

## Spatial-temporal adaptive transient stability assessment for power system under missing data



Bendong Tan<sup>a</sup>, Jun Yang<sup>a,\*</sup>, Ting Zhou<sup>a</sup>, Xiangpeng Zhan<sup>a</sup>, Yuan Liu<sup>b</sup>, Shengbo Jiang<sup>c</sup>, Chao Luo<sup>d</sup>

<sup>a</sup> School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China

<sup>b</sup> Jiangxi Electric Power Construction Branch Office of State Grid, Nanchang 330000, China

<sup>c</sup> State Grid Qingdao Electric Power Company, Qingdao 266000, China

<sup>d</sup> Central Southern China Electric Power Design Institute, Wuhan 430071, China

### ARTICLE INFO

#### Keywords:

Transient stability assessment  
Phasor measurement units  
Machine learning  
Optimal PMU clusters searching model  
Ensemble mechanism

### ABSTRACT

Transient stability assessment (TSA) plays an important role in the design and operation of power system. With the widespread deployment of phasor measurement units (PMUs), the machine learning-based method has attracted much attention for its speed and generalization. However, the generalization will deteriorate if some features are missing due to PMU failure. In this paper, a spatial-temporal adaptive TSA method is proposed to handle the missing data issue. By developing an optimal PMU clusters searching model based on temporal feature importance, and by constructing an ensemble mechanism of long short-term memory (LSTM) for the optimal PMU clusters, the spatial-temporal information is utilized adaptively. Therefore, the aim of maintaining the robustness of TSA performance under any possible PMU failure event is achieved. The proposed approach is demonstrated on New England 39-bus power system. Compared with existing methods, the proposed method achieves state-of-art performance in both accuracy and response time under missing data conditions. In addition, the proposed method is more robust in the case of PMU failure than others.

### 1. Introduction

Transient stability refers to the ability of power system to remain synchronous when subjected to a severe disturbance [1]. The integration of intermittent renewable energy and changes of the grid structure are significantly expanding the state space of power system operation. Electricity market liberalization reduces the controllability of the dispatcher for the power system as well [2]. Consequently, the operation of the power system is getting closer to its stability limit, and power blackouts are more likely to occur in case of a large disturbance [3]. To prevent the economic loss and social impact caused by losing transient stability, real-time transient stability assessment (TSA) is essentially implemented to leave enough time for post contingency remedial actions.

To achieve rapid and accurate TSA, time domain simulation (TDS) is a classical method to assess the stability of power system by solving a set of high dimensional differential algebraic equations [4]. Transient energy function (TEF) [5–8] is also employed to predict the post-fault stability by comparing the acquired transient energy at fault clearing instant with the maximum potential energy that can be assimilated by the power system. However, the speed and accuracy of those methods

cannot meet the requirement of TSA.

With the development of synchronous measurement technology, the phasor measurement unit (PMU) has been applied successfully in the supervisory control and data acquisition (SCADA) system due to its capability for obtaining real-time measurements of the power system. By using PMU, many data-driven TSA methods are proposed to capture the post-fault dynamics behavior of power system [3]. Based on dissipation theory in dynamic system, Lyapunov exponent (LE) [9] refers to the exponential rate of convergence or separation between adjacent trajectories in the operation space, and the maximal Lyapunov exponent (MLE) [10] can be used to quantitatively describe the short-term stability of power system to detect the synchronism of generators. MLE estimated by the recursive least-squares-based method is utilized to assess the angle stability of the power system in a model-free way [11], but it needs a large observation window to obtain reliable assessment results. There are also methods using trajectory prediction to assess transient stability. In [12], generator rotor speed-acceleration trajectories from PMUs are used to predict and identify out-of-step status. [13] utilizes linear autoregressive with exogenous input model to perform dynamic response estimation, which can be used as an effective instability indicator. [14,15] propose an improved grey model to

\* Corresponding author.

E-mail address: [JYang@whu.edu.cn](mailto:JYang@whu.edu.cn) (J. Yang).

predict rotor angle dynamic trajectories so that transient stability can be detected earlier than traditional TDS-based methods. However, trajectory prediction can only perform accurate prediction in limited time interval which cannot meet the requirement of power system transient stability.

In addition, researches on artificial intelligence (AI) and big data analysis technology are developing rapidly. Machine learning-based methods (MLBM) for TSA present promising potential of application. Once measurements are obtained from PMUs, the post-fault power system stability can be assessed online by the machine learning model trained in an offline manner, where inputs are selected or extracted features. Therefore, MLBM is able to implement TSA in a real-time manner to promote awareness of transient stability for the power system and leave enough time to perform remedial control. Related studies have shown the superiority of real-time TSA [16]. The support vector machine (SVM) and its variants [17–20], such as the core vector machine (CVM) [20], are employed to perform TSA and demonstrate the excellent generalization ability and scalability of SVM/CVM to predict the stability status of practical power system. Decision tree (DT) has been widely applied to TSA, and DT-based rules are mined for the out-of-step prediction of synchronous generators [21]. Refs. [22,23] present probabilistic frameworks for TSA of power system by considering renewable generation with DT. Neural network (NN) is also utilized to construct the TSA system with the rapid development of deep learning [24,25]. [25] develops an end-to-end TSA model to simplify the feature engineering process and a time-adaptive method is used to implement transient stability prediction in [4,26]. The mapping relationship between transient stability status and input features can be built by training the above machine learning classifiers. Therefore, transient stability status can be predicted once the constructed features are obtained.

In practice, PMU failure, phasor data concentrator (PDC) failure and communication delay may cause missing data in the power system [27]. In this case, the performance of MLBM to implement TSA may deteriorate drastically owing to the lack of input features. On the topic of TSA with missing data, considerable efforts have been made by using various techniques to deal with missing data. One feasible way is to recover the missing data by learning the system model and real-time measurements [28]. However, this strategy suffers computation burdens, and it may not meet the required assessment speed of online TSA. In [29], the DT with surrogate spilt (DTSS) method is proposed to handle the unavailable measurement of PMUs through impairing missing ones with highly correlated collocated features, which could lead to the deteriorated accuracy for TSA possibly as the number of missing data increases. In [30–32], an observability-constrained feature subsets algorithm is designed, and a PMU clustering-based ensemble classification model is then utilized to sustain the performance of stability assessment considering different missing-data scenarios. However, the above process is used to cluster PMUs under complete observability (with all features) of the power system, and it is not concerned with the importance of PMU, which is unreasonable to some extent. In addition, the temporal information is not utilized to further improve the accuracy and response time in the above references.

To overcome the drawback of the above existing researches, a spatial-temporal adaptive TSA to handle missing data is proposed in this paper. The main contributions of this paper are as follows.

- (1) A temporal feature importance calculation algorithm is designed to perform quantitative analysis for the capability of temporal features distinguishing classes in TSA, and the results of the algorithm is further utilized to find the critical PMU clusters.
- (2) The optimal PMU clusters searching model is constructed to obtain observability-constrained feature subsets in this work. The model maintains high generalization for TSA learning, and the risk of missing data is minimized by capturing the critical PMU clusters.
- (3) An ensemble mechanism of LSTMs is designed to integrate the TSA

classifiers with weighted averaging and implement spatial-temporal TSA online adaptively, so it can maintain great robustness to missing data.

- (4) The proposed method attempts to employ spatial and temporal data to make use of the remaining information provided by PMU measurements adaptively after PMU failure occurs.

The rest of this paper is organized as follows. Section 2 builds the optimal PMU clusters searching model based on the importance of temporal features. The spatial-temporal adaptive TSA method is presented in Section 3 in detail. Section 4 demonstrates the effectiveness of the proposed method on the New England 39-bus power system. Conclusions are drawn in Section 5.

## 2. Optimal PMU clustering

To find the critical PMU clusters to minimize the risk under any possible PMU failure, the approach of temporal feature importance calculation is first introduced, and an optimal PMU cluster-searching model is then built to retain the maximum expectation of feature importance under different PMU failure scenarios in this section.

### 2.1. Importance distribution of PMUs

The PMU is a kind of device to measure the alternating current (AC) voltages and power flow in the branch synchronously with the common time reference that is provided by global positioning system (GPS) [33]. With such a device installed at a specific bus, the voltage of that bus and the power flow of all branches connected to it are measured. Then, the observability of that bus and its incident buses can be confirmed. To obtain complete observability of the power system, several PMUs need to be installed at specific buses, which has been studied in many works such as optimal PMU placement with minimal cost [34].

In the dynamic process of the power system after suffering disturbance, some of generator rotors that are out-of-step cause transient instability. The operation condition is defined by a set of system variables, or features, e.g., active and reactive power of generator, bus magnitude and angle, power flow of branch. Therefore, some features are relatively more important for TSA, which are able to represent the dynamic characteristic of the power system. In other words, there are some critical features in the power system. Consequently, this characteristic can be described quantitatively by the importance of these features. Since the physical measurements such as bus voltage magnitudes and angles that each PMU can acquire are different, the feature importance of each PMU is also distinct for TSA.

In recent studies of TSA, feature importance is calculated to select critical feature subsets that are necessary and sufficient to describe transient stability, which reduces the dimensionality of the inputs and enhances generalization by avoiding overfitting. The traditional TSA generally assesses the power system measurements obtained at a certain instant after a fault in an online manner, so the existing feature importance calculation algorithm that can only handle vectors is not suitable for data in the form of a matrix, such as multivariate time series. However, inputs of the model in this paper is post-fault multivariate time series since the theme is time-adaptive TSA. If feature importance is calculated separately for each moment, it will not reflect the overall characteristics of features during the dynamic changes after fault. Therefore, a temporal feature importance calculation method for multivariate time series of TSA is proposed in this paper.

