

# Physical Mechanism Enabled Neural Network for Power System Dynamic Security Assessment

Guozheng Wang, *Member, IEEE, Member, CSEE*, Jianbo Guo, *Senior Member, IEEE, Fellow, CSEE*, Shicong Ma, *Member, IEEE, Senior Member, CSEE*, Kui Luo, *Member, CSEE*, Xi Zhang, Qinglai Guo, *Senior Member, IEEE*, Shixiong Fan, Tiezhu Wang, *Member, IEEE*, and Weilin Hou, *Member, CSEE*

**Abstract**—Data-driven artificial intelligence technologies have emerged as increasingly fascinating tools for assessing power system security. However, their inherent mechanism of inexplicability and unreliability now limits their scalability in power systems. To address this problem, this paper proposes a neural network design method empowered by physical mechanisms for power system security assessment. It incorporates geometric characteristics of dynamic security regions into the network training process and constructs connections between network structure and system's unstable mode, which can perform security assessment with a neural network efficiently while ensuring physical plausibility. Furthermore, a credibility evaluation mechanism is established to ensure credibility of neural network predictions and reduce misclassifications. Finally, effectiveness of the proposed method is verified on test systems. Methods and considerations in building a neural network with interpretable structures and credible predictions can provide a reference for machine intelligence applied in other industrial systems.

**Index Terms**—Credibility index, machine intelligence, neural network structure, physical properties, power system, security assessment.

## I. INTRODUCTION

IN recent years, data-driven artificial intelligence (AI) technologies have achieved astonishing results in science and engineering with large amounts of data [1]. For a physically embodied system (such as industrial systems), it is not feasible to rely only on data-driven intelligence technologies, and additional prior knowledge of physical mechanisms is required to improve adaptability and reliability of intelligence technologies [2]. However, this lag is somewhat obvious in comparison

to application of AI technologies, especially popular neural networks.

In the field of power systems, various neural networks, such as DNN [3], CNN [4], GNN [5], and LSTM [6], have been used for security assessment, which demonstrates better performance than conventional methods in computing efficiency and accuracy within proximity of training data. However, most studies focus on how to improve efficiency and performance of the neural network model from complex/new structures [3]–[6] and data sets [7]–[9] aspects, while few consider the incorporation of physics-prior [10]. Almost all such neural networks are developed based on experience of training and errors, and new kinds of neural networks are still proposed every year. Recently, extensive random search techniques have been used to search for more effective neural network structures [11]–[13]. Despite obvious advances, the structure design of networks still lacks rigorous theoretical basis. As a result, networks are usually designed and trained heuristically and then used as “black box” models, and now, application in power systems is increasingly limited.

In addition, for power system security assessment, model errors may lead to huge economic losses and even social unrest. Accordingly, neural networks in power systems require not only higher interpretability of model structure but also credibility of their prediction [14]. To achieve this goal, several issues need to be addressed: for a given security assessment task, what kind of neural networks needs to be designed? How much width and depth does the network need? What is the relationship between network structure and the physical mechanisms of security assessment? How to improve credibility of the neural network predictions?

Implementing complex models and large data sets is not sufficient to solve the above problems. Thus, proposals to realize a neural network by physics-prior have appeared [15], [16]. The first-principle-based constructive networks performed well at robot tracking control [15] and modeling for fluid flows [16]. These inspiring proposals nonetheless are difficult to apply directly in power systems due to the specialized nature of physical knowledge. Fortunately, Dynamic Security Region (DSR) methods [17]–[19] play an important role in power system security assessment while ensuring rigorous theoretical justification. Its physical properties should be highly correlated with the data-driven model and can be integrated into neural network design. With this consideration, we propose a neural network design method empowered by the physical properties

Manuscript received December 15, 2022; revised February 25, 2023; accepted March 15, 2023. Date of online publication June 27, 2023; date of current version September 4, 2024. This work was supported by the National Key R&D Program of China (2018AAA0101500).

G. Z. Wang (corresponding author, email: guozhengwang@foxmail.com) is with the Beijing Huairou Laboratory, Beijing, China, and the Department of Power System, China Electric Power Research Institute, Beijing 100192, China.

J. B. Guo, S. C. Ma, K. Luo, S. X. Fan, T. Z. Wang and W. L. Hou are with the Department of Power System, China Electric Power Research Institute, Beijing 100192, China.

X. Zhang is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China.

Q. L. Guo is with the Department of Electrical Engineering, Tsinghua University, Beijing 100084, China.

DOI: 10.17775/CSEEJPES.2022.08800

of DSR for power system security assessment. It incorporates the geometric characteristics of DSR into the neural network training process and constructs connections between network structure and unstable mode of power systems. Furthermore, a Credibility Index (CI) is established to ensure credibility of neural network predictions and reduce misclassifications. Incorporating the physical properties of DSR, CI inevitably ensures credibility of predictions. The proposed method performs very well at security assessment with a neural network efficiently while ensuring physical plausibility.

In the rest of the paper, design principles of neural network structures are proposed in Section II, and the CI for network predictions is constructed in Section III. In Section IV, physical significance of design principles and CI is illustrated using theoretical derivation of the DSR, and the proposed method is verified in Section V. Finally, a summary is provided in Section VI.

## II. DESIGN PRINCIPLES OF NEURAL NETWORK STRUCTURES BASED ON PHYSICAL PROPERTIES OF DSR

### A. Overview of DSR

Region-based security analysis methods play an important role in power system analysis, where the DSR determines a set of operating points that can ensure transient angle stability of the studied power system. The DSR is usually defined in power injection space [17]. It can be described as:

$$\begin{cases} \Omega_d(T_0, T_1, F) \{ \mathbf{x}_\beta | \mathbf{x}_\tau(\mathbf{x}_\beta) \in S(\mathbf{x}_\beta) \} \\ \mathbf{x}_\beta(P_1, P_2, \dots, P_n)^T \end{cases} \quad (1)$$

where  $T_0$  and  $T_1$  are grid structures before and after a fault, respectively,  $F$  is fault set of interest,  $\mathbf{x}_\beta$  is operating point described by a vector of power injections,  $\mathbf{x}_\tau(\mathbf{x}_\beta)$  is state of the power system after a fault,  $S(\mathbf{x}_\beta)$  is stability region of the post-fault power system,  $P_k$  is injected active power on bus  $k$ .

Both  $\mathbf{x}_\tau(\mathbf{x}_\beta)$  and  $S(\mathbf{x}_\beta)$  are determined by  $\mathbf{x}_\beta$ . For a given fault set  $F$ , each operating point  $\mathbf{x}_\beta$  corresponds to a transient stability region  $S(\mathbf{x}_\beta)$ . If  $\mathbf{x}_\tau(\mathbf{x}_\beta)$  is within  $S(\mathbf{x}_\beta)$ , it means  $\mathbf{x}_\beta$  is a security operating point, i.e.,  $\mathbf{x}_\beta$  is within  $\Omega_d$ . Otherwise,  $\mathbf{x}_\beta$  is an unsafe operating point, i.e.,  $\mathbf{x}_\beta$  is outside  $\Omega_d$ .

