A Hybrid Data-Driven Method for Online Power System Dynamic Security Assessment with Incomplete PMU Measurements

Qiaoqiao Li, Yan Xu
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
Email: qiaoqiao001@e.ntu.edu.sg; xuyan@ntu.edu.sg

Chao Ren
Interdisciplinary Graduate School
Nanyang Technological University
Singapore
Email: renc0003@e.ntu.edu.sg

Junhua Zhao
School of Science and Engineering
Chinese University of Hong Kong (Shenzhen)
China
Email: fuxiharp@gmail.com

***Abstract*—With widely deployment of phasor measurement units (PMU), data-driven dynamic security assessment (DSA) has been a promising tool to help power system counteract large disturbance and avoid instability. In case of PMU lost events, the DSA performance may be dramatically degraded. This paper designs a new online DSA model based on hybrid ensemble learning to address the missing data problem. In this method, a set of missing data estimators is proposed to restore the system measurement, and potential inaccurately estimated features are detected in feature validation stage to further ensure the model performance. After excluding inaccurate estimations, the filled-up dataset is provided to support accurate DSA under PMU lost situation. A hybrid ensemble of extreme learning machine (ELM) and random vector functional link networks (RVFL) is designed as learning engine for estimators and DSA classifiers. Simulation results show this approach achieves low estimation error as well as high DSA accuracy and strong robustness.**

***Index Terms*— Data-driven, dynamic Security Assessment, ensemble learning, phasor measurement units.**

I. Introduction

On-line power system dynamic security assessment (DSA) is one of the most effective approaches to evaluate online system security status. It could help to prevent system losing synchronism or even suffering cascading failure under various disturbances [1]. Unlike conventional time-consuming and information-requiring analytical method, such as time-domain (T-D) simulation, data-driven methods have increasingly attracted research attentions by its fast speed, less data requirement and knowledge discovery ability [1], [2]. Some data analytics algorithms are widely utilized for DSA, including decision tree (DT), neuron network (NN), extreme learning machine (ELM), etc. [3]-[6]. The main principle of these approaches is to capture the complex non-liner relationship between system operating states (called inputs or features) and stability information (called outputs or targets). The database consists of instances labelled by off-line simulated post-disturbance stability results. When applied online, the trained DSA model could map the collected features to a pre-contingency security status immediately [3]. Thus, it could support operators and system level controllers to trigger rescheduling or emergency controls to avoid instability against dangerous disturbances. More importantly, the wide-spread deployment of PMU can provide data foundation for applying data-driven methods in online DSA decision and control actions.

However, the performance of those highly data-dependent methods would be greatly challenged by synchronized measurement data loss. In practice, the completeness of system observation would suffer from unexpected PMU failure, communication link delay, signal interruption, cyber-attack etc. [3]-[6]. The incomplete feature input would detrimentally influence the accuracy of individual learning models [5]. To handle this problem, an effective online DSA method needs to be developed for missing data against any PMU loss events.

On this topic, some researchers have proposed decision tree with *surrogate* (DTWS) to solve missing features [3]. However, due to complex coupling characteristics of power system features and limitation of tree structure, the DSA accuracy dramatically degraded. Further study shows that during one PMU loss scenario, most of co-located features measured by the same PMU will all be unavailable. Whereas, viable non-co-located surrogate has low association with lost one, which leads to sequential decision failure of DT [4]. By assembling small DTs which are trained by location-separated random subspace, the approach in literature [4] achieves better DSA viability compared to single DT methods. However, the available measurements are still not fully utilized. In [5], [6], a robust ensemble model is designed to strategically collect observation-constrained PMU clusters as feature subsets. These subsets are further used to train single DSA classifiers respectively. During various PMUs missing scenarios, the ensemble of visible classifiers could achieve maximum availability of system observation with minimum single classifier number, thus maintaining high accuracy and computational efficiency.

However, all of these works could not remit the adverse impact to the completeness of the grid operating information. As a result, some critical features that suggest the system's weak points and thus supports further control actions may be unfortunately lost.

