An Integrated Transfer Learning Method for Power System Dynamic Security Assessment of Unlearned Faults With Missing Data

Chao Ren , *Student Member, IEEE*, Yan Xu , *Senior Member, IEEE*, Bijian Dai, and Rui Zhang, *Member, IEEE*

Abstract—**This letter proposes an integrated transfer learning (TL) method for pre-fault dynamic security assessment (DSA) of power systems, which aims to simultaneously achieve fault transfer and address missing data issue for unlearned faults. Moreover, this letter provides the tight mathematic proof for the guaranteed DSA performance of the proposed integrated TL method. Comprehensive simulation results on the benchmark testing system have shown that the proposed integrated TL method can achieve a high DSA accuracy for unknown faults with complete data and can also maintain a satisfactory DSA accuracy for unknown faults even with incomplete data inputs.**

Index Terms—**Adversarial training, data-driven, dynamic security assessment, missing data, transfer learning.**

I. INTRODUCTION

DATA-Driven power system dynamic security assessment (DSA) has shown promising performance in recent years [1]. The principle is to train a machine learning model with a DSA database in the offline stage and apply it online with real-time measurements. However, there are two practical issues for real-world applications.

1) The offline training database only covers a limited number of faults. While in practice, other different faults may appear, which have a different data distribution from the training data. As a result, the trained DSA model may not be able to accurately work for such unlearned faults.

2) The data inputs may be missing due to some practical issues such as PMU malfunction, communication congestion/failure, or even cyber-attacks, etc. Once happens, the DSA model cannot work correctly.

To solve the first issue, Ref. [2] proposes a transfer learning (TL)-based method, aiming to use the existing trained DSA model to assess unlearned faults. Such TL-based method optimizes the joint marginal and conditional distributions between the training data and the unlearned data via reducing maximum

mean discrepancy. For the second issue, Ref. [3] proposes an ensemble learning model of decision trees (DT) by using the neighboring tree nodes to replace the missing data. The robust feature selection method via ensemble learning (RFEL) in Ref. [4] aims to strategically select observation-constrained PMU clusters as training database. Besides, Ref. [5] proposes a generative adversarial network (GAN)-based method to fill up the missing data, which is independent on PMU observability and network topologies.

In above works, the two issues are addressed separately, i.e., DSA for different faults with complete data or DSA for same faults with incomplete data. In practice, these two practical issues may occur at the same time, that is, the unlearned fault with incomplete data. Under this scenario, the existing feature-based TL method [2] will be ineffective. Besides, the existing ensemble method [3], RFEL method [4] and GAN-based method [5] cannot be directly applied since the incomplete data inputs are from the unlearned faults.

This letter proposes an integrated TL method to simultaneously address the above two issues. Through the adversarial training process, the DSA model trained by one fault can work for a different but related fault by confusing domain discriminator. With the property of feature extractor network, the proposed integrated TL method can solve the missing data issue for the unknown faults. In this way, the extensibility of the DSA models can be greatly enhanced, since the limited DSA models can cover more unlearned faults no matter with complete or incomplete data.

II. PROPOSED METHODOLOGY

A. Preliminary Definitions

For a DSA model, the inputs $\mathbf{x} \in \mathcal{X}$ are operating variables of the power system (e.g., power generation/load, and bus voltage magnitude) where $\mathcal{X}$ indicates the feature space, and corresponding labels y (e.g., secure or insecure) from the label space $\mathcal{P}$. Source domain $\mathcal{D}^s$ and target domain $\mathcal{D}^t$ represent different but related domains over $\mathcal{X} \otimes \mathcal{P}$ [6]. In general, source domain $\mathcal{D}^s(\mathbf{x}) = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n^s}$ has the full knowledge of a database (i.e., fully labeled instances), but target domain $\mathcal{D}^t(\mathbf{x}) = \{\mathbf{x}_i\}_{i=1}^{n^t}$ is only with limited knowledge (i.e., only with unlabeled instances or with a very small number of labeled instances), where n^s and n^t represent the number of source and target domain samples, respectively.