Equation (1) is given by rigorous theoretical derivation of physical models, and obtaining a visible representation of DSR for security assessment. Instead, data-driven neural networks are trained with large amounts of data to get general rules for security assessment. These rules can be regarded as an invisible representation of DSR. Therefore, there is an intrinsic connection between the physical-driven method and data-driven method. Neural network design can be empowered by the physical properties of DSR. The relationship described above is shown in Fig. 1.

### B. A Mathematical Description of Network Classification

Mathematically, neural networks can achieve arbitrary classifications. Clarifying the classification process of networks can facilitate design of an interpretable model based on physical properties of DSR in power systems. This paper takes a simple two-dimensional classification (as shown in Fig. 2(b)) as an example to illustrate the classification process

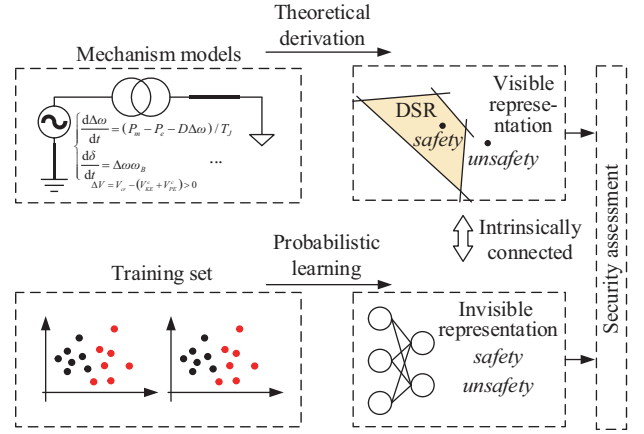


Fig. 1. Intrinsic correlation between physical-driven method and data-driven method.

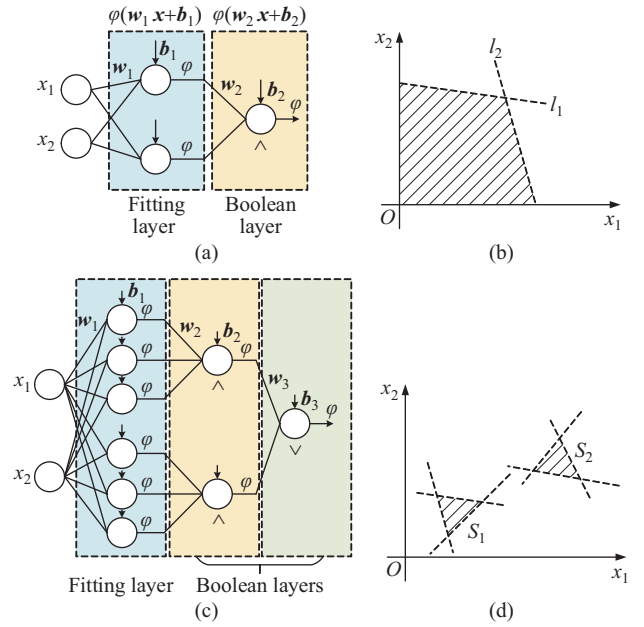


Fig. 2. Schematic diagram of network classification. (a) Schematic diagram of the network structure to realize the classification of (b). (b) A classification with two surfaces. (c) Schematic diagram of the network structure to realize the classification shown in (d). (d) A classification with six surfaces and two subsets.

of neural networks. Network structure can be designed as shown in Fig. 2(a), where  $\varphi$  is an activation function,  $w$  and  $b$  are weights and biases, respectively. The first layer can be described as:

$$\begin{pmatrix} y_1^1 \\ y_2^1 \end{pmatrix} = \varphi \left( \begin{pmatrix} w_1^{11} & w_1^{12} \\ w_1^{21} & w_1^{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1^1 \\ b_1^2 \end{pmatrix} \right) \quad (2)$$

If  $\varphi$  is a step function, a set of decision boundaries can be determined by  $w_1 \cdot x + b_1$ . Weight vectors  $(w_1^{11}, w_1^{12})$  and  $(w_1^{21}, w_1^{22})$  correspond to boundary normal vectors, i.e., normal vectors of  $l_1$  and  $l_2$  in Fig. 2(b). To determine the shaded region in Fig. 2(b), just the intersection of the sub-regions enclosed by  $l_1$  and  $l_2$  needs to be taken. That is, the role of the second layer in Fig. 2(a) is for intersection operation. For convenience, the first layer determining decision boundaries is called the Fitting Layer (FL), the second layer performing

Boolean operations is called the Boolean Layer (BL), and each neuron in the FL is called a Channel.

To achieve a more complex classification, we can add more FL channels and BLs, as shown in Fig. 2(c) and (d). For classification, the top three neurons of the FL are employed to determine three surfaces of subset  $S_1$ , and the bottom three neurons to determine three surfaces of  $S_2$ . The first BL is employed to determine subsets  $S_1$  and  $S_2$ , respectively, and the second BL to take the union of subsets  $S_1$  and  $S_2$ .

Based on the above analysis, the constructing process of a classification neural network can be divided into two parts: 1) constructing the FL to describe the decision boundary; 2) and constructing the BL to perform Boolean operations on the decision boundary determined by the FL. For a specific classification task in physically embodied systems, we can construct an interpretable neural network according to specific physical properties.

### C. DSR Boundaries and Determination of FL

For a given fault set, if all machines in an unstable system are equivalently divided into two clusters: stable machine cluster and unstable machine cluster, the DSR boundary can be described as:

$$\begin{aligned} P_{\text{emax}} &= \frac{1}{M_T} \left[ \underbrace{M_A, \dots, M_A}_{\#\{B\}}, \underbrace{-M_B, \dots, -M_B}_{\#\{A\}} \right] \cdots \\ &\quad [\mathbf{P}_{ei}, \mathbf{P}_{ej}]^T \quad (i \in B, j \in A) \\ &= \mathbf{K} \cdot \mathbf{P}_{\partial\Omega} \end{aligned} \quad (3)$$

where  $\#\{A\}$  and  $\#\{B\}$  are number of generators in stable machine cluster  $A$  and unstable machine cluster  $B$ , respectively,  $\mathbf{P}_{ei}$  and  $\mathbf{P}_{ej}$  are electromagnetic power of generators in stable and unstable clusters, and they are  $\#\{B\}$  and  $\#\{A\}$  dimensional vectors, respectively,  $M_B$ ,  $M_A$ , and  $M_T$  are inertia constants of stable cluster  $B$ , unstable machine cluster  $A$ , and all machines in the system, respectively:  $M_B = \sum_{i \in B} M_i$ ,  $M_A = \sum_{j \in A} M_j$ ,  $M_T = M_A + M_B$ ,  $P_{\text{emax}}$  is upper limit of equivalent power to ensure transient stability,  $\mathbf{K}$  is normal vector of the DSR boundary,  $\mathbf{P}_{\partial\Omega}$  is operating point on the DSR boundary.