To handle this problem, this paper designs a new DSA method with feature-validation-based missing data estimators. Firstly, a set of estimators is proposed to predict missing data by other PMU measurement, and then more complete state information of the grid could be provided to evaluate the security status. In doing so, not only the lost features are restored, high DSA accuracy is also sustained in various PMU lost scenarios. Considering the possibility that some features may have low correlation with other measurement, a validation stage is designed to exclude potential inaccurately estimated features in further DSA classifier training and online application. In this way, a low online estimation error can be achieved with a higher DSA accuracy. Moreover, by integrating multiple randomized learning algorithms including ELM and RVFL in ensemble learning, more diversified machine learning outcome can be obtained, thus further ensuring model performance. The proposed method is tested on New England 10-machine 39-bus system with 8 PMUs, which verifies its high accuracy of estimation and DSA classification even under the worst PMU missing event, which outperforms DTWS method.

II. RANDOMLIZED LEARNING

This paper applies the randomized learning algorithms, ELM and RVFL as learning algorithms for developing the data-driven DSA model [1], [5], [6]. The fundamentals of these two learning algorithms are introduced in this section.

A. Extreme Learning Machine (ELM)

ELM belongs to single-hidden layer feedforward networks (SLFN) with randomized input weights and bias, and analytically determined output weights, which was proposed by Huang in 2006 [7]. Fig. 1 shows the structure of ELM. It includes three layers: input layer, hidden layer and output layer. For instances with n-dimension input vector and m-dimension target vector, Eq. 1 demonstrates the output function of ELM with L hidden nodes and activation function ϑ.

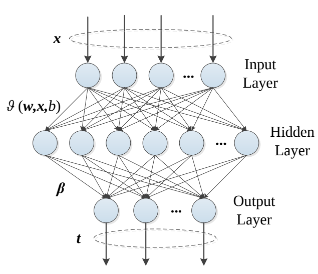

Fig. 1 Structure of ELM

$$f_L(\boldsymbol{x}_j)=\sum_{i=1}^{L}\boldsymbol{\beta}_i\cdot\vartheta(\boldsymbol{w}_i\cdot\boldsymbol{x}_j+b_i)=\boldsymbol{t}_j \quad j=1,2,\dots,N \tag{1}$$

where the $\boldsymbol{x}_j$ is the j th instance input vector out of N total instances, $\boldsymbol{x}\in R^n$; $\boldsymbol{t}_j$ is the j th instance target vector, $\boldsymbol{t}\in R^m$; $\boldsymbol{w}_i$ is the vector of weight linking input nodes to the i th hidden node; b_i is the threshold of the i th hidden node, which could also be written in the form of $\boldsymbol{w}_0$; $\boldsymbol{\beta}_i$ is the weight vector from i th hidden nodes to the output nodes.

In ELM theory, $\boldsymbol{w}$ and $\boldsymbol{b}$ could be randomly assigned, then the unknown variable vector only remains $\boldsymbol{\beta}$. In that way, by minimizing the training error between output and target value, the output of hidden layer could be determined, which dramatically simplify the training process. More specifically, a compact format of Eq.1 is given as follows.

$$\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{T} \tag{2}$$

where $\boldsymbol{H}$, called hidden layer output matrix, is determined by input weights and input vectors. If instance number N is larger than the hidden layer number L, which mostly occurs in practice, the output parameter could be obtained by $\boldsymbol{\beta}=\boldsymbol{H}^{\dagger}\boldsymbol{T}$. In this equation, $\boldsymbol{H}^{\dagger}$ is Moor-Penrose generalized inverse of $\boldsymbol{H}$, and the minimum square error could be achieved.

ELM outperforms the conventional learning schemes like gradient descent especially in terms of learning speed by skipping the iterative learning process and complex parameter turning stage. Thus, ELM is perfect candidate to implement very fast DSA in real time scale.