A popular feature importance calculation algorithm applied in the power system is Relief-F [35], which estimates the ability of each feature to distinguish samples with different classes by assigning weights. Relief-F not only considers the correlation between feature values and target but also the difference between samples on specific features. The difference can be defined as:

$$\text{diff}(a, R_1, R_2) = \frac{|R_1[a] - R_2[a]|}{\max(a) - \min(a)} \quad (1)$$

where  $R_1$  and  $R_2$  are different samples in the database respectively, and  $a$  is one of the features. Obviously, Relief-F cannot be applied if the feature  $a$  is time series, so the temporal feature importance of the time series calculation method by improving Relief-F is presented in this paper, which is later called Relief-FT. Considering the different transient characteristics of various disturbances, such features as voltage would be very low after unstable short-circuit fault while it can restore to a relatively high level after stable short-circuit fault, so the formula of difference between samples on the time series feature  $a$  can be represented by Euclidean distance:

$$\text{diff}(a_s, R_1, R_2) = \frac{\|R_1[a_s] - R_2[a_s]\|}{\max(a_s) - \min(a_s)} \quad (2)$$

where  $a_s$  is in the format of sequence for the feature  $a$ . The concrete theoretic of the relief family can be seen in [35] and the pseudo code of Relief-FT is shown in Algorithm 1, where  $m$  is a user-defined parameter to decide the repeated times for the whole process, and  $I(C)$  is the ratio of class  $C$  in all samples.

---

**Algorithm 1** Relief-FT feature importance calculation algorithm

---

**Input:** For each training sample  $x \in R$ , a vector of features values  $A$  and length of observation window  $T$ .

**Output:** A vector  $W \in \mathbb{R}^{|A|}$  of estimation of importance of features  $a \in A$ .

1. set all  $W$  to 0;
  2. for  $i = 1: m$
  3. randomly select a sample  $x_i$ ;
  4. for each class  $C = \text{class}(x_i)$
  5. from class  $C$  select  $k$  nearest hits  $H$ ;
  6. end
  7. for each class  $C \neq \text{class}(x_i)$
  8. from class  $C$  select  $k$  nearest misses  $M$ ;
  9. end
  10. for each  $a$  in  $A$
  11. 
$$W[a] = W[a] - \sum_{j=1}^k \frac{\text{diff}(a, x_i, H_j)}{m-k} + \sum_{C \neq \text{class}(x_i)} \frac{I(C)}{1-I(C)} \sum_{j=1}^k \frac{\text{diff}(a, x_i, M_j)}{m-k} \quad (3)$$
  12. end
  13. end
- 

In Relief-FT, each feature is assigned with weight in the range of  $[-1, 1]$ , which is the so-called importance to represent the ability to distinguish samples with different targets. A positive value of the weight indicates that the feature is very beneficial for classifying different patterns in TSA, while the negative value of the weight means the feature overlaps the samples.

## 2.2. Optimal PMU clusters searching model

In recent work [30–32], the ensemble data-analytics model with feature subsets sampling method is proposed to construct feasible inputs for PMU-based dynamic stability assessment considering incomplete measurement data. Under any PMU loss scenario, the ensemble model is able to maintain robustness against data loss for TSA with minimal number of PMUs. However, existing researches do not consider the importance of the PMU clusters in power systems to further decrease the risk of lost PMUs. On the other hand, while data loss occurs, temporal information can't be effectively utilized to circumvent such incidents in [30–32].

In this section, an optimal PMU clusters searching model is constructed not only to reduce the number of PMU clusters but also to capture the critical PMU clusters to minimize the risk from missing data. The power system observability description algorithm is first introduced to build observability constraints of the optimal model, and it can be accomplished by iterating the following observability rules [34]:

- (1) Rule 1: The voltage and branch currents of the bus where the PMU is installed are observable, and its incident bus is also observable.
- (2) Rule 2: The voltages at both ends of the branch are observable, and the current of the branch is observable.
- (3) Rule 3: If the branch current and voltage at one end are known, then the voltage at the other end of the branch is observable.
- (4) Rule 4: If ZIB is not installed with PMU, and only one incident branch current is unknown, then the branch current is observable.
- (5) Rule 5: If ZIB is not installed with PMU, and voltages of all adjacent buses are known, then the voltage of ZIB is observable.

When the observability status of each bus is unchanged, the iteration is finished. Once the placement of PMUs in the power system is given, the observability of each bus can be determined.

As introduced previously, MLBM will fail if some features of the inputs are missing. Optimal PMU clusters are necessary to be searched to integrate the MLBM models to reduce the impact with several inputs that are composed of these PMU clusters. For a power system with  $N$  PMUs,  $2^N - 1$  possible PMU combinations can be found. In such situation,  $M$  PMU clusters are assumed to be searched:

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1M} \\ q_{21} & q_{22} & \cdots & q_{2M} \\ \vdots & \vdots & q_{lk} & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nM} \end{bmatrix} = [q^{(1)}, q^{(2)}, \dots, q^{(M)}] \quad (4)$$

where  $n$  is the number of buses,  $q_{lk}$  is 1 if PMU is installed at bus  $l$  in  $k$ -th PMU cluster. Otherwise,  $q_{lk}$  equals to 0,  $l = 1, 2, \dots, n$ ,  $k = 1, 2, \dots, M$ .

In  $2^N - 1$  possible PMU combinations, the probability of multiple PMU failures occurring at the same time is very small. Therefore, the risk for TSA with  $M$  PMU clusters can be defined as the expectation of the ratio of feature importance missing by failure in arbitrary PMU combination. To construct ensemble MLBM with low risk for TSA, the objective function of the optimal PMU clusters searching model can be formulated as follows:

$$\min - \frac{\sum_{k=1}^M \sum_{i=1}^N \sum_{j=1}^{C_N^i} P^j \cdot \beta(\sigma_{i,j}, \psi^{(k)}) \cdot (\sum W_{\psi^{(k)}})}{\sum_{k=1}^M \sum_{i=1}^N \sum_{j=1}^{C_N^i} P^j \cdot (\sum W_{\psi^{(k)}})} \quad (5)$$

where  $W_{\psi^{(k)}}$  is the importance set of observable temporal features that consist of bus voltage magnitude, bus voltage angle, active power of branch, reactive power of branch, active power of load, reactive power of load, active power of generator and reactive power of generator in the  $k$ -th PMU cluster.  $P^j$  is the unavailability rate of  $j$ -th combination of PMUs.  $C_N^i$  is the formula of permutation.  $\sigma_{i,j}$  is a set vector that represents the  $j$ -th combinations with  $i$  PMUs.  $\psi^{(k)}$  is a set vector showing PMU locations in the  $k$ -th PMU cluster.  $\beta$  is an indication function that is formulated as:

$$\beta(x, y) = \begin{cases} 1 & x \cap y = \emptyset \\ 0 & x \cap y \neq \emptyset \end{cases} \quad (6)$$

For the optimal PMU clusters searching model, the only constraint is that union observability of all the PMU clusters can maintain full observability of the whole buses in the power system, which can be written as:

$$s. t. \begin{cases} f(q^{(1)}) + f(q^{(2)}) + \cdots + f(q^{(M)}) \geq 1 \\ 0 \text{ else} \\ 1 \text{ if bus } i \text{ is observable in } k\text{-th PMU cluster} \\ f(q^{(k)}) = V^{(k)}, k = 1, 2, \dots, M \end{cases} \quad (7)$$

where  $q^{(k)} = [q_{1k}, q_{2k}, \dots, q_{nk}]^T$  is a set vector that indicates PMU locations of the  $k$ -th PMU cluster,  $V^{(k)} = [V_{1k}, V_{2k}, \dots, V_{nk}]^T$  is a set vector that shows the observability of all the buses in the  $k$ -th PMU cluster and  $f(\cdot)$  is a function iterating the power system observability description

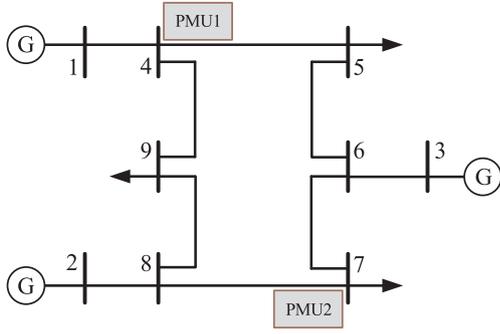


Fig. 1. IEEE 9-bus power system.

algorithm to obtain observable features under certain PMU clusters. Taking IEEE 9-bus power system as an illustration for the optimal PMU clusters searching model, as shown in Fig. 1. PMU1 and PMU2 installed at bus 4 and bus 7 respectively can get full observability for the system [34]. In the objective function, it is assumed that optimal PMU clusters  $\psi$  is  $\{\{4\},\{7\},\{4,7\}\}$ , so  $W_{\psi^{(1)}}$  is the sum of feature importance in PMU set  $\psi_1 = \{4\}$ ,  $W_{\psi^{(2)}}$  is the sum of feature importance in PMU set  $\psi_2 = \{7\}$  and  $W_{\psi^{(3)}}$  is the sum of feature importance in PMU set  $\psi_3 = \{4, 7\}$ . The unavailability rate of different PMU combinations can be denoted as  $p^1 = p^{PMU1}$ ,  $p^2 = p^{PMU2}$  and  $p^3 = p^{PMU1}p^{PMU2}$  respectively. In addition,  $\sigma_{1,1}$  is  $\{4\}$ ,  $\sigma_{2,1}$  is  $\{7\}$  and  $\sigma_{1,2}$  is  $\{4,7\}$ . In constrains, PMU clusters matrix  $Q$  can be shown as:

$$Q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (8)$$

With the above illustrations, it can be concluded that formulation (5) presents the sum of the negative ratio of available feature importance under different PMU failures. Therefore, the minimization of the objective function can represent the goal to decrease the risk of PMU failures.