Equation (3) indicates the DSR boundary consists of a set of hyperplanes with different unstable modes. According to the classification process of neural networks (i.e., (2)), the FL channels need to fit the hyperplanes one by one. The so-constructed neural network can be decoupled by the unstable mode and gives a specific physical meaning to one/all modules. If we want to study all unstable models of two-cluster types, theoretically, the minimum number of FL channels that can describe every unstable mode is  $C_N^1 + C_N^2 + \dots + C_N^{N-1}$ , where  $N$  is total number of generators in the studied power system, and  $C$  is the combinatorial number. In real systems, there is no corresponding boundary for some unstable modes, or such boundary is substituted by a more strict one. Thus, only hyperplanes of interest need to be modeled for FL channels. In this paper, unstable mode of power systems can be determined based on the simulation curve when preparing the training set, which provides a reference for interested hyperplanes.

### D. Determination of BL

Generally, the DSR must satisfy all security constraints. As shown in Fig. 3(a), the DSR is intersection of all security sub-regions determined by surfaces  $l_1$ ,  $l_2$ ,  $l_3$ , and  $l_4$ . In the Karnaugh map, the DSR is the blue area shown in Fig. 3(b) (where 0 represents *safety*, 1 represents *unsafety*). In this case, the BL needs only one Boolean operation (OR operation) on the results of FL to determine final DSR. Interestingly, since the function of the BL is pre-determined, parameters  $w_2$  and  $b_2$  of the BL can be determined directly without the backpropagation algorithm to adjust. For example, if activation functions of the FL are step functions (similarly, 0 represents *safety*, 1 represents *unsafety*), and final outputs of BL,  $y < T_h$  and  $y > T_h$  ( $T_h$  is threshold of final classification), represent *safety* and *unsafety*, respectively,  $w_2$  and  $b_2$  can be directly determined as:

$$\begin{cases} w_2 = \underbrace{(1, 1, \dots, 1)}_{N_c}^T \\ b_2 \in (T_h - 1T_h) \end{cases} \quad (4)$$

where  $N_c$  is channel number of the FL. In practice, the step function is not used as activation functions, generally. If ReLU function is adopted and  $w_2$  is set using (4),  $b_2$  will be close to zero, which can still represent OR operation.

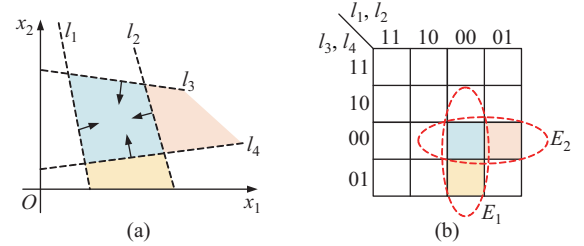


Fig. 3. Schematic of SR and Karnaugh map. (a) Schematic of SR with four surfaces  $l_1$ ,  $l_2$ ,  $l_3$ , and  $l_4$ , where the arrow side is safe. (b) Karnaugh map of (a), where 0 represents *safety*, and 1 represents *unsafety*.

When considering control measures of power systems (e.g., power-transfer schemes), convexity of the security region may change. As shown in Fig. 3, if a control measure is adopted to release restriction on  $l_4$ , the security region extends to the blue and yellow areas in Fig. 3(a). In this case, the security region is still a convex set and the area in the Karnaugh map can be enclosed by one loop  $E_1$ . That is, the security region can still be determined by one BL. If two control measures are adopted in a power system, and another control measure functions to release restriction on  $l_2$ , the security region extends to the union of blue, yellow, and red areas, and the corresponding area in the Karnaugh map is enclosed by  $E_1$  and  $E_2$  in Fig. 3(b) and cannot be further simplified. In this case, the BL needs two layers, where the function of the first layer is an OR operation and the second layer is an AND operation. Parameters of the second layer  $w_3$  and  $b_3$  can be determined as:

$$\begin{cases} w_3 = \underbrace{(1, 1, \dots, 1)}_{N_a}^T \\ b_3 \in (T_h - N_a, T_h + 1 - N_a) \end{cases} \quad (5)$$

where  $N_a$  is number of subsets determined by OR operation layer, and in Fig. 2(c),  $N_a = 2$ .

In summary, if the security region is convex, the BL of the network needs at least one layer; if control measures are considered, especially multiple control measures are adopted, the security region may be changed to a concave region and the BL needs at least two layers. As DSR of transmission grids is considered generally a region of absolute safety requiring no control measures, only weights of BL from (4) is considered in this paper.

### E. Construction of Loss Function

Loss function of networks is another important factor to improve training efficiency. Generally, loss function is determined by the difference between sample label  $y_{\text{label}}$  and network prediction  $y_{\text{pred}}$  [20]. For fitting or classification tasks, loss function usually adopts mean square error or cross-entropy loss, respectively:

$$\text{loss} = \begin{cases} \sum (y_{\text{pred}} - y_{\text{label}})^2 \\ -\sum (y_{\text{label}} \log y_{\text{pred}} + (1 - y_{\text{label}}) \log(1 - y_{\text{pred}})) \end{cases} \quad (6)$$

DSR describes a set of operating points [17] while including both information of each operating point and geometric characteristic. Nevertheless, construction of the loss function in (6) is based only on information of operating points, which can hardly reflect geometric characteristics of DSR.

In (3),  $\mathbf{K}$  is normal vector of the DSR boundary and can be directly determined by the unstable mode and generator's inertia constant. If a network is built according to principles of Sections II.B–II.D, weights of  $l^{\text{th}}$  FL channel  $\mathbf{w}_l$  describe the normal vector of learned decision boundary. Thus,  $\mathbf{w}_l$  and  $\mathbf{K}$  should be collinear. Based on (6), improved loss function considering the DSR geometric characteristics is:

$$\text{loss}_{\text{imp}} = \text{loss} - (\kappa \log n/n) \sum_l |\mathbf{w}_l \cdot \mathbf{K}| \quad (7)$$

where the 2<sup>nd</sup> term is Geometric Loss (GL) related to weights of  $l^{\text{th}}$  FL channel and normal vector of DSR boundaries.  $|\mathbf{w}_l \cdot \mathbf{K}|$  obtains maximum value when  $\mathbf{w}_l$  is parallel to  $\mathbf{K}$ . To obtain optimal value of  $\text{loss}_{\text{imp}}$  along the direction of gradient descent, a negative sign is added in front of  $|\mathbf{w}_l \cdot \mathbf{K}|$ , which is meant to make  $\mathbf{w}_l$  and  $\mathbf{K}$  as parallel as possible.  $\kappa \log n/n$  is coefficient of the GL, where  $n$  is number of training samples. It should be noted that if  $n$  is large enough to fit the DSR boundary exactly, there is no need to consider the role of the GL. At this point, (7) is simplified to (6) with  $\lim_{n \rightarrow +\infty} \kappa \log n/n \rightarrow 0$ .

