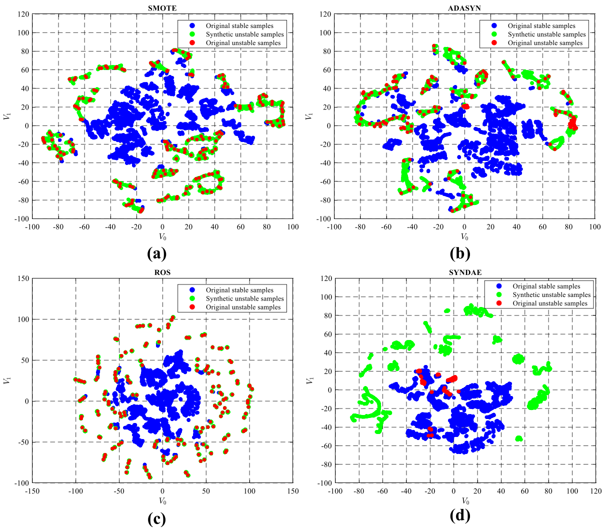


FIGURE 8. (a) Two-dimensional spatial distribution of SMOTE. (b) Two-dimensional spatial distribution of ADASYN. (c) Two-dimensional spatial distribution of ROS. (d) Two-dimensional spatial distribution of SYNDAE.

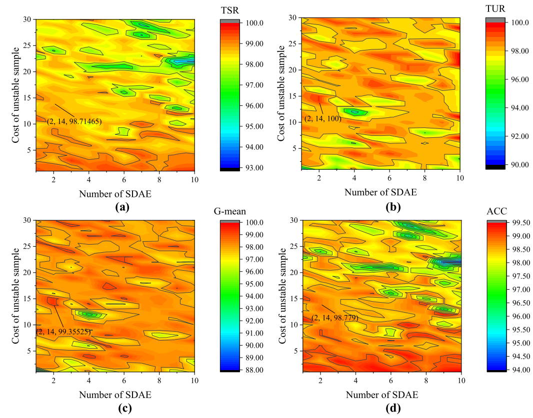


FIGURE 9. (a) Distribution of TSR. (b) Distribution of TUR. (c) Distribution of G-mean. (d) Distribution of ACC.

two-dimensional spatial distribution as shown in Fig. 8. Fig. 8 (d) also shows the final training result of SYNDAE. Compared with these other three methods, SYNDAE clearly belongs to nonlinear data synthesis methods by mining complex patterns from original data. The synthetic unstable data of SYNDAE and the original data form a stable boundary with a significantly larger interval than other methods. This is because SMOTE and ADASYN are based on linear interpolation, so then the synthetic unstable samples are distributed between the original unstable samples. In other words, the data synthesized by SMOTE and ADASYN are not implemented according to the transient patterns, so the correlation with the physical characteristics of the power system is not strong enough to form a stable boundary as large as SYNDAE. ROS randomly copies unstable samples to balance the original data results in the dispersion of the spatial distribution of unstable patterns, which undermines the generalization to a certain extent. Therefore, SYNDAE can use the synthetic unstable patterns to help classifiers more effectively distinguish transient data.

D. CLASSIFICATION PERFORMANCE ANALYSIS

In order to verify the effectiveness of the proposed method on TSA, this paper compares several traditional methods which are SMOTE, ADASYN, ROS, random under sampling (RUS), conditional generative adversarial network (CGAN), to highlight its advantages. All results are obtained by using training and testing data.

The hyperparameter of the model is determined by grid search optimization, which includes the number of SDAE and the cost of the unstable samples. Fig. 9 shows the optimal parameter search process of the proposed method. In order to highlight the distribution of various evaluation indices, contour is plotted. It can be observed that TUR increases as the cost increases, but as attention to the stable samples decreases, TSR also decreases. And because the number of stable samples is much larger than unstable samples, so the overall accuracy of samples is diminished. G-mean can be used to make trade-offs between TUR and TSR. On the premise of ensuring of maximum G-mean, a lightweight model is obtained in this paper. Finally, the number of SDAEs is set to 2, and the cost of unstable samples is set to 14.