B. Principle of Proposed Method

The proposed integrated TL method is based on adversarial training. Its principle is to firstly perform feature learning to
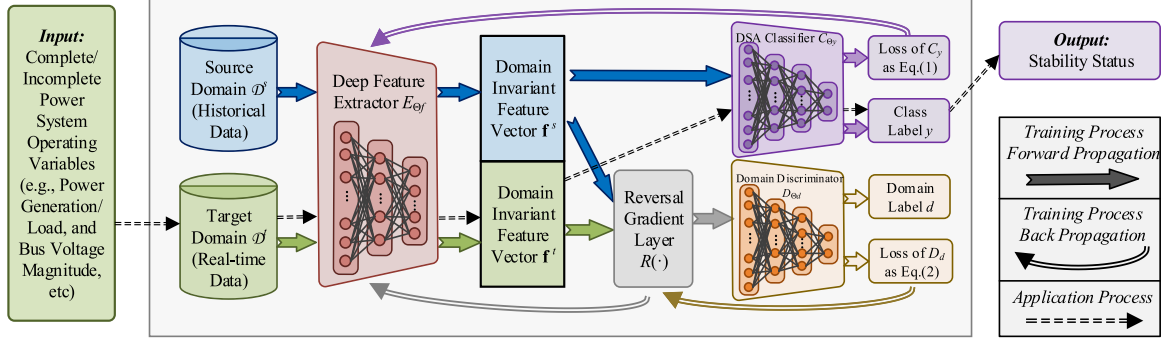
Fig. 1.   Framework of the proposed integrated TL method.

extract the *domain-invariant features* (i.e., common impact features of two domains) in a feature space from the input data for different faults. Then, by fooling the domain discriminator with such extracted domain-invariant features, the distribution of source domain and target domain becomes more similar. Thus, the DSA classifier trained by source domain can be used for unlabeled instances in target domain. Since the feature learning stage can also extract the domain-invariant features by incomplete target domain data, the proposed method can also accurately work with missing data.

C. Framework

The framework of the proposed integrated TL method is illustrated in Fig. 1, which consists of three parts: a deep feature extractor E_{Θ_f} which aims to extract the *domain-invariant features* from the source and target domain input data in a feature space, a DSA classifier C_{Θ_y} which aims to classify the power system stability status ($secure = 1/insecure = -1$), and a domain discriminator D_{Θ_d} which aims to distinguish whether the input data is from the source domain ($d = 0$) or the target domain ($d = 1$).

$E_{\Theta_f}(\cdot)$ is a deep neural network with model parameters Θ_f. The input $\mathbf{x}$ is mapped to a *M*-dimensional feature vector $\mathbf{f} \in \mathbb{R}^M$, such that $\mathbf{f} = E_{\Theta_f}(\mathbf{x})$. By mapping input data to feature space, $C_{\Theta_y}(\cdot)$ can predict the stability status of the domain data with domain-invariant features, and $D_{\Theta_d}(\cdot)$ cannot distinguish which domain the input fault data comes from.

$C_{\Theta_y}(\cdot)$ is a deep neural network with model parameters Θ_y. The feature vector $\mathbf{f}$ generated by $E_{\Theta_f}(\cdot)$ is used to predict the stability status y, such that $y = C_{\Theta_y}(E_{\Theta_f}(\mathbf{x})) \in \{-1, 1\}$. The loss function of DSA classifier, $\mathcal{L}_{C_y}^i(\Theta_f, \Theta_y)$, is formulated as Eq. (1).

$$\mathcal{L}_{C_y}^i(\Theta_f, \Theta_y) = \mathcal{L}_{C_y}\left(C_{\Theta_y}\left(E_{\Theta_f}(\mathbf{x_i})\right), y_i\right) \quad (1)$$

$D_{\Theta_d}(\cdot)$ is a *K*-layer fully connected neural network with model parameters Θ_d. The feature vector $\mathbf{f}$ generated by $E_{\Theta_f}(\cdot)$ is used to predict the domain label d, such that $d = D_{\Theta_d}(E_{\Theta_f}(\mathbf{x})) \in \{0, 1\}$. The loss function of domain discriminator, $\mathcal{L}_{D_d}^i(\Theta_f, \Theta_d)$, measures the dissimilarity of the different domains and can be formulated as Eq. (2).

$$\mathcal{L}_{D_d}^i(\Theta_f, \Theta_d) = \mathcal{L}_{D_d}\left(D_{\Theta_d}\left(E_{\Theta_f}(\mathbf{x_i})\right), d_i\right) \quad (2)$$

If the $D_{\Theta_d}(\cdot)$ cannot distinguish where the input data comes from, it means that the source domain and target domain are similar and target domain data can directly apply the DSA classifier trained by the source domain data.