In the proposed optimal PMU clusters searching model, objective function and constraints are nonlinear and not derivable, so the genetic algorithm (GA) is employed to search the optimal solutions in this paper. When the optimal PMU clusters are searched, each PMU cluster provides its observable features (voltage magnitude, voltage angle, et al) as inputs for each LSTM, which will be introduced in the next section, respectively.

### 3. Temporal adaptive ensemble TSA

The PMU failure is assumed to be permanent, and it happens at every instant with the same probability after the fault clearance in this paper. Based on this assumption, in case some PMUs fail, the remaining PMU clusters without missing data are still viable for TSA while the temporal data before the time PMU failure occurrence is also available. Therefore, an ensemble mechanism of LSTM is designed to make use of the remaining spatial-temporal data that are not missing adaptively to sustain the performance of accuracy and response time under any possible PMU failure event.

#### 3.1. Introduction to the LSTM Algorithm

LSTM is a variant of ANN, as shown in Fig. 2, which is designed to construct a memory cell that can keep long term memory to maintain long temporal dependence in the time domain [4]. The memory cell

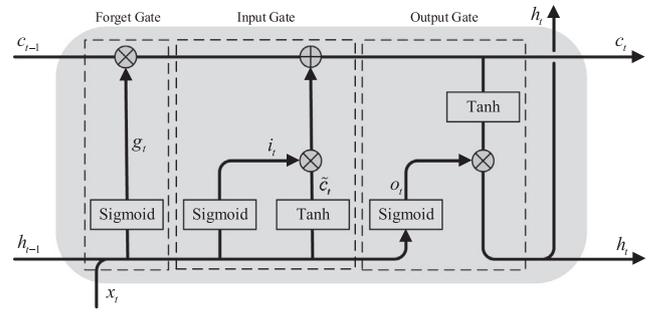


Fig. 2. The structure of LSTM.

consists of forget gate, input gate and output gate. These gates handle historical information, input data and output information, respectively. The memory cell of LSTM merges historical information and current data to implement feature extraction. In this way, the memory state with a specific timestamp  $c_t$  can be acquired and passed to the next time slot. The expression in mathematical form is presented as [4]:

$$g_t = \text{Sigmoid}(W_g x_t + U_g h_{t-1} + b_g) \quad (9)$$

$$i_t = \text{Sigmoid}(W_i x_t + U_i h_{t-1} + b_i) \quad (10)$$

$$\tilde{c}_t = \text{Sigmoid}(W_{\tilde{c}} x_t + U_{\tilde{c}} h_{t-1} + b_{\tilde{c}}) \quad (11)$$

$$c_t = g_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (12)$$

Current output  $h_t$  can be calculated with  $c_t$ :

$$o_t = \text{Sigmoid}(W_o x_t + U_o h_{t-1} + b_o) \quad (13)$$

$$h_t = o_t \odot \text{Tanh}(c_t) \quad (14)$$

where  $\text{Sigmoid}(x) = \frac{1}{1 + \exp(-x)}$ ,  $\text{Tanh} = \frac{1 - \exp(-x)}{1 + \exp(-x)}$ , and  $W_*$ ,  $U_*$  and  $b_*$  are the hyperparameters of the corresponding gate.

LSTM can be unrolled in time steps, as shown in Fig. 3. When time-stamped data  $X = [x_1, x_2, \dots, x_T]$  come, LSTM implements feature extraction for  $X$  to get hidden layer output  $H = [h_1, h_2, \dots, h_T]$  at every time step, and  $H$  is then transformed to  $Y = [y_1, y_2, \dots, y_T]$  in the range  $[0, 1]$  by sigmoid, where  $T$  is the length of the observation window. When LSTM is applied in TSA online, it is able to perform accurate assessment step by step through combining with historical and current information. For instance,  $h_1$  is calculated with  $x_1$ , and  $h_2$  is obtained with  $x_1, x_2$ , etc.

Therefore, even if most of the PMU data is missing, the power system transient stability can be assessed at the next instant. It is possible to reduce the impacts of missing data caused by PMU failure or communication delay via taking advantage of the temporal characteristics of LSTM, while the optimal PMU clusters model decreases the impacts of spatial characteristics in PMU placement.

#### 3.2. Offline training

##### 3.2.1. A single LSTM training

Before integrating LSTMs, each LSTM should be trained to acquire good performance in the generalization subspace so that the advantages of multiple LSTMs can be combined to perform reliable and rapid TSA.

In this paper, datasets, whose features consist of bus voltage magnitude, bus voltage angle, active power of branch, reactive power of

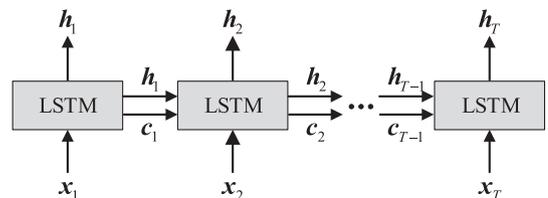


Fig. 3. LSTM unrolled in time steps.

branch, active power of load, reactive power of load, active power of generator and reactive power of generator since they present obvious transient characteristics after disturbance [20,25,36], are generated by TDS with a variety of different contingencies on the given power system. To obtain high generalization for LSTMs, datasets are randomly divided into training data, validating data and testing data at a certain ratio. According to the searched PMU clusters, training data, validating data and testing data are divided into several feature subsets respectively. It should be noticed that the number of the features subsets equals to the number of searched PMU clusters, so the same number of LSTMs shall be trained. To deploy each LSTM to TSA applications, it is necessary to train each LSTM with corresponding training data subset including certain features. For MLBM of TSA belongs to supervised learning, every sample in the datasets is calibrated with a stability label according to its transient stability status. The label for each sample in TSA can be defined as:

$$y = \begin{cases} 1 & \eta \leq 0 \\ 0 & \eta > 0 \end{cases} \quad (15)$$

where  $\eta = \frac{360 - \delta_{\max}}{360 + \delta_{\max}}$  [4], and  $\delta_{\max}$  is the maximum difference between angles of any two generators at the end of the post-fault power system simulation.

Given datasets  $\Delta$  with  $F$  training data, optimal  $W_*$ ,  $U_*$  and  $b_*$  of each LSTM can be found with the Adam optimizer [38]. The objective loss function of  $i$ -th LSTM is formulated as [4]:

$$\min - \sum_{k=1}^F [y^{(k)} \log \hat{y}_i^{(k)} + (1 - y_i^{(k)}) \log(1 - \hat{y}_i^{(k)})] \quad (16)$$

where  $y^{(k)}$  is the real label of  $k$ -th sample in training data, and  $\hat{y}_i^{(k)}$  is the assessment result of  $i$ -th LSTM for  $k$ -th sample in training data at last instant of simulation. The objective loss function is binary cross-entropy, which can evaluate the performance of the  $i$ -th LSTM in the training procedure.

### 3.2.2. Ensemble Mechanism for LSTM

In the supervised learning algorithm of machine learning, the goal is to learn a stable classifier (such as a single LSTM) that performs well in all scenarios, but the actual situation is often not so ideal. Sometimes it can only get a classifier with preferences (weak classifiers perform better only in few cases). Therefore, ensemble learning is employed to combine multiple weak classifiers to get a better and more comprehensively strong classifier. For an ensemble model with excellent generalization, the key point is the high diversity both in data and the structure of weak classifiers.

It is assumed that there are  $M$  PMU clusters obtained by the optimal PMU clusters searching model. Then,  $M$  LSTMs are needed to construct the ensemble model. To keep high diversity in the data,  $M$  training data subsets are sampled from the training data while training features in every training data subset are determined by observable features of each PMU cluster. The dropout technique [37], which randomly drops out some neurons of LSTM hidden layers, is employed in learning structure diversity. Therefore, diversity in data and structure for ensemble learning are both achieved. In a word, the proposed ensemble LSTM has the capability to merge multiple PMU clusters without losing generalization or comprehensiveness.

As  $M$  diverse LSTM classifiers have been trained, how to integrate them to make credible TSA is the critical step of the ensemble mechanism. At present, the common way of researches in the ensemble model to merge multiple classifiers is average integration or majority voting. However, since observable features of each PMU cluster are different, the performance of each LSTM classifier is distinct. As a result, the generalization of ensemble LSTM will be damaged just by way of average integration or majority voting. In this paper, an ensemble method based on weighted averaging is proposed to integrate LSTMs. Weight is assigned to each LSTM offline, and the final assessment result

of ensemble LSTM is the weighted averaging of all the LSTM outputs, which is formulated as:

$$\hat{y}_t^* = \frac{\sum_{i=1}^M \omega_i \hat{y}_{i,t}}{\sum_{i=1}^M \omega_i}, \quad t = 1, 2, \dots, T \quad (17)$$

where  $\omega_i$  is the weight of the  $i$ -th LSTM, it can be calculated by the reciprocal of validation error that is acquired by applying validating data to each trained LSTM respectively. The  $\hat{y}_{i,t}$  is the assessment result of the  $i$ -th LSTM for TSA at time  $t$ .

### 3.3. Online spatial-temporal adaptive TSA

A time-adaptive TSA method is proposed in [4] to combine historical information with current information to generate a reliable assessment as soon as possible, and it is further utilized to handle the TSA problem with time-delayed synchronized phasors in [27]. Specifically, once the fault is cleared, the stability of power system can be determined, so the violent dynamics will be gradually damped or just enhanced. If the angular difference can be damped, the later the time after fault clearance is, the less violent the dynamic behavior is, and the confidence of the model to the assessment result becomes greater. Otherwise, in the unstable scenario, the dynamic behavior would be more drastic over time after fault clearance and the assessment result of the model would be more credible. To obtain fast assessment speed, a time-adaptive mechanism is designed, and the judgement can be made if the assessment result is credible enough. However, once some features of inputs remain missing along the timeline, the performance of the time-adaptive model could deteriorate.