TABLE 2. TSA results for different methods.

Models	TSR%	TUR%	G-mean%	ACC%
Original + ELM	99.36	80.49	89.43	98.41
SYNDAE+ ELM	99.10	87.80	93.28	98.53
SMOTE + ELM	96.65	97.56	97.10	96.70
ADASYN+ ELM	96.91	95.12	96.82	96.01
ROS+ ELM	96.40	97.56	96.98	96.46
RUS+ELM	91.77	100.00	95.80	92.19
CGAN+ ELM	97.56	95.12	96.33	97.44
Proposed method	98.71	100.00	99.36	98.78

TSA results for several different methods are provided in TABLE 2. Extreme learning machine (ELM) [37] is used as a benchmark classifier for comparison and has been widely applied in TSA and gains excellent performance [38]–[40]. Compared to the original data, SYNDAE can significantly increase TSR without reducing TUR when the classifier is ELM, so its ACC is the highest compared to other data synthesis methods that sacrifice TSR to improve TUR. Compared with the imbalance of the learning methods listed in TABLE 2, the proposed method outperforms in terms of G-mean and ACC which realizes 100% of TUR with 98.81% of TSR. Other methods cannot achieve a good balance between TSR and TUR, resulting in a decrease of overall accuracy and failure to meet requirements for TSA.

As SMOTE and ADASYN only linearly interpolate the unstable samples, so transient characteristics of the data are not effectively extracted, and the recognition rate of the stable

TABLE 3. TSA results for different methods with wide-area noise.

Models	TSR%	TUR%	G-mean%	ACC%
Original + ELM	99.23	80.49	89.37	98.29
SYNDAE+ ELM	99.23	85.37	92.04	98.53
SMOTE + ELM	97.94	95.12	96.52	97.80
ADASYN+ ELM	97.42	92.68	95.02	97.19
ROS+ ELM	97.17	92.68	94.90	96.95
RUS+ ELM	90.35	100.00	90.74	95.06
CGAN+ ELM	98.07	92.68	95.34	97.80
Proposed method	98.71	100.00	99.36	98.78

TABLE 4. TSA results for a large power system.

Models	TSR%	TUR%	G-mean%	ACC%
Original + ELM	99.96	0.81	8.98	96.42
SYNDAE+ ELM	97.87	86.02	91.75	97.44
SMOTE + ELM	85.95	90.59	88.24	86.12
ADASYN+ ELM	89.45	93.01	91.21	89.58
ROS+ ELM	93.89	84.14	88.88	93.53
RUS+ ELM	86.99	91.13	89.03	87.14
CGAN+ ELM	98.43	93.01	95.68	98.24
Proposed method	98.21	95.16	96.67	98.10

sample is reduced with ELM after synthetically increasing the number of unstable samples. ROS simply repeats the unstable samples, resulting in a decrease of TSR and damage to the generalization of ELM. RUS delete some stable samples to balance the unstable samples, however, leading to harming the ability to recognize the stable samples. And CGAN is difficult to be trained to get the excellent performance for imbalanced TSA. While SYNDAE mines transient patterns effectively by synthesizing the unstable samples in hidden space, so that TUR can be improved while maintaining TSR. The ensemble cost-sensitive SDAE further enhances TUR with powerful feature extraction capabilities and increases the cost with misclassification of unstable samples. This indicates that the proposed method can effectively identify unstable samples and maintain a very high G-mean and ACC.

E. IMPACT OF WIDE-AREA NOISE

This paper assumes that data of WAMS is not polluted in the previous discussion; however, PMU measurements are accompanied by noise that affects the performance of the model used for TSA. Therefore, this section will discuss the impact of wide-area noise on the proposed methods and compare the robustness of different methods.