In Fig. 1, the forward and backward solid lines represent the forward propagation and back propagation of the training process, respectively. For the application process (shown as the dashed lines in Fig. 1), given the real-time measurements of the unlearned fault no matter with the complete or incomplete data, the power system stability status can be directly predicted by the DSA classifier with the extracted *domain-invariant features*.

D. Training Process

At the training process, there are three parts of model parameters to be computed, that is $\Theta_f, \Theta_y, \Theta_d$, which correspond to deep feature extractor E_{Θ_f}, DSA classifier C_{Θ_y} and domain discriminator D_{Θ_d}, respectively. Note that E_{Θ_f} aims to maximize the loss function $\mathcal{L}_{D_d}^i$, but D_{Θ_d} aims to minimize loss function $\mathcal{L}_{D_d}^i$. Thus, the whole training process will be achieved through adversarial training. Then, the whole objective function of the proposed method can be formulated via combining Eq. (1) with Eq. (2), as Eq. (3).

$$\mathcal{J}(\Theta_f, \Theta_y, \Theta_d) = \frac{1}{n^s} \sum_{\mathbf{x_i} \in \mathcal{D}^s} \mathcal{L}_{C_y}^i(\Theta_f, \Theta_y)$$
$$- \frac{\sigma}{n^s + n^t} \sum_{\mathbf{x_i} \in (\mathcal{D}^s + \mathcal{D}^t)} \mathcal{L}_{D_d}^i(\Theta_f, \Theta_d) \quad (3)$$

where σ represents the trade-off parameters between two parts. The first part is the classification loss of source domain; the second part is the domain loss of source and target domain. Based on Eq. (3), the optimal parameters can be computed by searching the saddle points $(\Theta_f^*, \Theta_y^*, \Theta_d^*)$ as follows:

$$(\Theta_f^*, \Theta_y^*) = \arg\min_{\Theta_f, \Theta_y} \mathcal{J}(\Theta_f, \Theta_y, \Theta_d^*) \quad (4)$$

$$\Theta_d^* = \arg\max_{\Theta_d} \mathcal{J}(\Theta_f^*, \Theta_y^*, \Theta_d) \quad (5)$$

By adversarial training the Eq. (4) and Eq. (5) iteratively, the proposed method can achieve an equilibrium that D_{Θ_d} cannot distinguish different domain data, while C_{Θ_y} can accurately predict the stability status. Then, stochastic estimates of gradients

can be used to update the parameters as follows:

$$\Theta_y{}^* \leftarrow \Theta_y - \gamma \cdot \left(\partial \, \mathcal{L}_{C_y}^i / \partial \Theta_y \right) \tag{6}$$

$$\Theta_d{}^* \leftarrow \Theta_d - \gamma \cdot \sigma \cdot \left(\partial \, \mathcal{L}_{D_d}^i / \partial \Theta_d \right) \tag{7}$$

$$\Theta_f{}^* \leftarrow \Theta_f - \gamma \cdot \left[\left(\partial \, \mathcal{L}_{C_y}^i / \partial \Theta_f \right) - \sigma \cdot \left(\partial \, \mathcal{L}_{D_d}^i / \partial \Theta_f \right) \right] \tag{8}$$

where γ denotes the learning rate. It can be seen that the only difference between the process of Eq. (6-8) and stochastic gradient descent (SGD) is the sign factor in Eq. (8), the gradients from E_{Θ_f} and C_{Θ_y} should be subtracted, rather than being summed. To directly implement SGD, a special reversal gradient layer $R(\cdot)$ between E_{Θ_f} and D_{Θ_d} as Eq. (9) is used [7]. For the forward propagation, $R(\cdot)$ is an identity transformation; for the backpropagation, $R(\cdot)$ takes the gradient from the D_d and change its sign, multiplies it by '-1', (i.e., $\partial \, \mathcal{L}_{D_d}^i / \partial \Theta_f$ is replaced by $-\partial \, \mathcal{L}_{D_d}^i / \partial \Theta_f$) before passing to E_{Θ_f}.

$$\begin{cases} R\left(\mathbf{x}\right) = \mathbf{x} & \text{if forward propagation} \\ dR/d\mathbf{x} = -\mathbf{I} & \text{if backpropagation} \end{cases} \tag{9}$$

where $\mathbf{I}$ represents an identity matrix. Based on the $R(\cdot)$, Eq. (3) can be reformulated to directly implement SGD as follows:

$$\tilde{\mathcal{J}} \left(\Theta_f, \Theta_y, \Theta_d \right) = \frac{1}{n^s} \sum_{\mathbf{x_i} \in \mathcal{D}^s} \mathcal{L}_{C_y} \left(C_{\Theta_y} \left(E_{\Theta_f} \left(\mathbf{x_i} \right) \right), y_i \right)$$

$$- \frac{\sigma}{n^s + n^t} \sum_{\mathbf{x_i} \in \left(\mathcal{D}^s + \mathcal{D}^t \right)} \mathcal{L}_{D_d} \left(D_{\Theta_d} \left(R \left(E_{\Theta_f} \left(\mathbf{x_i} \right) \right) \right), d_i \right) \tag{10}$$