Although the DSR is described in vector space spanned by  $\mathbf{x}_\beta$ , more input features other than  $\mathbf{x}_\beta$  can be input to networks. Function of FL can be reformed as:

$$\mathbf{y} = \varphi((\mathbf{w}_\beta, \mathbf{w}_\alpha) \cdot (\mathbf{x}_\beta, \mathbf{x}_\alpha)^T + (\mathbf{b}_\beta, \mathbf{b}_\alpha)) \quad (8)$$

where  $\mathbf{x}_\alpha$  is features of the power system other than  $\mathbf{x}_\beta$ . When constructing  $\text{loss}_{\text{imp}}$ , only the relationship between  $\mathbf{w}_\beta$  and  $\mathbf{K}$  needs to be considered, that is,  $\mathbf{w}_l$  is replaced by  $\mathbf{w}_\beta$  in (7).

## III. CREDIBILITY INDEX OF NEURAL NETWORK PREDICTIONS

In application, network prediction can only answer “yes” or “no”, and cannot provide a “good” or “bad” credibility measure for predictions. If an unsafe operating point is hidden in the learned DSR, it may affect stability of the whole power system. Therefore, setting a CI for network predictions can overcome the above problems [21].

### A. Credibility Index of Predictions

#### 1) Probability Factor

In general, samples with the same classification label are similar, so distance between them will be relatively small. In this paper, the probability factor is established using  $k$ -nearest neighbors' algorithm. In the neighborhood of a test sample  $X_p$ , proportions of samples with different classes are calculated to assess credibility. Before any details, we need to clarify the following two concepts:

- 1)  $k$ -order neighborhood  $\Theta_k(p)$  of  $X_p$ :  $\Theta_k(p)$  contains at least  $k$  samples in the neighborhood of  $X_p$ .
- 2)  $k$ -order nearest neighbor distance  $D_k(p)$  of  $X_p$ : radius of  $\Theta_k(p)$ . When Euclidean Distance is adopted,  $D_k(p)$  is calculated as:

$$D_k(p) = \max_{q \in \Theta(p)} (D(p, q)) = \max_{q \in \Theta(p)} \left( \sqrt{\sum_{i=1}^m (x_{pi} - x_{qi})^2} \right) \quad (9)$$

s.t.  $\begin{cases} \#\{D(p, q) \leq D_k(p)\} \geq k \\ \#\{D(p, q) < D_k(p)\} \leq k - 1 \end{cases}$

where  $x_{pi}$  and  $x_{qi}$  are the  $i^{\text{th}}$  feature of  $X_p$  and  $X_q$ , respectively,  $D(p, q)$  is Euclidean Distance between  $X_p$  and  $X_q$ .

Calculation of  $D_k(p)$  must ensure that: 1) at least  $k$  samples in the neighborhood  $\Theta_k(p)$  holds that  $D(p, q) \leq D_k(p)$ , 2) and at most  $k - 1$  samples  $X_p$  in  $\Theta_k(p)$  holds that  $D(p, q) < D_k(p)$ . An example of  $\Theta_5(p)$  is shown in Fig. 4.

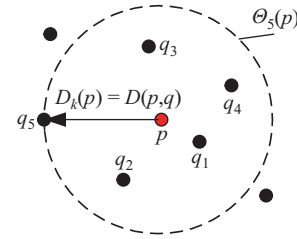


Fig. 4. The 5-order neighborhood of  $X_p$ .

According to the algorithmic randomness theory [22] and the classification label, probability factor  $\eta_k$  is calculated as:

$$\eta_k = \#\{S_0 \cap \Theta_k(p)\} / \#\{\Theta_k(p)\} \times 100\% \quad (10)$$

where  $S_0$  is the set with a *safety* label. It is noticed the denominator of (10) cannot be directly replaced by  $k$  since there are at least  $k$  points in  $\Theta_k(p)$ , i.e.,  $\#\{\Theta_k(p)\} \geq k$ .

## 2) Density Factor

In practice, samples are vulnerable to environmental contamination and attacks. It is impossible to distinguish samples with abnormal distribution, such as outlier points, with  $\eta_k$  only. Therefore, this sub-section proposes a density factor based on local relative density to overcome this limitation. First, the following concepts need to be defined:

a) Reachability distance:  $k$ -order reachability distance from  $X_p$  to  $X_o$  is the maximum of  $D_k(p)$  and  $D(p, o)$ :  $RD_k(p, o) = \max\{D_k(o), D(p, o)\}$ . Refer to Appendix A for advantages of using reachability distances.

b) Local reachability density:  $k$ -order local reachability density of  $X_p$  is the reciprocal of the mean of  $RD_k(p, o)$  in  $\Theta_k(p)$ , i.e.:  $LD_k(p) = \#\{\Theta_k(p)\} / \sum_{o \in \Theta_k(p)} RD_k(p, o)$ .

c) Local relative density:  $k$ -order local relative density of  $X_p$  is the ratio of  $LD_k(p)$  to the mean of  $LD_k(o)$  in  $\Theta_k(p)$ , i.e.:  $LRD_k(p) = LD_k(p) \cdot \#\{\Theta_k(p)\} / \sum_{o \in \Theta_k(p)} LD_k(o)$ .

The meaning of  $LRD$  is: 1)  $LRD_k(p) \approx 1$  indicates local density of  $X_p$  is similar to neighbors in  $\Theta_k(p)$ ; 2)  $LRD_k(p) < 1$  indicates  $X_p$  is in a relatively sparse area; 3) and  $LRD_k(p) > 1$  indicates  $X_p$  is in a relatively dense region.

## 3) Credibility Index $\alpha$

According to (10) and  $LRD$ , the  $k$ -order CI of  $X_p$  can be described as:

$$\alpha_k = LRD_k \eta_k \quad (11)$$

which is used to evaluate credibility of network predictions. Here,  $\eta_k$  and  $LRD_k$  can be interpreted as follows:

a) A larger value of  $\eta_k$  indicates there are more samples with *safety* labels in  $\Theta_k(p)$ , giving us more reason to believe the label of  $X_p$  is also safe;

b) A larger value of  $LRD_k$  indicates sample density in  $\Theta_k(p)$  is higher, i.e., neighbors of  $\Theta_k(p)$  are closer to  $X_p$ , making the calculated value of  $\eta_k$  more trustworthy.

In this paper,  $LRD_k(p) > 0.8$  means density around  $X_p$  is acceptable. Thus, value of  $LRD_k(p)$  is set to 1 in case of  $LRD_k(p) > 0.8$ , to ensure value of  $\alpha_k$  will not be greater than 1.

The above analysis explains how the CI is constructed from a mathematical view, and inherently also has a clear physical meaning. Detailed physical meaning is presented in Section IV combined with DSR.