B. Random Vector Functional Link (RVFL)

RVFL was proposed by Pao [8] in 1994 and has been widely applied in various fields. Although the rule to train the network of RVFL is similar to ELM, RVFL has distinctive direct links from input layer to output layer, and it plays an important role in performance enhancing. The basic structure of RVFL is demonstrated in Fig. 2.

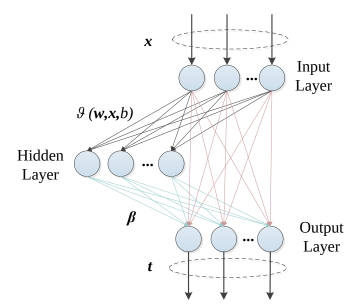

Fig. 2 Structure of RVFL

Similar to the format of ELM, the output weights could be determined by solving the following equation:

$$\boldsymbol{t}_i=\boldsymbol{h}_i^T\boldsymbol{\beta}, \quad i=1,2\dots N \tag{3}$$

where $\boldsymbol{t}$ is the instance target vector, N is the number of input instances, $\boldsymbol{h}$ is the vector version of the concatenation of the input layer nodes and the random features in hidden layer, $\boldsymbol{\beta}$ is the output weight vector.

As mentioned before, the weights connecting input features and hidden nodes are randomly generated from a uniform distributed interval. Moreover, the scope of interval is turned to achieve a trade-off between discrimination power and neuronal saturation in hidden layer. After calculating $\boldsymbol{h}, \boldsymbol{\beta}$, as weights connecting hidden layer to output layer and input layer directly to output layer, could be obtained in a single step. Studies shows these direct links could serve as regularization

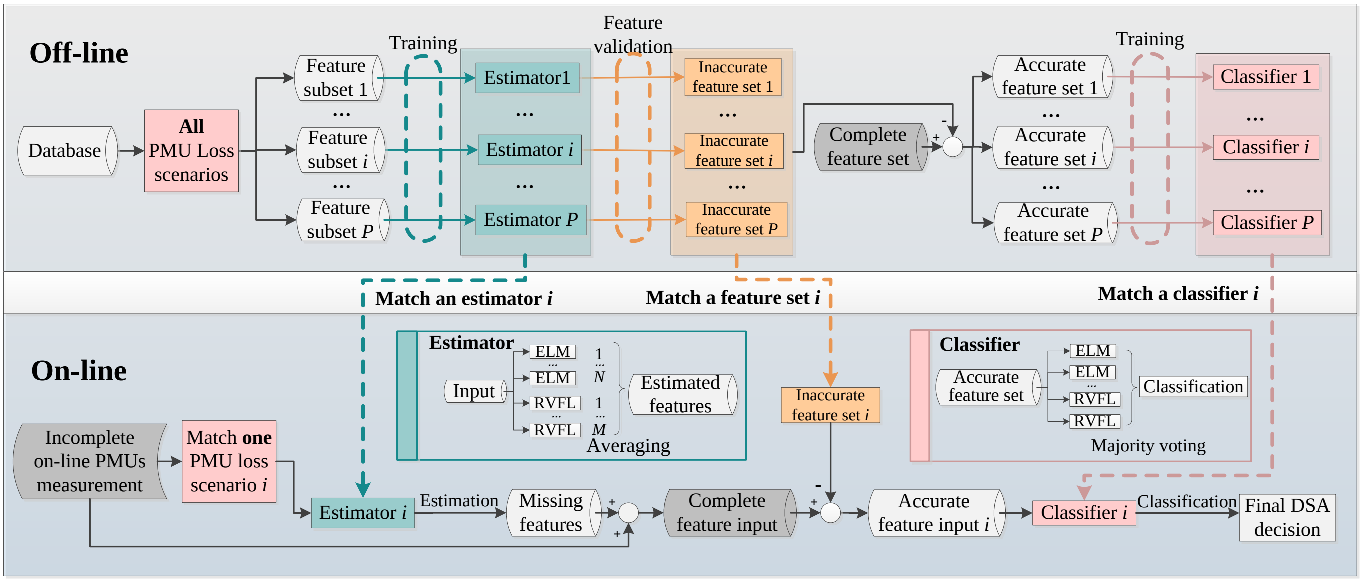

Fig. 3 Framework of proposed DSA model

items and lead to a thinner network with less hidden neurons, thus having a high possibility to achieve better performance than those without [9].