E. Mathematic Proof for Guaranteed Performance

This section provides the mathematic proof for the guaranteed performance of the proposed integrated TL method. The proposed method can be considered to build a classifier $h(\cdot)$ to predict labels y given the input $\mathbf{x}$ for the target domain with a target error $\varepsilon_{\mathcal{D}^t}(h(\cdot))$.

$$\varepsilon_{\mathcal{D}^t}(h) = \Pr_{(\mathbf{x},y) \sim \mathcal{D}^t} \left(h\left(\mathbf{x}\right) \neq y \right) \tag{11}$$

Given the source and target domain $\mathcal{D}^s$, $\mathcal{D}^t$ and a hypothesis class $\mathcal{H}$, the $\mathcal{H}\Delta\mathcal{H}$-divergence can be formulated as:

$$d_{\mathcal{H}\Delta\,\mathcal{H}}\left(\mathcal{D}^s, \mathcal{D}^t\right) = 2 \sup_{h_1, h_2 \in \mathcal{H}}$$

$$\left| \Pr_{\mathbf{f} \sim \mathcal{D}^s} \left(h_1\left(\mathbf{f}\right) \neq h_2\left(\mathbf{f}\right) \right) - \Pr_{\mathbf{f} \sim \mathcal{D}^t} \left(h_1\left(\mathbf{f}\right) \neq h_2\left(\mathbf{f}\right) \right) \right| \tag{12}$$

where $\mathbf{f}$ is the feature vector generated by E_{Θ_f}. Even with incomplete target domain data, E_{Θ_f} can also extract the domain-invariant features. The target error $\varepsilon_{\mathcal{D}^t}(h)$ can be bounded by the source error $\varepsilon_{\mathcal{D}^s}(h)$ and $\mathcal{H}\Delta\mathcal{H}$-divergence as Eq. (13) [8]:

$$\varepsilon_{\mathcal{D}^t}(h) \leq \varepsilon_{\mathcal{D}^s}(h) + \frac{1}{2} \cdot d_{\mathcal{H}\Delta\,\mathcal{H}}\left(\mathcal{D}^s, \mathcal{D}^t\right) + \tau \tag{13}$$

where τ represents the shared error of the ideal joint hypothesis and is a constant term. It can be seen that if want to decrease $\varepsilon_{\mathcal{D}^t}(h)$, E_{Θ_f} should decrease $\varepsilon_{\mathcal{D}^s}(h)$ and $d_{\mathcal{H}\Delta\,\mathcal{H}}(\mathcal{D}^s, \mathcal{D}^t)$, C_{Θ_y} should decrease $\varepsilon_{\mathcal{D}^s}(h)$, and D_{Θ_d} should obtain the upper bound of $d_{\mathcal{H}\Delta\,\mathcal{H}}(\mathcal{D}^s, \mathcal{D}^t)$ representing the domain distance.

Consider fixed $\mathcal{D}^s$ and $\mathcal{D}^t$ over the feature space by E_{Θ_f} and a family of $\mathcal{H}_{C_y}$. In order to guarantee that Eq. (13) always holds, the hypothesis class $\mathcal{H}_{D_d}$ generated by D_{Θ_d} should be rich enough to include the hypothesis class $\mathcal{H}_{C_y}$ generated by C_{Θ_y}, such that $\mathcal{H}_{C_y} \subseteq \mathcal{H}_{D_d}$. Then E_{Θ_f} can be related to as follow:

$$d_{\mathcal{H}_{C_y}\Delta\,\mathcal{H}_{C_y}} \left(\mathcal{D}^s, \mathcal{D}^t \right)$$

$$= 2 \sup_{h \in \mathcal{H}_{C_y}\Delta\,\mathcal{H}_{C_y}} \left| \Pr_{\mathbf{f}\sim\mathcal{D}^s} \left(h\left(\mathbf{f}\right) = 1 \right) - \Pr_{\mathbf{f}\sim\mathcal{D}^t} \left(h\left(\mathbf{f}\right) = 1 \right) \right|$$