In this section, a spatial-temporal adaptive TSA method is proposed to reduce the impacts on reliability of assessment under missing data, which includes PMU failures and communication delay, et al. Specifically, spatial information, which is represented by PMU clusters, and temporal information are employed to make use of the remaining information in the available PMUs to assess stability status rapidly. Consequently, remedial action can be conducted in a timely manner to maintain the security of the power system.

As Fig. 4 shows, the online spatial-temporal adaptive TSA can be carried out when offline training is finished, where  $x_1$ ,  $x_{t_f}$ ,  $x_T$  are transient features at time 1,  $t_f$ ,  $T$  respectively after fault clearance. It is assumed that the PMU failure happens at time  $t_f$  after the fault is cleared. If  $t_f = 1$ , the remaining available LSTMs, which are determined with missing data, will be integrated to assess the stability of the power system reliably. If  $t_f > 1$ , the decision will be made by fusing assessment results at  $t$  of remaining available LSTMs and assessment results at  $t_f - 1$  of affected LSTMs. Therefore, the method of integration is rewritten as:

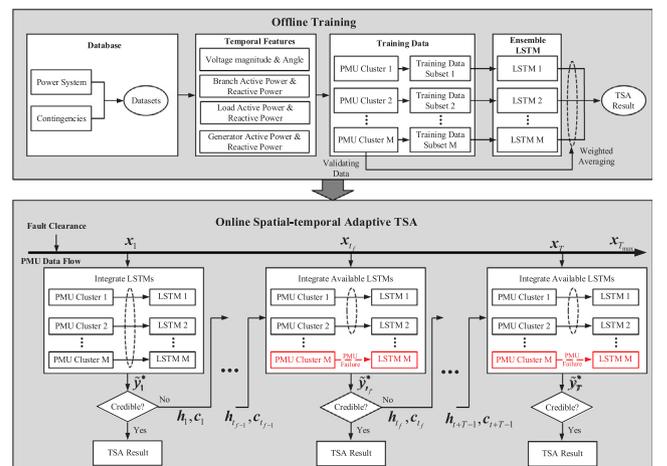


Fig. 4. Training and online application of spatial-temporal adaptive TSA.

$$\tilde{y}_t^* = \begin{cases} \frac{\sum_{i \in \Omega} \omega_i \tilde{y}_{i,t} + \sum_{j \notin \Omega} \omega_j \tilde{y}_{j,t_f-1}}{\sum_{i=1}^M \omega_i} t_f > 1 \\ \frac{\sum_{i \in \Omega} \omega_i \tilde{y}_{i,t}}{\sum_{i \in \Omega} \omega_i} t_f = 1 \end{cases} \quad (18)$$

where  $\Omega$  is the set of remaining available LSTMs under missing data event,  $\tilde{y}_{j,t_f-1}$  is assessment result of the  $j$ -th LSTM for TSA at time  $t_f - 1$  and  $\tilde{y}_{i,t}$  is assessment result of the  $i$ -th LSTM for TSA at time  $t$ . From formula (18), if  $t_f > 1$ , it can be seen that the final TSA result  $\tilde{y}_t^*$  is weighted averaging of assessment results at time  $t_f - 1$  of the unavailable LSTMs affected by PMU failure and assessment results at time  $t$  of available LSTMs. In that way, the spatial-temporal information can be efficiently utilized since the information of affected LSTMs is integrated into unaffected LSTMs adaptively to perform TSA, while existing methods can only make use of unaffected predictors at a fixed observation window.

To acquire a credible TSA result, the stability index can be defined as [4]:

$$\text{Stability} = \begin{cases} \text{Unstable if } 1 \geq \tilde{y}_t^* \geq \delta \\ \text{Unknown if } \delta > \tilde{y}_t^* > 1 - \delta, t = 1, 2, \dots, T \\ \text{Stable if } 1 - \delta \geq \tilde{y}_t^* \geq 0 \end{cases} \quad (19)$$

where  $\delta$  is the stable threshold that needs to be determined by parameter searching to balance the speed and accuracy of TSA. In the adaptive procedure, if the stability index is ‘‘Stable’’ or ‘‘Unstable’’ at time  $t$ , the spatial-temporal TSA result  $\tilde{y}_t^*$  is regarded as credible to stop the assessment manner. If the stability index is ‘‘Unknown’’, the historical information  $h_t$  and  $c_t$  will be transferred to available LSTMs at the next instant and combined with  $x_{t+1}$  to conduct TSA until the TSA result is credible or the max assessment time  $T_{\max}$  is reached.

#### 4. Simulation results

In this section, the proposed approach is demonstrated with New England 39-bus power system, and all simulations are conducted on a laptop with Intel Core i5-7300HQ CPU, 8 GB RAM and a 1050 2D/3D graphics card with 2 GB memory.

##### 4.1. Data generation

New England 39-bus power system [39], as shown in Fig. 5, is

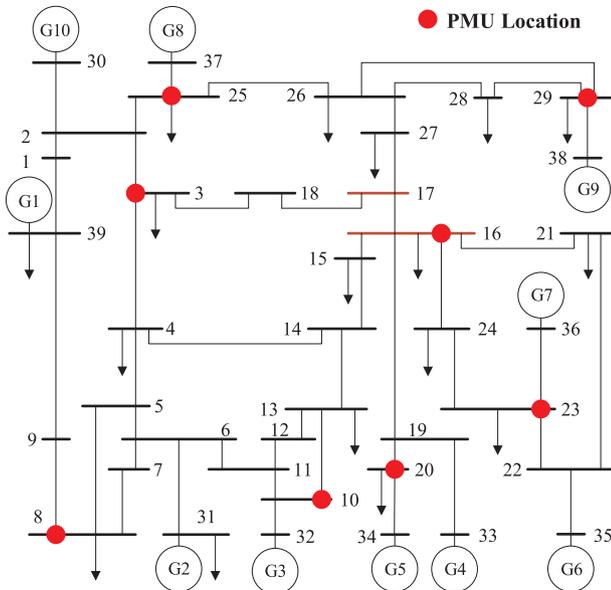


Fig. 5. New England 39-bus power system.

utilized to train and test the spatial-temporal adaptive TSA model. The test system consists of 39 buses, 46 branches, 10 generators and 19 loads. The reference power is 100 MVA, and the reference voltage is 345 kV. The PMUs are placed at bus 3, 8, 10, 16, 20, 23, 25 and 29, which assure full observability with minimum number of PMUs. Parameters of dynamic models are taken from [39] except subtransient parameters of generators from [40], where loads are considered to have a voltage dependency according to Eqs. (21) and (22) and the parameter  $k_{pu}$  is set to 1 (constant current behaviour for active power) and  $k_{qu}$  to 2 (constant impedance behaviour for reactive power).

$$P = P_0 \left( \frac{U}{U_0} \right)^{k_{pu}} \quad (20)$$

$$Q = Q_0 \left( \frac{U}{U_0} \right)^{k_{qu}} \quad (21)$$

where the subscript 0 indicates the initial operating condition.  $P$  is active power,  $Q$  is reactive power and  $U$  is voltage.

Before data generation, the physical facts of input features reflecting how rotor angle stability is affected in time is discussed, so physical insight on the actual dynamic behavior of the system can be further mined. Since the imbalanced active power in generators is the main factor affecting the transient stability, the active power of each generator is plotted in Fig. 6, which shows that the large imbalanced active power of generators leads to instability. Reactive power and voltage also determine the distribution of power flow in power system and thereby has impacts on the imbalanced power, so bus voltage magnitude, bus voltage angle, active power of branch, reactive power of branch, active power of load, reactive power of load, active power of generator and reactive power of generator are considered as input features.

Based on the New England 39-bus power system, batch TDS is performed on PSS/E, which is the commercial power system analysis

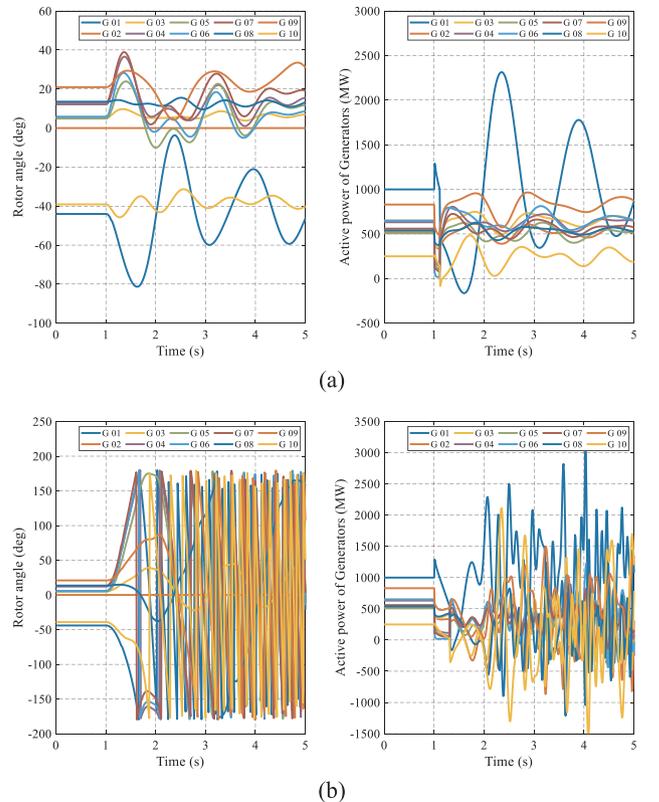


Fig. 6. (a) Dynamic behavior in stable scenario; (b) dynamic behavior in unstable scenario.