## B. Credibility Evaluation Mechanism Incorporating Transductive and Inductive Methods

Although the transductive method derived CI can be used to predict classification of new test samples, it will fail once test samples are outside the range of the initial training set, such as shown in Fig. 6(a). In contrast, the neural network introduced in Section II, essentially an inductive learning method, can reason from observed training samples to general rules, which are then applied to test samples. Thus, network (inductive method) can be used to learn general rules from the initial training set, and then combined with calculation of CI (transductive method) to give a final prediction. Advantages of incorporating transductive and inductive methods include:

1) obtained CI is based on a different mathematical method from neural networks, which confirms independence of CI calculation; 2) and process of transductive method is transparent and outcome is interpretable. The proposed credibility evaluation mechanism in this paper is shown in Fig. 5.

Workflow as shown in Fig. 5 consists of four steps and is explained in Table I. First, the meaning of Step 2 needs to be explained specifically.

1) Expansion of  $TS_0$ : to calculate the CI of a test sample using (11), there must be other reference samples surrounding the test sample. If test sample is located within  $TS_0$ , the CI can be calculated directly. But (11) will fail when the test sample is located outside  $TS_0$ , such as the blue area in Fig. 6(a). Thus, the trained network in Step 1 is used as a label generator for the new sample region with similar distribution and density as  $TS_0$ . Expanded sample set is shown as  $TS_1$  in Fig. 6(b).

2) Correction of  $TS_0$ : if sample labels in  $TS_0$  are tampered with, as the error sample in Fig. 6(a), it will significantly affect probability factor calculation of (10). Nevertheless, the trained network in Step 1 can be used to identify and correct error samples [23]. First, the trained network is used to obtain  $TS'_0$  by label correction of  $TS_0$ , and then CL is calculated in  $TS'_0$ . The correction approach is well adapted for dirty samples with low quality.

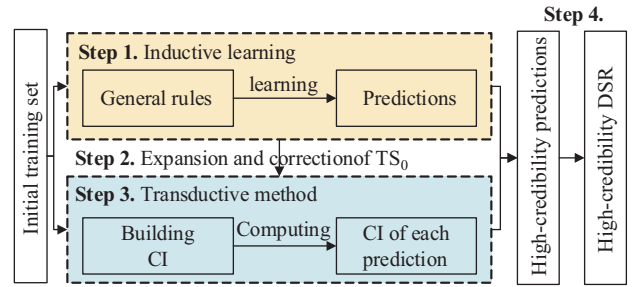


Fig. 5. Credibility evaluation mechanism incorporating transductive and inductive methods.

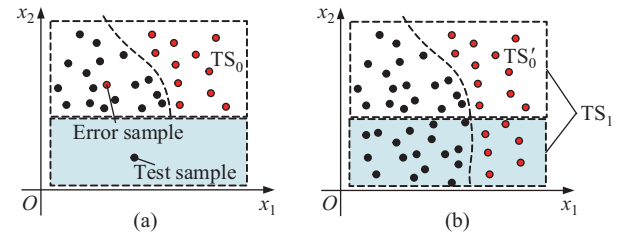


Fig. 6. Schematic diagram of a sample set expansion and correction. (a) The initial sample set. (b) The sample set after expansion and correction.

TABLE I  
WORKFLOW OF CREDIBILITY EVALUATION MECHANISM

|                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------|
| <b>Require:</b> Obtain the initial training set $TS_0$ and construct a neural network with the approach proposed in Section II. |
| <b>Step 1.</b> Training the network (or other inductive learning methods) on $TS_0$ ;                                           |
| <b>Step 2.</b> Using the trained network to expand and correct $TS_0$ , and getting $TS_1$ ;                                    |
| <b>Step 3.</b> Calculating CI on $TS_1$ according to (11);                                                                      |
| <b>Step 4.</b> Incorporating network predictions and obtained CI, and getting a high-credibility DSR.                           |

#### IV. PHYSICAL PROPERTIES AND INTERPRETATION

This section introduces characteristics of DSRs to help readers understand the physical meaning of the proposed network design principles and the CI. Rigorous theoretical derivation and technical details of DSR are studied in [24]. In this section, we introduce only a few important equations and properties for further physical interpretation.

##### A. Properties of DSR

Transient energy function methods provide a measurement criterion for stability of power systems. If system is stable, potential energy  $V_d$  should be greater than kinetic energy  $V_{KE}^c$ , i.e.,  $\Delta V = V_d - V_{KE}^c > 0$ . In the power injection space, the relationship between transient energy and distance from the operating point to the DSR boundary  $S$  is described as:

$$\begin{cases} S = (\mathbf{K} \cdot \mathbf{P}_{\partial\Omega} - \mathbf{K} \cdot \mathbf{P}) / \|\mathbf{K}\| \\ d\Delta V/dS > 0 \end{cases} \quad (12)$$

According to the One-Machine Infinite Bus equivalence, normal vector of DSR boundaries  $\mathbf{K}$  can be calculated by [24]:

$$\mathbf{K} = \frac{1}{M_T} \left[ \underbrace{M_A, \dots, M_A}_{\#\{B\}}, \underbrace{-M_B, \dots, -M_B}_{\#\{A\}} \right] \quad (13)$$

Within the DSR, (12) indicates the operating point farther from DSR boundaries would be safer. If the operating point is exactly on the DSR boundary, they have the following relationship:

$$\mathbf{K} \cdot \mathbf{P}_{\partial\Omega} = P_{\text{emax}} \quad (14)$$

According to (12)–(14), we further obtain the following three properties [24]:

**Property 1:** DSR boundaries consist of a set of hyperplanes, and different unstable modes correspond to different hyperplanes.

**Property 2:** Vector  $\mathbf{K}$  can be used to quickly calculate normal vector of DSR boundaries.

**Property 3:** For a given unstable mode, each level set of  $\Delta V$  forms a cluster of hyperplanes parallel to the DSR boundary, and within the DSR, the farther the operating point is from DSR boundaries, the safer it is.

##### B. Physical Interpretation

The DSR boundary described by (14) is a set of hyperplanes, and  $\mathbf{K}$  is normal vector of the hyperplane. For a given grid structure, different  $\mathbf{K}$  corresponds to different unstable modes. Thus, if FL channels in Section II are used to fit decision hyperplanes, they should correspond to unstable modes one-to-one (Property 1).

Since  $\mathbf{K}$  and weight of FL channels,  $w_l$  represents the normal vector of the DSR boundary and learned decision boundary, respectively, vectors  $w_l$  and  $\mathbf{K}$  should be collinear. That is, 2<sup>nd</sup> term of loss function (7) is constructed by geometric characteristics of DSR boundaries. Besides, approximate calculation of  $\mathbf{K}$  is only related to unstable mode and inertia constant of generators, which ensures  $\mathbf{K}$  is computable during generation of training samples, thereby ensuring  $\mathbf{K}$  can take part in the training process of the neural network (Property 2).