III. Proposed DSA Model

To support accurate and robust DSA against missing PMU events, this developed method proposes a set of missing data estimators to fully utilize information provided by available measurement and remit the adverse impact of PMU loss to the completeness of the grid observation. According to Kirchhoff's circuit laws, AC power flow equations and the grid topology, most of the missing features have complex coupling correlation with other system operating features, thus, it is theoretically feasible for them to be estimated by other available synchronous measurement. Since the potential inaccurately estimated features are detected and abandoned by validation step, not only online estimation error could be decreased, DSA accuracy and robustness are also improved. Otherwise, some estimated feature may dramatically deviate from the real value especially in severe PMU lost event, thus misleading the DSA decision and even control action. Also, it is necessary to emphasis that this proposed method implements accurate missing data recovery with available PMUs measurement, which is not able to deal with situation that all PMUs are destroyed.

A. Model Design

The whole scheme of this method is illustrated in Fig.3, which could be divided in off-line training stage and on-line application stage. At the first part of off-line stage, a set of estimators are trained by the available feature subset under each possible PMU missing scenario, respectively. Meanwhile, a validation dataset is utilized to test the trained estimators and accuracy information is obtained. At the second part, after excluding inaccurate features from the full feature set, these rearranged datasets are used to train DSA classifiers. At on-line stage, the missing features are filled up by matched estimator and inaccurate features are abandoned. Then the security status is obtained by corresponding DSA classifier. To increase learning diversity, a hybrid ensemble of ELM and RVFL is designed for feature estimation and DSA classification.

B. Missing Data Estimator

As is illustrated in Fig.3, the full feature set is rearranged into P different training subset. Each of them is the maximum available measurement under one possible PMU loss scenario. If existing L PMUs in the grid, then P=2^L-1. Then situation that all the PMU are lost is not considered. Then each of the P training dataset and their corresponding missing features are responsible to train one estimator as input and target. After that the well-trained estimators are tested in validation set respectively, where the missing features that have relatively low accuracy are considered as can-not-be-estimated features in particular PMU loss situation, which will not be utilized as input features in corresponding DSA classifier training and classification. The process is described as follows:

Missing Data Estimator Training and Feature Validation

For a power grid including L PMUs, given a database of $F \times I$ size (where F and I are number of total features and instances) , N single ELM and M single RVFL.

for i=1**to** L-1 **do**
  Define a set of vectors $\{\boldsymbol{u}_j, j=1\ldots C_L^i\}$, where $\boldsymbol{u}_j$ is a vector denoting a feature subset measured by i out of L PMUs.
  for j=1**to** C_L^i **do**
    Extract $F_k=\|\boldsymbol{u}_j\|$ (k=1…2^L-1) columns (corresponding to features in $\boldsymbol{u}_j$) from database to dataset $\boldsymbol{R}_{I\times F_k}$, and the rest "missing" features as target set $\boldsymbol{V}_{I\times(F\text{-}F_k)}$.
    Divide dataset and target set into training set $\boldsymbol{R}_{I_t\times F_k}$, $\boldsymbol{V}_{I_t\times(F\text{-}F_k)}$ and validation set $\boldsymbol{R}_{I_v\times F_k}$, $\boldsymbol{V}_{I_v\times(F\text{-}F_k)}$.
    for n=1 **to** N+M **do**
      Train ELM or RVFL on training set.
      Test the trained ELM or RVFL on validation set, and obtain the accuracy vector of target features.
    end for
    Obtain mean error vector $\boldsymbol{O}_k$ of these N+M learning units, define inaccurate feature vector $\boldsymbol{m}_k=\{m|o_m\in\boldsymbol{O}_k \,\&\, o_m>r\}$, where r is the threshold of error.
  end for

12: **end for**
13: Return the inaccurate feature set $\boldsymbol{D}=\bigcup_{k=1...P}\boldsymbol{m}_k$, $\boldsymbol{D}$ is a set of vectors.