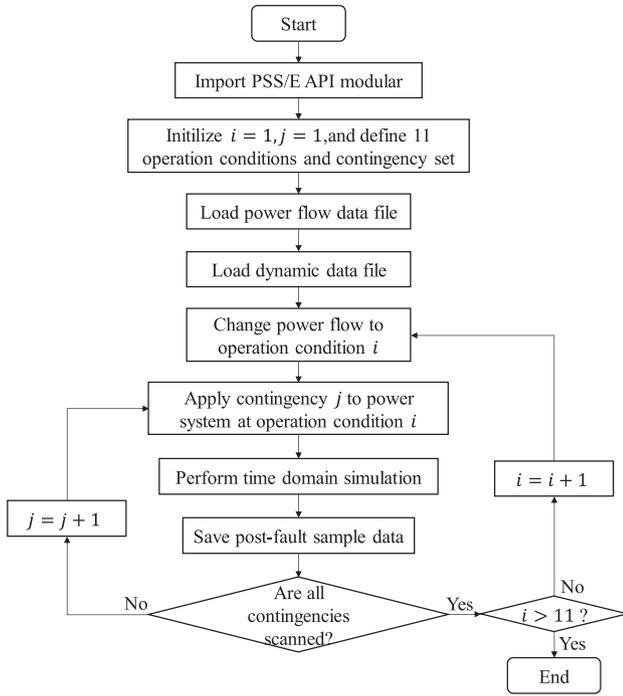


Fig. 7. Flowchart of batch program of time domain simulation.

software, to generate massive reasonable datasets to train the proposed model by considering different operation conditions and contingencies. Consequently, 5775 samples are generated according to the following principles:

- 11 different kinds of load levels (75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 125% of basic load level) are considered, and the powers of the generator are adjusted to the same ratio of basic power as load.
- For each power flow level, three-phase short-circuit fault is applied to every bus and four locations (20%, 40%, 60%, 80% of the length) on the transmission line, respectively.
- In the TDS, the fault duration is set to 0.1 s, 0.3 s or 0.5 s for a specific fault to clear after the disturbance.
- The length of the simulation time is set to 10 s to label the transient stability status accurately.

PSS/E has application programming interface (API) of Python, so we can use Python to perform batch program of time domain simulation with PSS/E. What we need to do is to scan contingency set for each operation condition, and the above process is shown in Fig. 7.

All the samples are divided into training data, validating data and testing data in the ratio of 3:1:1, where the training data are utilized to train the single LSTM for each PMU cluster, the validating data are used to implement weighted averaging for ensemble LSTM, and the performance of the spatial-temporal adaptive TSA model is assessed with testing data.

In the simulation, the main control parameters shall be defined to train the proposed model, and all parameters are selected according to the performance of response time and accuracy of TSA. Two LSTM layers with 100 memory cells in each layer are introduced to build the single LSTM for every PMU cluster. To obtain the structure diversity of the model, the dropout rate for each LSTM layer is set to 0.05. The training epoch is 200, and the input size of each single LSTM is the number of features in the corresponding PMU cluster. Referring to [4], the maximum assessment time  $T_{\max}$  is set to 20. Main parameters of GA, generations and population size, are 40 and 40 respectively, other parameters of GA are assigned with default values in Matlab.

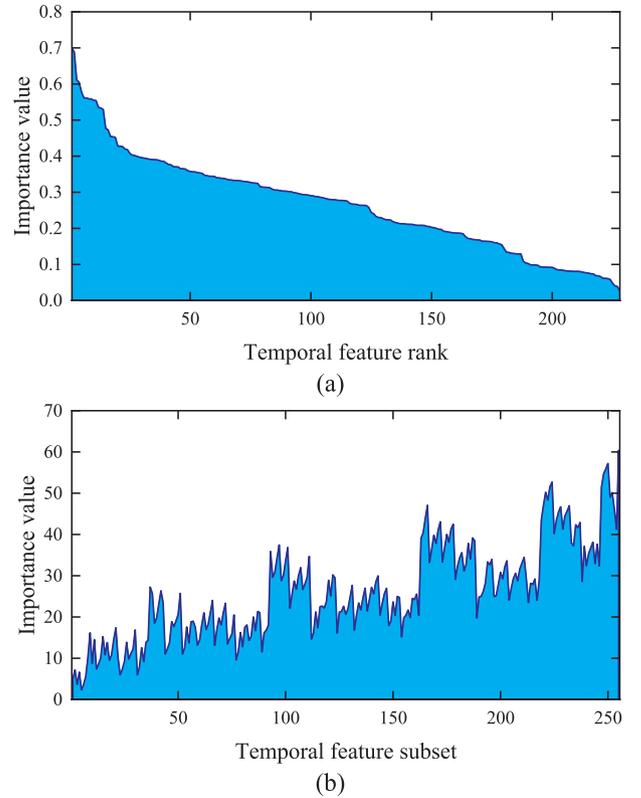


Fig. 8. (a) The importance of each temporal feature; (b) The importance of each temporal feature subset.

## 4.2. Results of optimal PMU clusters searching

### 4.2.1. Distribution of temporal feature importance

Since the optimal PMU clusters are optimized based on the importance of temporal features, the temporal feature with larger importance has a stronger capability to distinguish the different classes. By applying the proposed Relief-FT algorithm, the importance values of all temporal features are estimated as shown in Fig. 8(a) and ranked in descending order. All the temporal features have positive importance values to separate samples among classes.

The temporal feature importance of various PMU clusters is different due to their different observability, and the importance distribution among various PMU clusters is shown in Fig. 8(b). The X-axis numbers temporal feature subsets of different PMU clusters, and the Y-axis shows the sum of the importance of each temporal feature subset. Generally, the more PMUs in the PMU cluster, the larger the sum of their importance. It will lead to more excellent generalization but higher risk when affected by missing data. Therefore, suitable PMU clusters shall be found to maintain generalization with minimum risk.

### 4.2.2. Optimal PMU clusters

In the WAMS of practical power systems, the probability of a single PMU failure event is very small, so the unavailability of a single PMU is assumed to be 0.02 according to the previous evaluation results of the reference [29]. Therefore, the availability of a single PMU is 0.98. To quantitatively analyze the risk of PMU clusters under missing data, the robustness index is defined as the absolute value of the objective function in the optimal PMU clusters searching model proposed in Section 2.

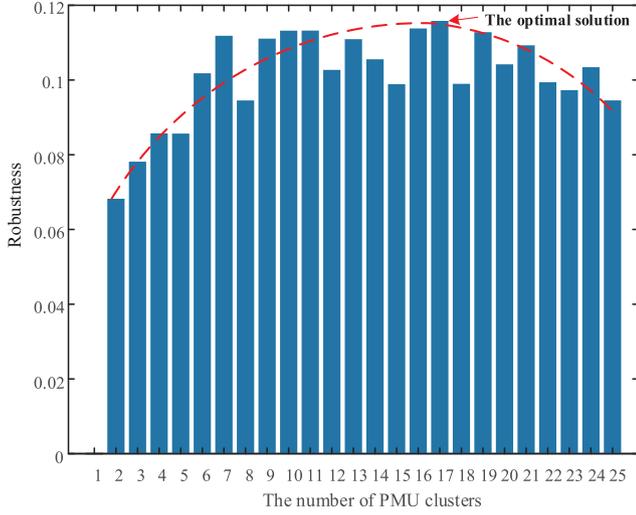


Fig. 9. The relationship between robustness and the number of PMU clusters.

$$Robustness = \frac{\sum_{k=1}^M \sum_{i=1}^N \sum_{j=1}^{C_N^i} P^j \cdot \beta(\sigma_{i,j}, \psi^{(k)}) \cdot (\sum W_{\psi^{(k)}})}{\sum_{k=1}^M \sum_{i=1}^N \sum_{j=1}^{C_N^i} P^j \cdot (\sum W_{\psi^{(k)}})} \quad (22)$$

As Fig. 9 shows, the robustness of the PMU clusters under missing data is the highest when the number of PMU clusters is 17. In other words, the risk of PMU Clusters under missing data is lowest with 17 PMU Clusters. Intuitively, robustness is defined as the sum of the ratio of the temporal feature importance in the available PMU clusters to the sum of the temporal feature importance in all the PMU clusters after data loss occurs. Considering two extreme scenarios: (1) If there is only one PMU cluster with just one input for the model, the performance of the model will be destroyed when PMU failures happen. Then the robustness is zero; (2) If there are too many PMU clusters, some redundancies are inevitable, and the robustness will be reduced. Therefore, the more PMU clusters, the greater the robustness. But the robustness will decrease after the number of PMU clusters reaches a certain amount, and there must be a suitable number of PMU clusters as obtained in this paper.

Consequently, the optimal number of PMU clusters with the highest robustness are chosen to reduce the impacts of missing data. All the PMU clusters are listed in detail in Table 1. Compared with [30] and [31], the number of PMU clusters in this paper is less, which results in reducing the computation consumption, since the proposed method only needs to satisfy the full observability of buses while other methods require the full observability of all features in [30] and [31].

#### 4.3. LSTM ensemble training

Based on the temporal features observed of each PMU cluster obtained by the optimal PMU clusters searching model, every single LSTM in ensemble LSTM is trained separately. The control parameters of each single LSTM training are following as Section 4.1. Finally, 17 LSTMs are trained, and integrated with the respective reciprocal of validation error to form ensemble LSTM.

To make the trade-off between accuracy and speed for TSA, the

Table 1  
Optimal PMU clusters.