In calculation of CI, the probability factor of a test sample  $X_p$  is determined based on its neighbors within  $\Theta(p)$ . According to Property 3,  $\Delta V$  is monotonically decreasing along the  $\mathbf{K}$  direction within  $\Theta(p)$  (As the circles shown in Fig. 7, red indicates a larger value of  $\Delta V$ ). Generally,  $X_p$  is located in the center of  $\Theta(p)$ .  $\Delta V$  of  $X_p$  must be greater than zero as long as samples on both sides of  $\mathbf{K}_\perp$  satisfy  $\Delta V > 0$ , i.e.,  $X_p$  must be a safe operating point. In other words, using (11) to calculate CI is not only an aggregation of similar data points in a mathematical sense, but also approximates the upper and lower bounds of  $\Delta V$  of  $X_p$  are used for security assessment in a physical sense (with a certain inevitability).

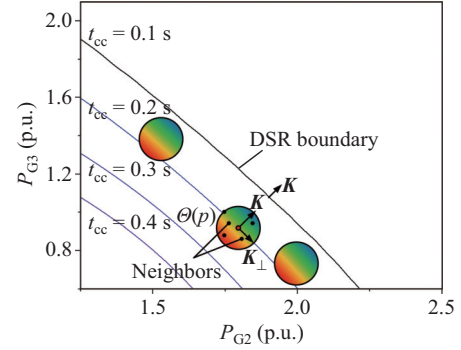


Fig. 7. Critical clearing time  $t_{cc}$  of 3-generator 9-bus system (Fig. 8(a)) in the power injection space of  $P_{G2}$  and  $P_{G3}$ .  $P_{G2}$  and  $P_{G3}$  represent the active power of  $G_2$  and  $G_3$ , respectively. Since  $t_{cc}$  can be easily obtained by simulation, we use  $t_{cc}$  of the studied faults to represent the stability margin  $\Delta V$ . A larger  $t_{cc}$  means a larger  $\Delta V$ .

#### V. CASE STUDY

In this section, validity of the proposed method is demonstrated through simulations on both the IEEE 9-bus system and a 300-bus AC/DC hybrid system. These tests are carried out on a computer equipped with an Intel Core i7-6700 processor (3.4 GHz) and 16 GB of memory. Simulations are performed using PSD-BPA power system transient stability analysis program, which is released by the China Electric Power Research Institute.

##### A. IEEE 9-bus System

###### 1) FL Channels and BL Weights

As shown in Fig. 8(a), fault set of interest is  $(N - 1)$  contingency for all transmission lines, and fault duration is 0.1 s. Training samples are obtained by changing power of generators from 0.15 p.u. to 2.15 p.u. (600 samples in total). Sample features contain active power, voltage amplitude, and phase angle, and sample label is determined by maximum angle deviation, represented by  $\Delta\delta_{\text{max}}$ , between any two generators after a fault [6], [25]. If the studied fault set contains  $n_f$  faults, the label of the sample  $y_{\text{Label}}$  is determined by taking logical OR operation of results for each fault, as shown in (15), where 0 and 1 represent *safety* and *unsafety*, respectively.

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

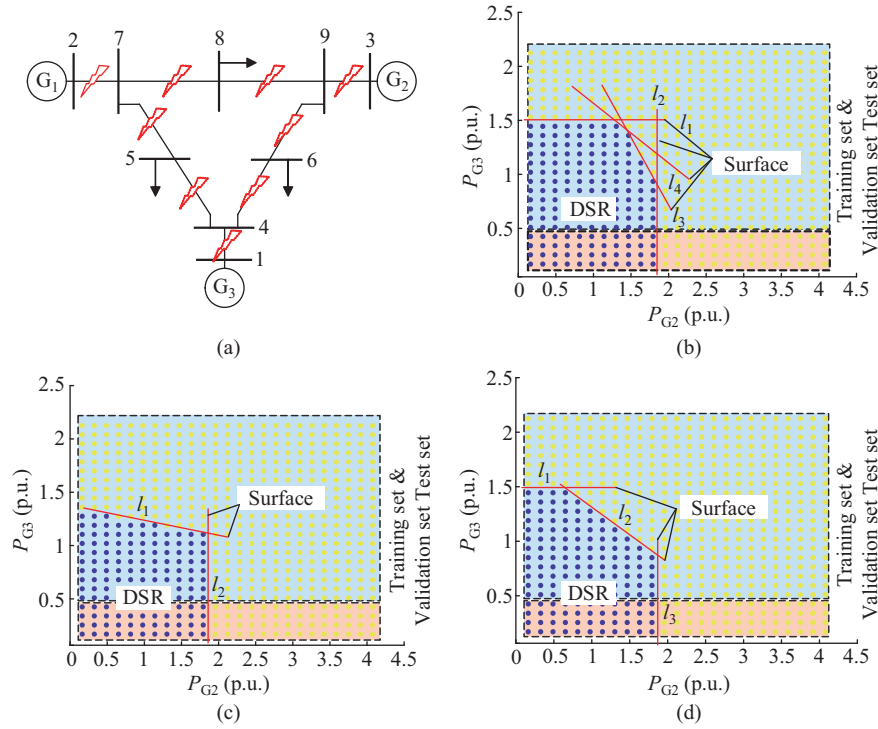


Fig. 8. Training results of networks with different numbers of FL channels. (a) Schematic of IEEE 9-bus system. (b) Simulation results in the space of  $P_{G2}$  and  $P_{G3}$ , i.e., the prepared sample set. (c) Training result of the network with two FL channels. (d) Training result of the network with three FL channels.

$$y_i = \begin{cases} 0, & \text{TSI} > 0 \\ 1, & \text{TSI} < 0 \end{cases}$$

$$y_{\text{Label}} = y_1 \vee y_2 \vee \dots \vee y_i \vee \dots \vee y_{nf} \quad (15)$$

To facilitate visualization, obtained samples in power injection space  $P_{G2}$  and  $P_{G3}$  are presented in Fig. 8(b), where blue points indicate *safety* and yellow one's *unsafety*. In Fig. 8(b), the DSR boundary is approximated to four hyperplanes (i.e.,  $l_1$ ,  $l_2$ ,  $l_3$ , and  $l_4$ ). In the upper sample set (with blue background), 80% and 20% of samples are used as the training set and validation set, respectively. Samples in the lower sample set (with purple background) are used as test set, which ensures test samples are not within distribution of the training set. The FL and BL of networks are constructed according to the principles proposed in Section II, and gradient descent is used for training the network with a learning rate of 0.01.

The relationship between network structure and DSR boundary is verified by varying number of FL channels. A typical training result is shown in Fig. 8(c) when FL channel number is two. The learned decision boundary contains two obvious surfaces. When FL channel number increases to three, the learned decision surfaces increase from two to three accordingly, as shown in Fig. 8(d), which indicates FL channels correspond to learned surfaces (i.e., hyperplane of the learned DSR boundary).