According to IEEE standards [13], the PMU measurement error cannot exceed 1% of the measured value. The generation of wide-area noise can be referenced in [41], which is summarized as:

$$\boldsymbol{x}_i^{meas} = \boldsymbol{x}_i^{real} + \boldsymbol{\varepsilon} \tag{16}$$

where $\boldsymbol{x}_i^{meas}$ is the measured phasor and $\boldsymbol{x}_i^{real}$ is the real value phasor without measurement error. To meet the standards of IEEE, reasonable white noise $\boldsymbol{\varepsilon}$ is added to the data in the time domain simulation. Training data and testing data are corrupted with white noise, and all parameters of the model are the same as the previous test.

TABLE 3 presents the results for different methods balancing classes for TSA, considering wide-area noise. SYNDAE is almost not affected by noise while the TSR of SMOTE, ADASYN, ROS, RUS and CGAN drops significantly with the ELM classifier. Performance of the proposed method does not decrease due to wide-area noise, so it performs better than other methods in terms of G-mean and ACC. As SMOTE, ADASYN and ROS will add wide-area noise in the synthesized data for wide-area noise outputs in the original data. In addition, CGAN and RUS are affected by noise slightly due since they do not interpolate. However, SYNDAE and the proposed method consist of DAEs which are designed considering wide-area noise and can denoise the measurement error.

F. LARGER POWER SYSTEM CASE

In order to test the scalability of the proposed method, a larger power system is employed to be analyzed, namely South Carolina 500-Bus System. The same database generation method used in the IEEE 39-bus system is utilized to construct datasets that includes 34725 samples with 1243 unstable samples, and the datasets are split into training data (80% of datasets) and testing data (20% of datasets) for the proposed method.

As seen in TABLE 4, the original data has the poorest ability to detect unstable samples while SYNDAE can maintain the generalization for stable samples with moderate performance of unstable samples; therefore, SYNDAE achieves the highest ACC among the eight imbalanced learning methods. When the ensemble cost-sensitive SDAE is trained with data synthesized by SYNDAE, the proposed method efficiently extracts stable and unstable patterns in the datasets, resulting in desirable performance of ACC and G-mean. CGAN shows good performance at larger power system, but its ability to capture the unstable situation is not stronger than proposed method. By intuitively repeating or deleting unstable samples, ROS and RUS creates a bottleneck by recognizing them. Additionally, methods based on linear interpolation, such as SMOTE and ADASYN, cannot capture enough transient characteristics of the power system to achieve state-of-the-art performance.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, to overcome the imbalanced-class problem in TSA, a deep imbalanced learning framework consisting of nonlinear data synthesis method SYNDAE and ensemble cost-sensitive SDAE is presented. Simulations of the proposed framework on benchmark power systems demonstrate superior performance for TSA compared to traditional imbalanced learning methods. The main conclusions can be drawn as follows:

1) Compared with the methods based on linear interpolation, the proposed nonlinear data synthesis method SYNDAE demonstrates improved capability in extracting transient characteristics for TSA with non-temporal data.

2) The proposed ensemble cost-sensitive SDAE can effectively mine transient patterns in data and increase attention to unstable samples so as to improve their recognition accuracy.

3) The proposed framework greatly improves the generalization of implementing TSA with nonlinear data synthesis and ensemble cost-sensitive learning compared with existing imbalanced learning methods such ADASYN, SMOTE and ROS, etc.

In practice, PMU failure, phasor data concentrator (PDC) failure and communication delay may cause missing data both in cyber and physical power systems, resulting in poor performance for TSA. Therefore, the information characteristics of WAMS data in real time will be considered in future works. In addition, the topology changes of power system is another problem to be studied, for which would affect the adaptability of the data-driven model.

REFERENCES

[1] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. Van Cutsem, and V. Vittal, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387–1401, Aug. 2004.