Methods	PMU clusters
Proposed method	{8}, {23}, {25}, {3,16}, {3,25}, {8,20}, {8,25}, {10,16}, {10,23}, {16,20}, {16,23}, {20,25}, {25,29}, {3,10,23}, {3,10,25}, {16,23,25}, {3,8,10,25}
Method in [31]	{3}, {8}, {10}, {16}, {20}, {23}, {25}, {29}, {3,8}, {3,16}, {8,25}, {16,20}, {16,23}, {3,8,10}, {3,8,25}, {3,10,16}, {3,16,25}, {3,16,29}, {3,16,25,29}

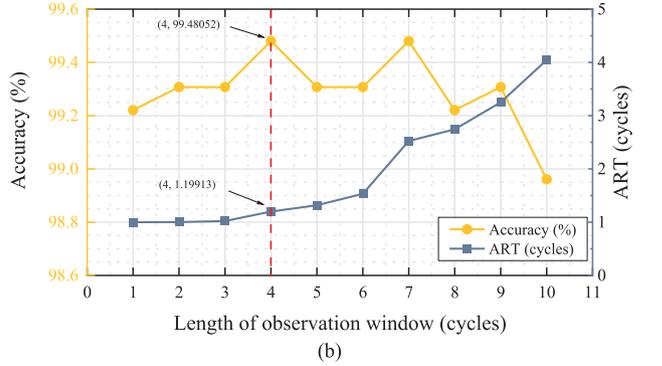
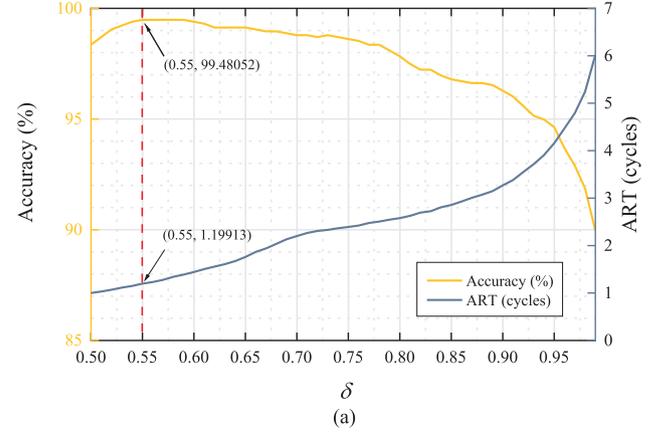


Fig. 10. (a)  $\delta$  Sensitivity analysis; (b)  $T$  sensitivity analysis.

reasonable stability threshold  $\delta$  and the length of the observation window  $T$  shall be defined. The 50  $\delta$  values in the range of [0.5,0.99] with a step size of 0.01 are simulated, the average response time (ART) [26] and accuracy against  $\delta$  are plotted in Fig. 10(a). ART increases with the  $\delta$  values due to the increase of the unknown interval.  $\delta = 0.55$  is preferred to acquire the most rapid assessment speed while keeping the highest TSA accuracy for ensemble LSTM. Fig. 10(b) indicates the relationship between  $T$  and accuracy, ART, respectively. The response time will get longer while the length of the observation window becomes larger, so  $T = 4$  with the highest accuracy is preferred. The accuracy of TSA is not higher with the increase of  $T$  owing to the information loss at the earlier time under larger  $T$ .

#### 4.4. Performance comparison under missing data

To demonstrate the superiority of the proposed method both in response time and accuracy, the performance of existing TSA methods considering PMU failure is compared. For a fair comparison, the length of the observation window is set to  $T = 4$  for all methods, and the same database is utilized to evaluate their performance. They are mean imputation (MI) [42], DTSS, random forest with surrogate split (RFSS) [43] and ensemble random vector functional link (Ensemble RVFL) [30] respectively.

##### 4.4.1. The expectation of response time analysis

The length of the observation window of the existing TSA methods considering PMU failure is fixed due to their mechanisms of receiving

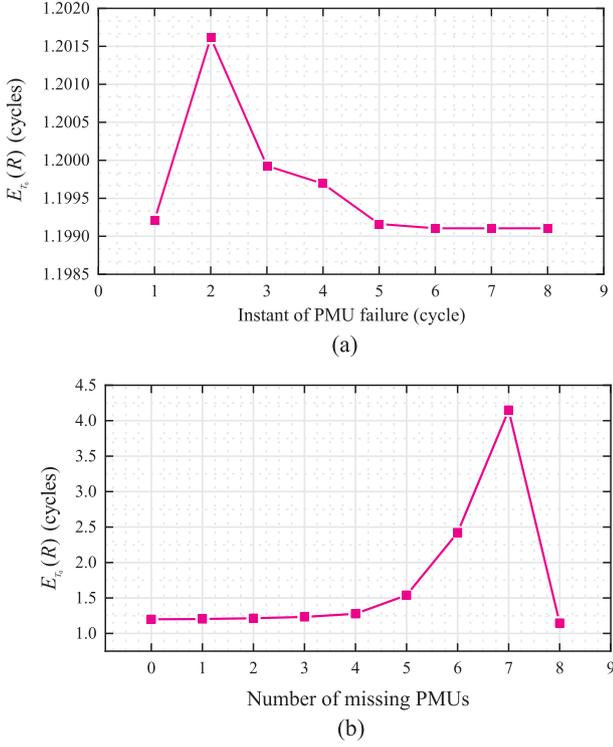


Fig. 11. (a) The expectation of ART with different PMU failure occurrence instant; (b) The expectation of ART with different amount of missing PMUs.

input features. However, the method presented in this paper can be used to assess the power system adaptively in timeline, which means it can reduce the response time of TSA. For the sake of studying the effect of PMU failure on the response time of the proposed method, two response time indexes are defined as:

- The expectation of ART for every timestamp:

$$E_{T_0}^t(R) = \sum_{i=0}^N C_N^i (1-P)^{N-i} P^i \bar{R}_{i,t} \quad (23)$$

- The expectation of ART for different number of missing PMUs:

$$E_{T_0}^i(R) = \frac{1}{T_0} \sum_{t=1}^{T_0} \bar{R}_{i,t} \quad (24)$$

where  $C_N^i$  is the number of possible missing PMU combinations for  $i$  PMUs, and  $\bar{R}_{i,t}$  is average ART for  $i$  missing PMUs if the PMU failure happen at time  $t$ . Besides,  $T_0$  represents the valid length of the assessment time; thus,  $T_0$  of other methods is fixed to 4, while it is 5 for the proposed method to implement adaptive assessment according to simulation result (the response time in Fig. 11(a) after time  $t = 5$  is not change any more).

As Fig. 11(a) shows, while the response time decreases from PMU failure occurrence time  $t = 2$  to  $t = 5$ , it demonstrates a reverse trend with the increase of missing PMUs in Fig. 11(b). Essentially, the response time increases due to the loss of information in the PMU data, but the proposed method can still sustain a quick response time through making use of historical temporal data that are not affected by PMU failure. However, PMU failure occurring at time  $t = 1$  presents a shorter response time compared with time  $t = 2$  to  $t = 4$  since there is no historical information to be utilized to further confirm the stability status of the power system, leading to a quicker decision but lower accuracy. In addition, 8 missing PMUs result in shorter response time because of sacrificing the accuracy (when there is no data for assessment, all scenarios are judged to be unstable).

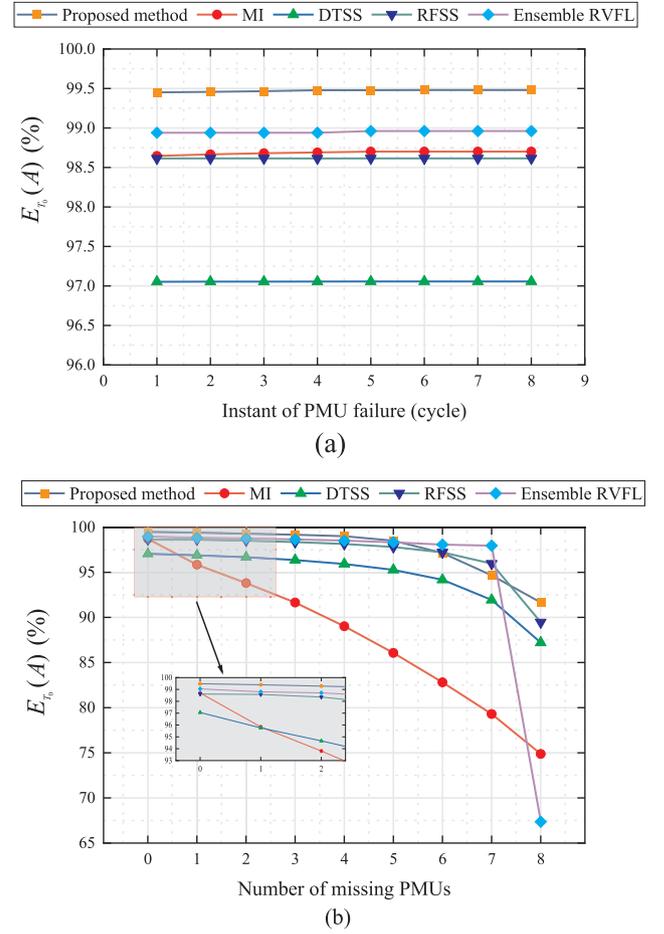


Fig. 12. (a) The expectation of accuracy with different PMU failure occurrence instants; (b) The expectation of accuracy with different amount of missing PMUs.