$$\leq 2 \sup_{h \in \mathcal{H}_{D_d}} \left| \Pr_{\mathbf{f}\sim\mathcal{D}^s} \left(h\left(\mathbf{f}\right) = 1 \right) - \Pr_{\mathbf{f}\sim\mathcal{D}^t} \left(h\left(\mathbf{f}\right) = 1 \right) \right|$$

$$= 2 \sup_{h \in \mathcal{H}_{D_d}} \left| \Pr_{\mathbf{f}\sim\mathcal{D}^s} \left(h\left(\mathbf{f}\right) = 0 \right) + \Pr_{\mathbf{f}\sim\mathcal{D}^t} \left(h\left(\mathbf{f}\right) = 1 \right) - 1 \right| \tag{14}$$

where $\mathcal{H}_{C_y}\Delta\,\mathcal{H}_{C_y} = \{h | h = h_1 \oplus h_2, \; h_1, h_2 \in \mathcal{H}_{C_y}\}$. Eq. (13) can be considered as the optimization by D_{Θ_d}. Eq. (14) indicates that the upper bound of $d_{\mathcal{H}_{C_y}\Delta\,\mathcal{H}_{C_y}}(\mathcal{D}^s, \mathcal{D}^t)$ can be obtained if can find a classifier $h(\cdot)$ in a family of $\mathcal{H}_{D_d}$ with sufficient complexity, Then $h(\cdot)$ will have the largest probability to judge $\mathcal{D}^s$ as '0', and $\mathcal{D}^t$ as '1', such as domain discriminator D_{Θ_d}. Finally, the proposed integrated TL method can obtain the better approximation of $\varepsilon_{\mathcal{D}^t}(h)$ as Eq. (13).

III. RESULTS AND CONCLUSION

The proposed integrated TL method is tested on New England 10-machine 39-bus system. A large number of operating points are generated based on Monte-Carlo method with randomly sampled generation and load within a certain range. Given the contingency set, the time-domain simulation is performed to evaluate the dynamic security label for each operating point by Transient Stability Analysis Tool (TSAT), when the maximum rotor angle separation is beyond 360 degrees, it is labelled as insecure. The contingencies considered in this study are the three-phase faults with inter-area corridor trip and cleared 0.25s after their occurrences [1]. Specifically, a total of 8 serious faults simulated on the total 40344 operating points with their security conditions are generated as Ref. [2], where 60% of samples are randomly sampled as training data and the remaining 40% serve as testing data for each fault.

Given the eight complete fault databases, the proposed integrated TL method is trained by complete $\mathcal{D}^s$ and complete/incomplete $\mathcal{D}^t$. Each of the DSA classifier trained by one known fault is transferred to the remaining unlearned seven faults, and the testing results for complete and incomplete target domain are shown in Fig. 2(a-c), where the symbol '→' in x-axis denotes transfer from the source domain to the target domain, e.g., 'F1→' means the DSA classifier trained by F1 is transferred to other faults; y-axis is the DSA accuracy of the proposed method. In Fig. 2(a), the DSA accuracy results are shown for the unlearned faults under the complete data scenario. Fig. 2(b) and Fig. 2(c) show the DSA accuracy results for the unlearned faults with 12.5% and 25% incomplete data, respectively. Note that we test several incomplete target domain data by randomly missing.

To further verify the superiority of the proposed TL method, the confusion matrix is tested, including true positive (TP), true negative (TN), false positive (FP) and false negative (FN), are
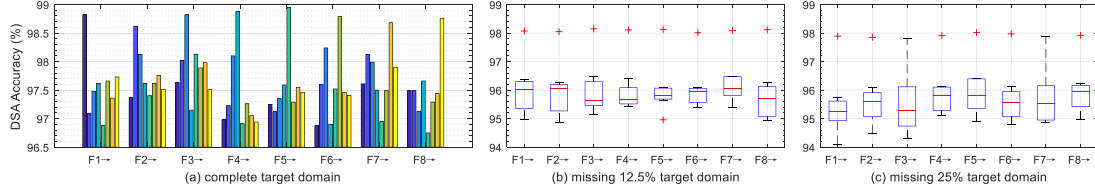
Fig. 2. Testing results of the proposed TL method: (a) complete target domain data, (b) 12.5% missing target domain data, and (c) 25% missing target domain data.