[2] C. Zhang, Y. Li, Z. Yu, and F. Tian, "A weighted random forest approach to improve predictive performance for power system transient stability assessment," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Beijing, China, Oct. 2016, pp. 1259–1263.

[3] *Power System Stability and Control*. New York, NY, USA: McGraw-hill, 1994, pp. 848–872.

[4] 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. PWRS-3, no. 2, pp. 400–412, May 1988.

[5] *Energy Function Analysis for Power System Stability*. Berlin, Germany: Springer, 2012.

[6] R. Zhang, Y. Xu, Z. Y. Dong, and K. P. Wong, "Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system," *IET Gener., Transmiss. Distrib.*, vol. 9, no. 3, pp. 296–305, Feb. 2015.

[7] S. Wei, M. Yang, J. Qi, J. Wang, S. Ma, and X. Han, "Model-free MLE estimation for online rotor angle stability assessment with PMU data," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2463–2476, May 2018.

[8] *Transient Stability of Power Systems: A Unified Approach to Assessment and Control*. Berlin, Germany: Springer, 2012.

[9] D. R. Gurusinghe and A. D. Rajapakse, "Post-disturbance transient stability status prediction using synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3656–3664, Sep. 2016.

[10] Y. Tang, F. Li, Q. Wang, and Y. Xu, "Hybrid method for power system transient stability prediction based on two-stage computing resources," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 8, pp. 1697–1703, Apr. 2018.

[11] Q. Zhu, J. Chen, L. Zhu, D. Shi, X. Bai, X. Duan, and Y. Liu, "A deep end-to-end model for transient stability assessment with PMU data," *IEEE Access*, vol. 6, pp. 65474–65487, 2018.

[12] T. Guo and J. V. Milanović, "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.

[13] M. He, V. Vittal, and J. Zhang, "Online dynamic security assessment with missing PMU measurements: A data mining approach," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1969–1977, May 2013.

[14] F. R. Gomez, A. Rajapakse, U. Annakkage, and I. 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.

[15] X. Gu, Y. Li, and J. Jia, "Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm," *Int. J. Elect. Power Energy Syst.*, vol. 64, pp. 664–670, Jan. 2015.

[16] S. K. Tso, X. P. Gu, Q. Y. Zeng, and K. L. Lo, "An ANN-based multilevel classification approach using decomposed input space for transient stability assessment," *Electr. Power Syst. Res.*, vol. 46, no. 3, pp. 259–266, Sep. 1998.

[17] 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.

[18] J. J. Q. Yu, A. Y. S. Lam, D. J. Hill, and V. O. K. Li, "Delay aware intelligent transient stability assessment system," *IEEE Access*, vol. 5, pp. 17230–17239, 2017.

[19] X. Liu, Q. Jiang, and Y. Cao, "A novel fast transient stability prediction method based on perturbed trajectories fitting of rotor angle," (in Chinese), *Autom. Electr. Power Syst.*, vol. 32, no. 19, pp. 5–9, Oct. 2008.

[20] T.-Y. Liu, "Easyensemble and feature selection for imbalance data sets," in *Proc. Int. Joint Conf. Bioinf., Syst. Biology Intell. Comput.*, Shanghai, China, Aug. 2009, pp. 517–520.

[21] W. Fan, S. J. Stolfo, and J. Zhang, "AdaCost: Misclassification cost-sensitive boosting," in *Proc. 16th Int. Conf. Mach. Learn.*, Bled, Slovenia, Jan. 1999, pp. 97–105.

[22] L. Zhu, C. Lu, Z. Y. Dong, and C. Hong, "Imbalance learning machine-based power system short-term voltage stability assessment," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2533–2543, Oct. 2017.

[23] G. E. Batista, R. C. Prati, and M. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, 2004.

[24] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[25] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Hong Kong, Jun. 2008, pp. 1322–1328.

[26] 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.

[27] Y. Hirose, K. Yamashita, and S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," *Neural Netw.*, vol. 4, no. 1, pp. 61–66, 1991.