#### 4.4.2. The expectation of accuracy analysis

Two accuracy indexes are also defined to describe the effect of PMU failure on the accuracy of the proposed method as follows:

- The expectation of accuracy for every timestamp:

$$E_{T_0}^i(A) = \frac{1}{T_0} \sum_{t=1}^{T_0} \bar{A}_{i,t} \quad (25)$$

- The expectation of accuracy for different number of missing PMUs:

$$E_{T_0}^t(A) = \sum_{i=0}^N C_N^i (1-P)^{N-i} P^i \bar{A}_{i,t} \quad (26)$$

where  $\bar{A}_{i,t}$  is the average accuracy for  $i$  missing PMUs if the PMU failure happens at time  $t$ . Fig. 12 presents the expectation of accuracy result against different PMU failure occurrence instants and various numbers of missing PMUs, respectively. Accuracy of all methods is almost not affected by PMU failure occurrence instant in Fig. 12(a), while it shows a downward trend with the increase of missing PMUs in Fig. 12(b). However, the proposed method achieves the highest accuracy until 6 PMUs fail compared with other methods, because it is able to make rapid and credible decisions based on spatial-temporal information adaptively from LSTM and optimal PMU clusters. The proposed method can still sustain the average accuracy of 91.65% even if all PMUs fail, since all samples are assessed to be unstable when there is no any spatial-temporal information to be used to analysis. However, the average accuracy of other methods drops sharply in the same situation.

**Table 2**  
Overall performance comparison.

Method	$E(A)$	$E(R)$
Proposed method	99.47%	1.20 cycles
MI	98.67%	4 cycles
DTSS	97.04%	4 cycles
RFSS	98.61%	4 cycles
Ensemble RVFL	98.94%	4 cycles

4.4.3. Model performance comparison

To compare the overall performance of all methods with missing data, two indexes are defined as:

- The expectation of overall ART:

$$E(R) = \frac{1}{T_0} \sum_{t=1}^{T_0} \sum_{i=0}^N C_N^i (1 - P)^{N-i} P^i \bar{R}_{i,t} \tag{27}$$

- The expectation of overall accuracy:

$$E(A) = \frac{1}{T_0} \sum_{t=1}^{T_0} \sum_{i=0}^N C_N^i (1 - P)^{N-i} P^i \bar{A}_{i,t} \tag{28}$$

As shown in Table 2, the proposed method outperforms other methods both in accuracy  $E(R)$  and response time  $E(A)$  under all possible PMU failure events. Besides, the box plot in Fig. 13 illustrates the accuracy preference distribution of all methods, and the inter-quartile range (IQR) distribution (the red region), which is equal to the difference between the 75th and 25th percentiles, demonstrates that the proposed method is significantly more robust than other methods with missing data as indicated by the smallest IQR. The proposed method is able to obtain more important temporal features in dynamic response by the optimal PMU clusters searching method and it also utilizes LSTM to extract spatial-temporal information deeply. As a result, it can catch efficient transient characteristics and achieve high accuracy. Meanwhile, the proposed method adopts a spatial-temporally adaptive way to assess transient stability, so it can reduce the impacts of PMU failures and make reliable decision earlier. However, other methods, made of shallowed learning machines, handle missing data by simply padding average value (MI), similar feature (DTSS and RFSS), or do not take critical temporal features into consideration (Ensemble RVFL) so they suffer from PMU failures more seriously. And other methods make decision after a fixed observation window, while the proposed method can analysis stability of power system time-adaptively. Therefore, the proposed method is more robust and reliable for rapid and accurate TSA under any possible PMU failure events.

4.4.4. Model performance under unlearned faults

The proposed model is built with training datasets and validation datasets which only cover a certain number of faults, but there may be some potential faults the model does not learn at the online application stage. Since the unlearned faults may have greatly different dynamic distribution compared with learned faults, the test to verify the model robustness to unlearned faults is needed.

In this paper, single faults which are three-phase short-circuit on all buses and lines are considered, so the proposed model has enough capability of assessing these cases. However, multi-fault is not included. Since it is a rare case that several faults happen at the same time, we only consider two faults happening at the same time as an illustration to simplify analyses. If time interval between several faults are too long, we think it can be seen as several individual faults to be assessed separately. As removing bus 16 and bus 17 can divide New England 39-bus power system into two areas, the two bus are considered to be critical buses in dynamic behavior. And it is a very severe-case scenario if the two buses are imposed with faults at the same time. Therefore, bus 16 and bus 17 are applied with three-phase short-circuit fault simultaneously and simulation configuration is same as Section 4.1.

Simulation results are presented in Table 3. From the results, it can be seen that the proposed model achieves higher accuracy (97.96%) and shorter response time (1.191 cycles) than other methods. As for higher accuracy, the proposed model is constructed by PMU clusters based on temporal feature importance and LSTM which can extract spatial-temporal information from power system dynamic behavior, so the proposed model can catch critically common features of faults efficiently though dynamic distribution caused by unlearned faults is greatly different from that caused by learned faults. However, other methods based on shallowed learning machines perform badly under unlearned faults, since they don't have enough ability to extract critically common features between faults or don't take the importance of temporal features in power system into consideration.

On the other hand, the proposed model under faults simultaneously on bus 16 and bus 17 has a shorter response time than under learned faults. Since the unlearned fault is severer than single faults in training datasets (because the unlearned faults are applied to critical buses at the same time), the proposed model can make decision earlier with adaptive assessment mechanism rather than wait decision for the observation window of a certain length.

4.4.5. The effect of PMU noise

In previous analyses, we assume that PMUs can accurately sample system variables in high frequency. However, PMUs suffer from wide-area noise which leads to measurement error. According to IEEE Standard of Synchrophasor Data Transfer for Power Systems (C37.118.2-2011) [44], the measurement error of a vector cannot exceed 1% of its real value for all PMUs. To verify the robustness of the proposed method to wide-area noise, a numerical simulation is performed in this test.

In order to generate noise complying with IEEE standard, the approach in [29] is followed. Then, the simulated wide-area noise is added to datasets and other settings are same as the previous simulations. Results are summarized in Table 4 and it can be concluded that the proposed method achieves superior performance than existing methods. Accuracy of the proposed method only slightly influenced by

**Table 3**  
Overall performance comparison under unlearned faults.

Method	$E(A)$	$E(R)$
Proposed method	97.96%	1.191 cycles
MI	81.98%	4 cycles
DTSS	92.26%	4 cycles
RFSS	95.54%	4 cycles
Ensemble RVFL	73.39%	4 cycles

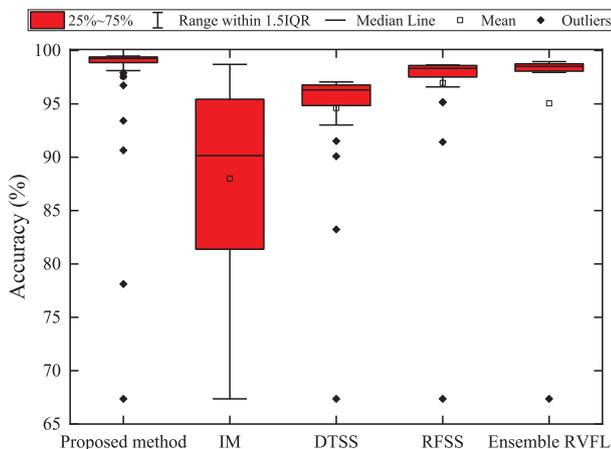


Fig. 13. The accuracy preference distribution of all methods.

**Table 4**  
Overall performance comparison under PMU noise.

Method	$E(A)$	$E(R)$
Proposed method	99.43%	1.188 cycles
MI	97.92%	4 cycles
DTSS	96.92%	4 cycles
RFSS	98.54%	4 cycles
Ensemble RVFL	98.80%	4 cycles

**Table 5**  
Overall performance comparison under random generation and demand.

Method	$E(A)$	$E(R)$
Proposed method	96.35%	1.399 cycles
MI	81.57%	4 cycles
DTSS	89.52%	4 cycles
RFSS	93.68%	4 cycles
Ensemble RVFL	73.46%	4 cycles

noise since it has great capacity of extracting spatial-temporal information from power system dynamic measurement, while its response time (1.188 cycles) is slightly shorter than noiseless data due to some sample around stability margin corrupted by noisy data.

#### 4.4.6. Model performance under random variation of generation and demand

To consider the random variation of generation and demand of a modern power system, we generate samples according to [45]. The output of the generator on bus 37 varies between zero and rated power, and each load is assumed to vary between 0.6 and 1.4 of their nominal values. The output of the generator on bus 37 and loads are assumed to be normally distributed, and the operation instances are obtained by Monte-Carlo technique [46] that randomly samples the output of the generator on bus 37 and each load within their variation range. Among these instances, the non-convergent ones are ignored and we only apply three-phase short-circuit fault to each bus in the New England 39-bus power system. Besides, the fault duration is set to 0.2 s. Finally, 200 operation instances are randomly sampled and 7800 samples are obtained.

The performance of the proposed method on these random samples is shown in Table 5. It can be seen that the proposed method performs TSA well in the highly variable generation and demand while other methods suffer overfitting, because the proposed method tries to depict the stability margin rather than just memorize samples [47,48]. The proposed method in this paper is robust to different operation conditions, since the stability margin of power system doesn't change unless the topology changes.

## 5. Conclusions

In this paper, a spatial-temporal adaptive method based on ensemble LSTM to implement rapid and credible TSA with missing data has been developed. The idea is to integrate LSTM classifiers automatically according to PMU clusters and implement TSA adaptively with the temporal data when PMU failure occurs. To do so, Relief-FT is proposed to calculate the importance of temporal features, and the optimal PMU clusters searching model is built to find observability-constrained temporal feature subsets to reduce the risk of missing data based on the feature importance. Furthermore, a novel ensemble mechanism for time-adaptive LSTM is designed to improve the performance of the accuracy and response time for TSA with the feature subsets. The simulation results using the New England 39-bus power system reveal that the number of missing PMUs is the major factor that deteriorates the performance of TSA in missing-data scenarios. Compared with existing methods, besides outperforming them both in

accuracy and response time, the proposed method shows more robustness on TSA with PMU failure events.