C. DSA Classifier

As further illustrated in Fig.3, after the inaccurate feature set $\boldsymbol{D}$ is obtained, the DSA database are also rearranged into P training datasets by respectively excluding the inaccurate features from the full feature set. Then each training dataset and corresponding security labels are utilized to train N ELMs and M RVFLs, which are further assembled as the final DSA classifier for one PMU lost scenario.

Ensemble-based DSA Classifier Training

Given a training dataset $F\times I_t$, inaccurate feature set $\boldsymbol{D}=\bigcup_{k=1...P}\boldsymbol{m}_k$, N single ELM and M single RVFL
1: **for** k=1 **to** P **do**
2: Exclude features in $\boldsymbol{m}_k$ from full feature set and obtain accurate feature subset $\boldsymbol{p}_k$.
3: **for** j=1 **to** N+M **do**
4: Randomly sample i instances from new dataset, f features from new dataset, $1\leq i\leq I_t$, $1\leq f\leq \|\boldsymbol{p}_k\|$
5: Select activation function and randomly assign hidden nodes from pre-turned range of ELM or RVFL.
6: Train this ELM or RVFL.
7: **end for**
8: **end for**

It is important to point out that to maximum the performance enhancement of ensemble learning, this model achieves learning diversity in the following three aspects [11],[12]: data diversity, structure diversity and parameter diversity. As is shown in Fig. 4. Data diversity is achieved by randomly sample instances and attributes while training. Structure diversity is achieved by integrating different single learners and randomly assigned hidden nodes number. Due to random nature of the input weight determination scheme of ELM and RVFL, parameter diversity is also achieved. Moreover, this scheme speeds up the training process of ensemble learning, making randomized learning algorithm a perfect base learner. By this ensemble-based learning scheme, the single learners could compensate for each other, thus the accuracy and robustness of this DSA model are further guaranteed.

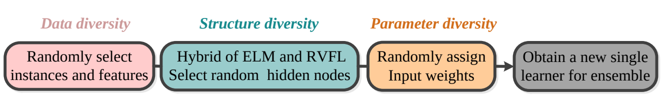

Fig.4 Diversity of ensemble learning

D. On-line Application

As shown in Fig.3, under incomplete PMU measurement scenarios, the corresponding estimator would fill up the lost part by estimating missing features and meanwhile abandoning the features that have a high possibility to be inaccurate to form up a relatively complete and accurate feature set for DSA input. As illustrated in Section III.A, the value of estimated features is obtained by averaging the prediction of N+M single learners in estimator. Finally, the DSA decision will be obtained by the matched classifier based on majority voting. Decision rule could be demonstrated as follows.

Rule for classification

1: **for** i=1**to** N+M **do**
2: **if** $y_i>0$, $y_i=1$→output of this single learner is "secure"
3: **else** y_i=-1→output of this single learner is "insecure"
4: **end if**
5: **end for**
6: **if** $Y=\sum_{i=1}^{N+M}y_i>0$ → final decision is "secure"
7: **else if** $Y=\sum_{i=1}^{N+M}y_i<0$ → final decision is "insecure"
8: **end if**
where y_i is the output of the i th single learner, i=1…N+M. Y is the final result of the ensemble-based classifier.

IV. SIMULATION RESULTS

The proposed method is conducted on New England 39-bus system, where 8 PMUs installed with the minimum number of PMUs required to realize global observability.

A. Database Generation

The renewable energy output and load bus demand are randomly selected from the pre-defined range by Monte-Carlo simulation. Other features are obtained by solving optimal power flow with objection to minimize generation cost. 10 typical three-phase faults at inter-area corridors are simulated on Transient Stability Analysis Tool. Finally, 5043 different instances with security labels are obtained. 80% of those instances are used for training, 10% to build validation set and the other 10% to imitate on-line PMU measurement as test set.