[28] D. Kinga and J. B. Adam, "A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–15.

[29] F. Hang, S. Huang, Y. Chen, and S. Mei, "Power system transient stability assessment based on dimension reduction and cost-sensitive ensemble learning," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Beijing, China, Nov. 2017, pp. 1–6.

[30] Y. Li and Z. Yang, "Application of EOS-ELM with binary Jaya-based feature selection to real-time transient stability assessment using PMU data," *IEEE Access*, vol. 5, pp. 23092–23101, 2017.

[31] D. Huang, Q. Chen, S. Ma, Y. Zhang, and S. Chen, "Wide-area measurement—Based model-free approach for online power system transient stability assessment," *Energies*, vol. 11, no. 4, p. 958, Apr. 2018.

[32] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, "Power system transient stability assessment based on big data and the core vector machine," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2561–2570, Sep. 2016.

[33] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3258–3265, Jul. 2017.

[34] *Keras: The Python Deep Learning Library*. Accessed: Dec. 1, 2018. [Online]. Available: https://keras.io/

[35] *PowerFactory_DIgSILENT Germany*. Accessed: Dec. 1, 2018. [Online]. Available: http://www.digsilent.de/index.php/products-powerfactory.html

[36] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[37] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.

[38] Y. Xu, Z. Y. Dong, K. Meng, R. Zhang, and K. P. Wong, "Real-time transient stability assessment model using extreme learning machine," *IET Gener., Transmiss. Distrib.*, vol. 5, no. 3, pp. 314–322, Mar. 2011.

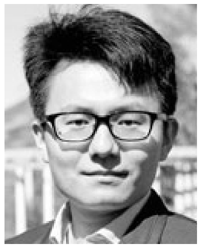
[39] Y. Zhang, Y. Xu, Z. Y. Dong, Z. Xu, and K. P. Wong, "Intelligent early warning of power system dynamic insecurity risk: Toward optimal accuracy-earliness tradeoff," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2544–2554, Oct. 2017.

[40] C. Ren, Y. Xu, and Y. Zhang, "Post-disturbance transient stability assessment of power systems towards optimal accuracy-speed tradeoff," *Protection Control Mod. Power Syst.*, vol. 3, no. 1, p. 19, Jun. 2018.

[41] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1978–1987, May 2013.

BENDONG TAN (M'18) received the B.S. degree in electrical engineering from Wuhan University (WHU), Wuhan, China, in 2017, where he is currently pursuing the master's degree in electrical engineering and automation. His research interests include machine learning, power system secure operation and control, and the applications of big data in smart grids.

JUN YANG (M'05) was born in Hubei, China, in 1977. He received the Ph.D. degree from the Huazhong University of Science and Technology, China, in 2006. He then joined Wuhan University, Wuhan, China, where he is currently a Professor. His research interests include power system security, electrical vehicles, and smart grids.



YUFEI TANG received the Ph.D. degree in electrical engineering from the University of Rhode Island, in 2016. He is currently an Assistant Professor with the Department of CEECS and a Faculty Fellow of I-SENSE, Florida Atlantic University (FAU), where he is also the Director of the Intelligent and Resilient Systems Research Group. His research interests include computational intelligence including machine learning, networked data mining, and cyber-physical systems, including the IoT, smart grids, and critical infrastructure systems.

SHENGBO JIANG was born in Weifang, China, in 1994. He received the B.S. degree from the School of Electrical Engineering, Wuhan University, in 2016. His research interests include security analysis and the control of cascading failure, and the applications of machine learning in power systems.

PEIYUAN XIE received the Ph.D. degree from the School of Electrical Engineering, Huazhong University of Science and Technology. His research interests include security analysis and the control of power systems.

WEN YUAN received the Ph.D. degree from the School of Electrical Engineering, Huazhong University of Science and Technology. His research interests include security analysis and the control of power systems.