During the training process, activation function of the FL is ReLU, weights of the BL are set to one based on (4) and not adjustable in the gradient descent process, bias  $b$  of the BL is adjustable, and  $T_h$  is set to 0.5. After training, values of BL bias  $b$  are recorded in Table II. It can be found  $b \in (-0.5, 0.5)$  and  $|b|$  is close to zero, which is consistent with analysis

in Section II-D: only when outputs of all FL channels are 0 (*safety*), DSR is obtained or, more precisely, BL performs the OR operation.

## 2) Effectiveness of GL

To illustrate effectiveness of GL, we train the two-channel and three-channel networks 200 times respectively, where each training process contains 300 epochs, and regularization by early stopping is adopted. We record number of times when accuracy of the two networks is above 85% and 90%, respectively.

TABLE II  
TRAINING RESULTS OF BL BIAS  $b$

| Channel numbers | Training No. |        |         |        |         |
|-----------------|--------------|--------|---------|--------|---------|
|                 | 1            | 2      | 3       | 4      | 5       |
| 2 channels      | -0.1191      | 0.0950 | -0.1312 | 0.0437 | -0.1815 |
| 3 channels      | 0.0334       | 0.0056 | -0.0619 | 0.1026 | -0.0003 |

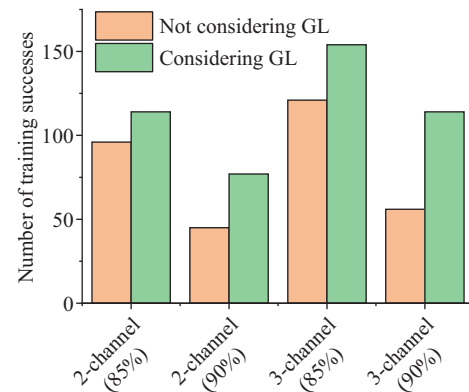


Fig. 9. Number of training successes considering GL vs. not.

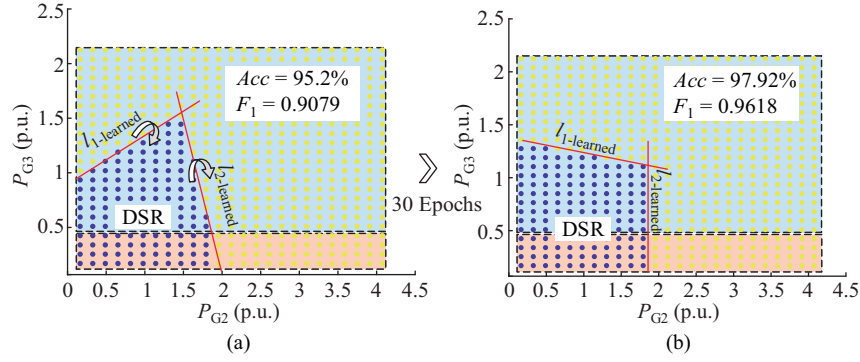


Fig. 10. Effect on the network decision boundary after considering the GL. (a) The network training results when the loss function does not contain GL. (b) Based on (a), the network training results after another 30 epochs when the loss function contains GL.

tively, as shown in Fig. 9. Comparison reveals considering GL in loss function can significantly improve number of training successes.

To further illustrate the mechanism of GL on performance improvement, we split the training process of the two-channel network. When loss function does not contain GL (i.e. (6)), a training result is shown in Fig. 10(a), with an accuracy of 95.2% and an  $F_1$ -score of 0.9079. On this basis, GL is added to loss function (i.e. (7)), and after another 30 epochs, training results are shown in Fig. 10(b). It can be found the learned surfaces  $l_{1\text{-learned}}$  and  $l_{2\text{-learned}}$  rotate along the arrow direction, as shown in Fig. 10(a), and become increasingly parallel to actual DSR boundary ( $l_1$  and  $l_2$  in Fig. 8(b)). Meanwhile, accuracy and  $F_1$  scores are improved to 97.92% and 0.9618, respectively.

### 3) Demand for Sample Size

To verify demand for training samples, we gradually increased number of samples (10 samples per stage), and Fig. 11 records average accuracy of 20 training sessions of the physical mechanism enabled neural network (PMENN) and a traditional ANN. It can be found PMENN can achieve higher accuracy with fewer samples, and this advantage is especially obvious when number of samples is less than 200, which is mainly due to guidance of DSR's prior knowledge on training of the intelligent model.

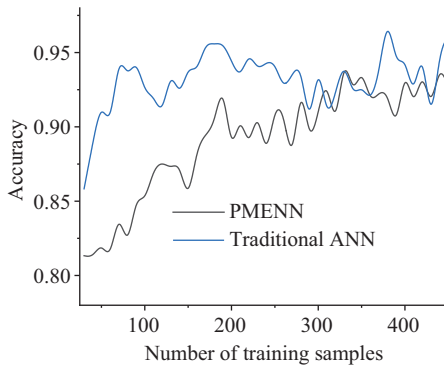


Fig. 11. Comparison of training effects with different sample sizes.

Changing operating scenario: Scenario 1, inertia of generator G2 is increased 1.5 times; Scenario 2, line 5~7 and line 7~8 are changed to double-circuit lines. Table III records

TABLE III  
COMPARISON OF REQUIRED SAMPLES FOR TRANSFER LEARNING UNDER DIFFERENT SCENARIOS

| Model | Scenario 1 | Scenario 2 |
|-------|------------|------------|
| PMENN | < 40       | < 40       |
| ANN   | > 90       | > 90       |
| FSNN  | > 150      | > 150      |
| CNN   | > 120      | > 150      |

number of samples required for different intelligent models to achieve an accuracy of 95% or higher when transferring to a new scenario, based on 5 trials. Results show that transfer learning can enable intelligent models to adapt to new scenarios again. However, in terms of number of samples required, PMENN requires the least number of new samples. This is mainly because PMENN integrates the physical properties of DSR, which can provide guidance for training, and the model has fewer parameters, so it can perform well from fewer samples. In comparison, structure of the CNN and feature-separated neural network (FSNN) [26] is more complex and has more parameters to be optimized, resulting in a noticeable increase in number of new samples required.

### 4) Effectiveness of CI

This sub-section verifies effectiveness of the proposed credibility evaluation mechanism (as shown in Fig. 5). Based on the trained network with three FL channels in Section V.0.1, test samples randomly selected in space of  $P_{G2}$  and  $P_{G3}$  are input to the network for security assessment. Incorporating network predictions, the CI is calculated by (11) with  $k = 5$ . Results are shown in Fig. 12(a), where points in yellow indicate operating points are more likely to be *safety*, and points in dark blue indicate operating points are more likely to be *unsafety*. Points in other colors form an untrusted band, which is usually parallel to the DSR boundary. DSR with  $k$ -order CI not lower than  $\alpha_T$  is denoted as  $\Omega_{k-\alpha_T}$ , that is,  $\Omega_{k-\alpha_T} = \{y_{\text{pred}} = 0 \wedge \alpha_k \geq \alpha_T\}$ . In  $\Omega_{k-1}$ , accuracy of network predictions for *safety* is described as:

$$\xi = \#\{S_{P\text{-safety}} \cap S_{T\text{-safety}}\} / \#\{S_{P\text{-safety}}\} \times 100\% \quad (16)$$

where  $S_{P\text{-safety}}$  is the sample set that network prediction is *safety* and  $\alpha_k$  equals 1,  $S_{T\text{-safety}}$  is the actual *safety* sample set.