TABLE I
AVERAGE DSA PERFORMANCE OF DIFFERENT METHODS.

Method		Average DSA performance on target domain with the complete data			
		accuracy	precision	specificity	F1-score
Without TL	Ensemble learning	82.25%	90.59%	84.14%	85.66%
	DT	78.84%	87.37%	78.53%	82.98%
	LSTM	85.76%	92.03%	86.06%	88.70%
TL-based	Ref [2]	97.27%	98.81%	97.81%	97.89%
	Proposed method	97.68%	99.18%	98.48%	98.21%

TABLE II
AVERAGE DSA ACCURACY OF DIFFERENT METHODS.

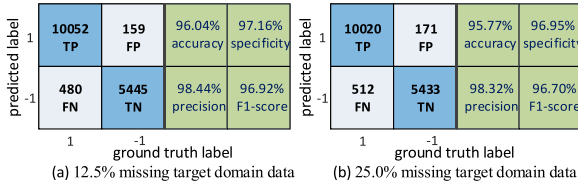| Method		Average DSA accuracy on target domain with the incomplete data	
		12.5% missing data	25% missing data
Without TL (e.g., DT, LSTM)		×	×
TL-based	Ref. [2]	×	×
	Proposed method	96.04%	95.77%

Fig. 3. Confusion Matrix of (a) 12.5% missing target domain data and (b) 25% missing target domain data.

TABLE III
GUARANTEED PERFORMANCE OF PROPOSED INTEGRATED TL METHOD.

Upper bound value	Complete	12.5% missing data	25% missing data
$d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}^s, \mathcal{D}^t)$	0.643	0.679	0.692
Target error $\varepsilon_{\mathcal{D}^t}(h(\cdot))$	0.356	0.395	0.423

method in Ref. [2] method will be ineffective, but the proposed integrated TL method can still work and maintain the DSA accuracy, i.e., achieving 96.04% and 95.77% with the 12.5% and 25% missing target domain data respectively. Overall, it can be seen that the proposed integrated TL method can accurately transfer the model to the unlearned faults with complete or incomplete data inputs.

Finally, the upper bound values of target errors as Eq. (13) are listed in Table III, which demonstrate that a higher DSA accuracy for unlearned fault corresponds to a smaller value of divergence $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}^s, \mathcal{D}^t)$ and target error $\varepsilon_{\mathcal{D}^t}(h(\cdot))$. The simulation is conducted on a computer with an Intel Core i7 CPU @ 3.3-GHz processor and 16-GB RAM. For the computation time, all of the above training processes can be finished in 1.2 seconds, and the proposed integrated TL method can be applied in real-time with the negligible processing time. The proposed method can be a very promising method to solve other similar safety-critical issues in power engineering.

III. REFERENCES

[1] Y. Xu, Y. Zhang, Z. Y. Dong, and R. Zhang, *Intelligent Systems for Stability Assessment and Control of Smart Power Grids*. Boca Raton, FL, USA: CRC Press, 2020.

[2] C. Ren and Y. Xu, "Transfer learning-based power system online DSA: Using one model to assess many unlearned faults," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 821–824, Jan. 2020.

[3] 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, no. 2, pp. 1969–1977, May 2013.

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

[5] C. Ren and Y. Xu, "A fully data-driven method based on generative adversarial networks for power system dynamic security assessment with missing data," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 5044–5052, Nov. 2019.

[6] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[7] Y. Ganin and V. Lempitsky. "Unsupervised domain adaptation by backpropagation," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[8] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, no. 1/2, pp. 151–175, May 2010.

[9] D. Powers, "Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

applied. For DSA problem, the most important concept is the number of FP cases, which means the insecure case is misclassified as the secure case, resulting in cascading failure or even wide-spread blackout. Then, based on FP, four relevant indices are utilized as Ref. [9], including precision, specificity, accuracy and F1-score. The larger the value of these four indices means the better the DSA performance. Under the complete target domain scenario, the average DSA performance of proposed method is compared with traditional methods without TL (ensemble learning with randomized learning models, LSTM and DT) and the existing TL method in Ref. [2], as shown in Table I. It can be seen that the proposed integrated TL method has the best performance in terms of the four indices, achieving the highest DSA accuracy up to 97.68% for the unlearned faults.

Besides, Table II shows the average DSA accuracy with the incomplete target domain data, and Fig. 3(a-b) show the detailed information for the plot of the confusion matrix [9], where the columns represent the ground truth label and the rows represent the predicted label. As Table II, the original DSA models and TL