B. Parameter Selection

1) Threhold of accuracy in feature validation: If the estimated error is higher than threhold r, the feature will not be utilized in DSA decision. r is defined as 0.3 in this paper.

2) Numbers of ELM and RVFL: In ensemble learning scheme, the generalization error will decrease as number of single learner increase. In this paper, the total number is set as 200 with 50 ELMs and 150 RVFLs.

3) Random parameters in ensemble learning: For ELM, activation function is chosen as Sigmoid function and the number of hidden nodes of ELM is better to be randomly selected from [200,400]; For RVFL, the activation function is Sine function, the optimal hidden nodes range is [100,300]. This could also verify that the direct link of RVFL could lead to a thinner model.

C. Testing Results

Considering all PMU loss scenarios, the proposed model is compared with DTWS in the same application environment, and the contribution of missing data estimation and feature validation step are also presented in this section. All the results are the average of result of ten contingencies.

1) Estimation performance comparision: the estimator error is evaluated by mean average percentage error (MAPE):

$$\text{MAPE}=\frac{1}{N_t}\cdot\sum_{i=1}^{N_t}\left|\frac{y_i^{real}\text{-}y_i^{predicted}}{y_i^{real}}\right| \quad (4)$$

where N_t is the total number of the testing instances, y^{real} and $y^{predicted}$ are the real and predicted feature value, respectively. The results in table I are the average MAPE of all PMU scenarios. It is obvious that without feature validation, the prediction of missing data will be far away from the real value, which demonstrates feature validation step effectively detects the features that cannot be estimated by the available measurements.

TABLE I. ESTIMATION WITH AND WITHOUT FEATURE VALIDATION

Average Computation Time	**Mean Missing Data MAPE**	
	Without feature validation	***With feature validation***
10.06 ms	55.5%	2.2%

2) Computational time: the average computational time for the implementation of such data recovery is 10.06ms, and 1.92ms of DSA decision. Moreover,as is indicated in table II, the missing data estiamtion time will grow up linearly with missing data number increases. However, even in the worst situation, the computational time is negligible for on-line PMU application. In that case, the reaction time could be saved to activate further control actions, thus the estimation method is feasible to conduct.

TABLE II. ESTIMATION TIME UNDER EACH PMU LOST SCENARIOS

Lost PMU number	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Computation time(ms)	5.29	6.13	7.78	9.45	11.37	14.78	15.62

3) Classification accuracy of DTWS and proposed model: The DSA accuracy is defined as the percentage of correctly classified instances in test set. The results corresponding to the average DSA accuracy on all situations for each fixed number of lost PMUs are plotted in Fig.5. It can be calculated that the accuracy degradation of proposed method from complete feature set to only one available PMU left is less than 1%, which demonstrates the strong robustness of this model. Whereas, the accuracy of DTWS is dramatically degraded.

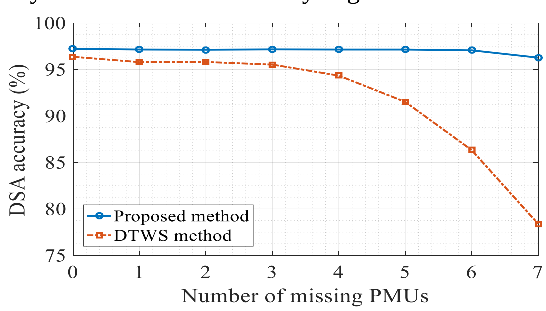

Fig 5. The performance of DTWS and popoesed method for DSA

4) Contribution of different strategies in proposed model: To further demonstrate contributions of different steps, three methods are tested for on-line DSA classification. As is shown in Fig. 6, the proposed method is the most robust with accuracy degradation at only 0.9%, and sustains the highest accuracy under any PMU lost numbers. While the method without feature estimation and validation has the least robustness at about 1.1% degradation and the least accuracy. It is calculated that the two steps of proposed methods contribute about 10% robustness improvement, respectively. In summary, the results illustrate that the two steps both play a role in DSA performance enhancement.