However, data injection attacks are becoming more and more threatening with the development of the smart grid. Therefore, the approach to deal with data incorrectness issues for TSA to avoid serious losses will be studied in future work. Also, if topology changes, the learnt patterns are different. And transfer learning will be studied and employed in future work to improve the generalization of the proposed method when the topology changes.

## CRedit authorship contribution statement

**Bendong Tan:** Conceptualization, Methodology, Software, Writing - original draft, Data curation. **Jun Yang:** Writing - review & editing, Supervision, Data curation, Funding acquisition. **Ting Zhou:** Methodology. **Xiangpeng Zhan:** Methodology. **Yuan Liu:** Methodology. **Shengbo Jiang:** Methodology. **Chao Luo:** Methodology.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The work is funded by the National Key R&D Program of China under Grant 2018AAA0101502.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijepes.2020.106237>.

## References

- [1] Liu T, Liu Y, Xu L, et al. Non-parametric statistics-based predictor enabling online transient stability assessment. *IET Gener Transm Distrib* Nov. 2018;12:5761–9.
- [2] Sun M, Konstantelos I, Strbac G. A deep learning-based feature extraction framework for system security assessment. *IEEE transactions on smart grid* 2019. <https://doi.org/10.1109/TSG.2018.2873001>.
- [3] Mazhari SM, Safari N, Chung CY, Kamwa I. A hybrid fault cluster and thévenin equivalent based framework for rotor angle stability prediction. *IEEE Trans Power Syst* 2018;33(5):5594–603.
- [4] Yu JQ, Hill DJ, Lam AYS, et al. Intelligent time-adaptive transient stability assessment system. *IEEE Trans Power Syst* 2018;33:1049–58.
- [5] Xue Yusheng, Van Cutsem Thierry, Ribbens-Pavella M. A simple direct method for fast transient stability assessment of large power systems. *IEEE Trans Power Syst* 1988;3:400–12.
- [6] Kakimoto N, Ohnogi Y, Matsuda H, et al. Transient stability analysis of large-scale power system by Lyapunov's direct method. *IEEE Power Eng. Rev.* 1984;PER-4(1):41–7.
- [7] Vittal V, Zhou EZ, Hwang C, et al. Derivation of stability limits using analytical sensitivity of the transient energy margin. *IEEE Trans Power Syst* 1989;9(11):33–4.
- [8] Xue Y, Van Cutsem CT, Ribbens-Pavella M. Extended equal area criterion justifications, generalizations, applications. *IEEE Trans Power Syst* 1989;4(1):44–52.
- [9] Liu CW, et al. Detection of transiently chaotic swings in power systems using real-time phasor measurements. *IEEE Trans Power Syst* 1994;9(3):1285–92.
- [10] Yan J, Liu CC, Vaidya U. PMU-based monitoring of rotor angle dynamics. *IEEE Trans Power Syst* Nov. 2011;26(4):2125–33.
- [11] Wei S, Yang M, Qi J, Wang J, Ma S, Han X. Model-free MLE estimation for online rotor angle stability assessment with PMU data. *IEEE Trans Power Syst* May 2018;33(3):2463–76.
- [12] Sobbouhi Ali Reza, Aghamohammadi Mohammad Reza. A new algorithm for predicting out-of-step using rotor speed-acceleration based on phasor measurement units (PMU) data. *Electric Power Compon Syst* 2015;43(13):1478–86.
- [13] Liu Yong, Sun Kai, Liu Yilu. A measurement-based power system model for dynamic response estimation and instability warning. *Electric Power Syst Res* 2015;124:1–9.
- [14] Shi Fang, Zhang Hengxu, Xue Guiting. Instability prediction of the inter-connected power grids based on rotor angle measurement. *Int J Electr Power Energy Syst* 2017;88:21–32.
- [15] Huang Dan, et al. Wide-area measurement system-based model-free approach of post-fault rotor angle trajectory prediction for on-line transient instability detection. *IET Gener Transm Distrib* March. 2018.;12(10):2425–35.

- [16] Zhang Y, Xu Y, Dong ZY, Xu Z, Wong KP. Intelligent early warning of power system dynamic insecurity risk: toward optimal accuracy-earliness tradeoff. *IEEE Trans Ind Inf* 2017;13:2544–54.
- [17] Gomez FR, Rajapakse AD, Annakkage UD, Fernando IT. Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. *IEEE Trans Power Syst* 2011;26(3):1474–83.
- [18] Bashiri Mosavi A, Amiri A, Hosseini H. A learning framework for size and type independent transient stability prediction of power system using twin convolutional support vector machine. *IEEE Access* 2018;6:69937–47.
- [19] Hashiesh F, Mostafa HE, Khatib A-R, Helal I, Mansour MM. An intelligent wide area synchrophasor based system for predicting and mitigating transient instabilities. *IEEE Trans Smart Grid* 2012;3(2):645–52.
- [20] Wang B, Fang B, Wang Y, Liu H, Liu Y. Power system transient stability assessment based on big data and the core vector machine. *IEEE Trans Smart Grid* 2016;7(5):2561–70.
- [21] Papadopoulos PN, Guo T, Milanović JV. Probabilistic framework for online identification of dynamic behavior of power systems with renewable generation. *IEEE Trans Power Syst* 2018;33(1):45–54.
- [22] Guo T, Milanović JV. Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability. *IEEE Trans Power Syst* 2014;29(1):377–85.
- [23] Papadopoulos PN, Milanović JV. Probabilistic framework for transient stability assessment of power systems with high penetration of renewable generation. *IEEE Trans Power Syst* 2017;32(4):3078–88.
- [24] Yan R, Geng G, Jiang Q, Li Y. Fast transient stability batch assessment using cascaded convolutional neural networks. *IEEE Trans Power Syst* 2019. <https://doi.org/10.1109/TPWRS.2019.2895592>.
- [25] Zhu Qiaomu, et al. A deep end-to-end model for transient stability assessment with PMU data. *IEEE Access*. 2018;6:65474–87.
- [26] Zhang R, Xu Y, Dong ZY, et al. Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system. *IET Gener Transm Distrib* 2015;9:296–305.
- [27] Yu JJQ, Lam AYS, Hill DJ, et al. Delay aware intelligent transient stability assessment system. *IEEE Access* 2017;5:17230–9.
- [28] James JQ, Hill David J, Lam Albert YS. Delay aware transient stability assessment with synchrophasor recovery and prediction framework. *Neurocomputing* 2018;322:187–94.
- [29] He M, Vittal V, Zhang J. Online dynamic security assessment with missing PMU measurements: A data mining approach. *IEEE Trans Power Syst* 2013;28(2):1969–77.
- [30] Zhang Y, Xu Y, Dong ZY. Robust classification model for PMU-based on-line power system DSA with missing data. *IET Generat, Transmiss Distrib* 2017;11(18):4484–91.
- [31] Zhang Y, Xu Y, Dong ZY. Robust ensemble data analytics for incomplete PMU measurements-based power system stability assessment. *IEEE Trans Power Syst Jan*. 2018;33(1):1124–6.
- [32] Zhang Y, Xu Y, Zhang R, Dong ZY. A missing-data tolerant method for data-driven short-term voltage stability assessment of power systems. *IEEE Trans Smart Grid* 2018. <https://doi.org/10.1109/TSG.2018.2889788>.
- [33] Nuqui RF, Phadke AG. Phasor measurement unit placement techniques for complete and incomplete observability. *IEEE Trans Power Deliv* 2005;20(4):2381–8.
- [34] Aminifar F, Khodaei A, Fotuhi-Firuzabad M, Shahidehpour M. Contingency-constrained PMU placement in power networks. *IEEE Trans Power Syst* 2010;25(1):516–23.
- [35] Urbanowicz RJ, Meeker M, La Cava W, et al. Relief-based feature selection: introduction and review. *J Biomed Inform* 2018;85:189–203.
- [36] Tang Yi, et al. Hybrid method for power system transient stability prediction based on two-stage computing resources. *IET Gener Transm Distrib* 2017.;12(8):1697–703.
- [37] Baldi Pierre, Sadowski Peter J. Understanding dropout. *Adv Neural Informat Process Syst* 2013:2814–22.
- [38] Diederik Kingma JB. Adam: A method for stochastic optimization. *Proc. Int. Conf. Learn. Representations, San Diego, CA, USA*. 2015. p. 1–15.
- [39] Pai MA. Energy function analysis for power system stability. Springer Science & Business Media; 1989.
- [40] Website of the IEEE PES PSDPCSCS, Task Force on Benchmark Systems for Stability Controls: “IEEE 10 Generator 39 Bus System” [Online]. Available: <http://www.sel.eesc.usp.br/ieee/IEEE39/main.htm>, Accessed on: April. 24, 2020.
- [42] Batista GE, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Appl Artif Intell* 2010;17(5–6):519–33.
- [43] Breiman L. Random forests. *Mach. Learn.* 2001;45(1):5–32.
- [44] IEEE Standards Association. IEEE standard for synchrophasor data transfer for power systems. *IEEE Std C* 2011;37:18–20.
- [45] Zhang Yuchen, et al. Online power system dynamic security assessment with incomplete PMU measurements: a robust white-box model. *IET Gener Transm Distrib* 2018;13:662–8.
- [46] Anders, George J. Probability concepts in electric power systems. 1989.
- [47] Zheng Le, et al. Deep belief network based nonlinear representation learning for transient stability. 2017 IEEE power & energy society general meeting. *IEEE*; 2017.
- [48] Wei HU, et al. Real-time transient stability assessment in power system based on improved SVM. *J Mod Power Syst Clean Energy* 2019;7(1):26–37.