V. CONCLUSION AND FURTURE WORK

Based on hybrid ensemble learning, this paper proposes a new DSA method with feature-validation based missing data estimators. This method could provide accurate and relatively complete synchronous measurement of the system as well as sustain high DSA accuracy under various PMU missing scenarios. Simulation on New England 39-bus system verifies that this method achieves excellent performance, in regard to estimation accuracy as well as DSA robustness, over existing methods. In future work, to further justify the effectiveness of the proposed model, practical PMU signals with noises introduced in will be investigated and tested.

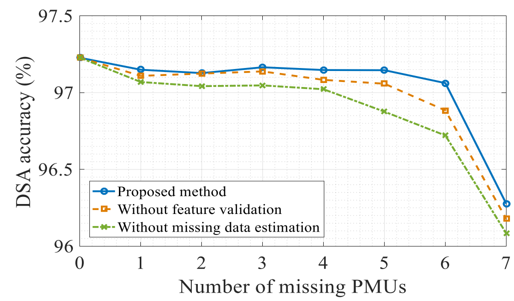

Fig 6. Contribution of different strategies in popoesed method for DSA

VI. ACKNOWLEDGEMENT

The work in this paper was supported by Ministry of Education, Republic of Singapore, under grant AcRF TIER 1 2017-T1-001-228 (RG92/17).

REFERENCES

[1] Y. Xu, Z.Y. Dong, K. Meng, R. Zhang and K.P. Wong, “Real-time transient stability assessment model using extreme learning machine,” *IET Gen. Trans. & Dist.*, vol. 5, pp. 314-322, Mar. 2011.

[2] Z.Y. Dong, Y. Xu, P. Zhang, and K.P. Wong, “Using intelligent system to assess an electric power system’s real-time stability,” *IEEE Intelligent Systems*, vol. 28, pp. 60-66, July-Aug. 2013.

[3] T. Y. Guo and J. V. Milanovic, “The effect of quality and availability of measurement signals on accuracy of on-line prediction of transient stability using decision tree method,” in *Proc. 4th IEEE/PES Conf.* pp. 1–5.

[4] M. He, V. Vittal, and J. S. Zhang, “Online dynamic security assessment with missing PMU measurements: A data mining approach,” *IEEE Trans. Power Syst.*, vol. 28, pp. 1969–1977, May 2013.

[5] Y.C. Zhang, Y. Xu, and Z.Y. Dong, "Robust Ensemble Data Analytics for Incomplete PMU Measurements-Based Power System Stability Assessment." *IEEE Trans. Power Syst.*, vol.33, pp. 1124-1126, Jan. 2018.

[6] Y.C. Zhang, Y. Xu, and Z.Y. Dong, "Robust classification model for PMU-based on-line power system DSA with missing data." *IET Gen. Trans. & Dist.*, vol. 11, pp. 4484-4491, Dec. 2017.

[7] G.B. Huang, “Extreme learning machine: Theory and applications,” *Neuro computing*, vol. 70, pp. 489–501, Dec .2006.

[8] Pao, Y.H. and Phillips, S. M. (1995). The functional link net and learning optimal control. *Neuro computing*, vol.9, pp149–164,1995.

[9] L. Zhang and P. N. Suganthan, "A comprehensive evaluation of random vector functional link networks," *Information sciences*, vol. 367, pp. 1094-1105, 2016.

[10] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees. London, U.K.: Chapman and Hall/CRC,1984.

[11] Y.C. Zhang, Y. Xu, Z.Y. Dong, et al. “Real-Time Assessment of Fault-Induced Delayed Voltage Recovery: A Probabilistic Self-Adaptive Data-driven Method”, *IEEE Trans. Smart Grid*, Feb. 2018.

[12] C. Ren, Y. Xu, Y. Zhang, C. Hu, “A multiple randomized learning-based ensemble model for power system dynamic security assessment,” *Proc. 2018 IEEE PES General Meeting*, Portland, OR, USA, Aug, 2018.