Variation of  $\xi$  with  $k$  is shown as the black curve in Fig. 12(b). It can be seen that  $\xi$  increases with  $k$ . When  $k = 5$ ,  $\xi$  is equal to 100%. In this case, there is no misclassification within obtained  $\Omega_{5-1}$ .

According to Fig. 7, a larger value of  $k$  means CI tends to push the decision boundary away from real DSR boundary (i.e., keeping a larger security margin). However, a larger  $k$  would cause a decrease in completeness of obtained  $\Omega_{k-1}$ , as the blue curve shown in Fig. 12(b). Completeness  $\zeta$  is the ratio of  $\Omega_{k-1}$  and real DSR, i.e.:

$$\zeta \propto \# \{S_{P\text{-safety}} \cap S_{T\text{-safety}}\} / \# \{S_{T\text{-safety}}\} \times 100\% \quad (17)$$

Setting an appropriate value of  $k$  and density of samples, completeness  $\zeta$  can meet engineering requirements.

### B. 300-bus AC/DC Hybrid System

300-bus AC/DC hybrid system, provided by China Electric Power Research Institute (CEPRI) based on actual operating parameters, is used to verify the proposed method in a large grid. System includes 93 generators (generation capacity: 26,857 MW), 168 loads (total load: 26,584 MW), and seven cross-area DC lines, as shown in Fig. 13(a). The PSD-BPA released by CEPRI is used to generate training samples (4961 samples in total), where total generations and loads fluctuate within 1000 MW. The fault set contains lines of bus 1A-7~bus 1A-12, bus 1A-20~bus 1B-2 (double-circuit line), and bus 1A-21~bus 1B-3 (double-circuit line) with  $(N - 1)$  contingencies, and fault duration is 0.1 s. Sample features contain

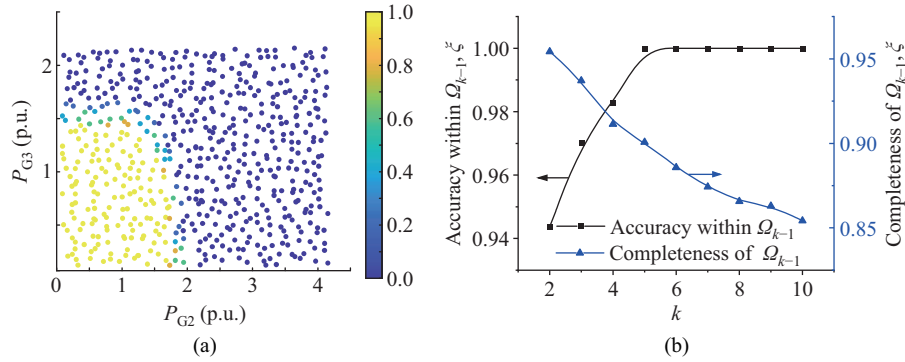


Fig. 12. Distribution of CI and performance. (a) The distribution of CI based on the credibility evaluation mechanism in Section III-B. (b) The accuracy and completeness of  $\Omega_{k-1}$  with different values of  $k$ .

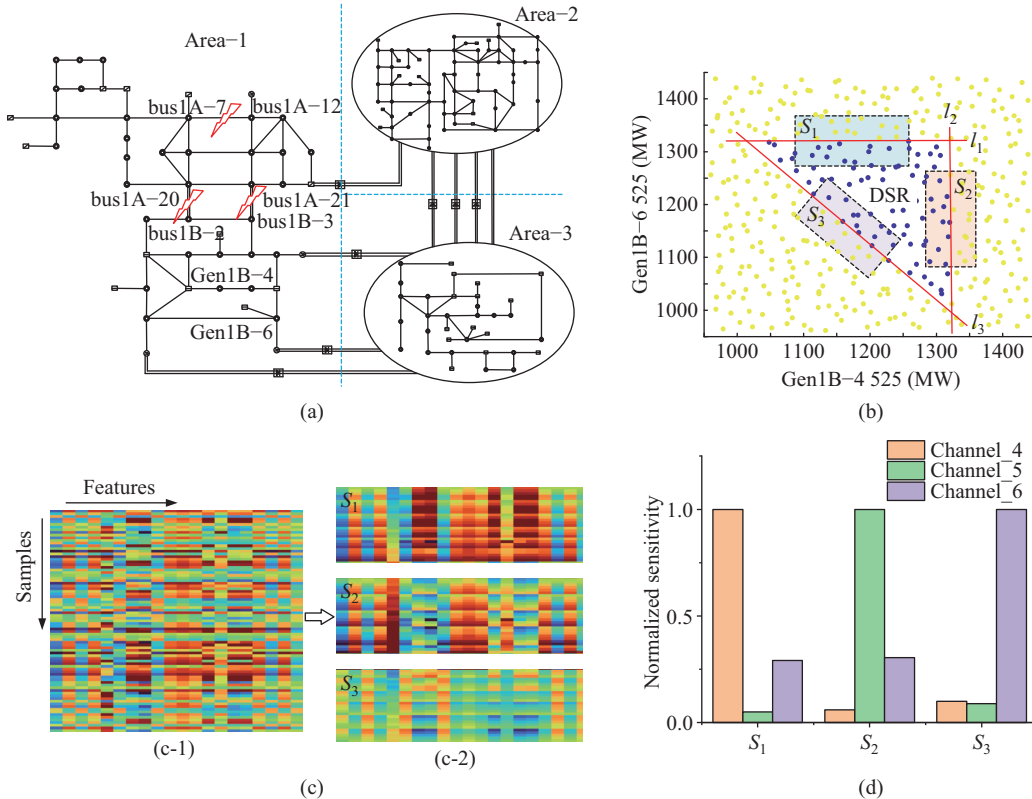


Fig. 13. 300-bus system and network interpretation results. (a) The schematic of the 300-bus power system. (b) The prepared test set of the 300-bus power system in the space of Gen 1B-4 and Gen 1B-6. (c) The heat map of the selected sample set. (d) The second-order sensitivity of channels 4, 5, and 6 in  $S_1$ ,  $S_2$ , and  $S_3$  (after normalization).

active power and reactive power, voltage amplitude and phase angle, and active power flow of lines. The neural network is constructed based on principles proposed in Section II with ten FL channels. The GL in the loss function (i.e., (7)) is constructed by weights of Channels 4, 5, and 6.

The constructed network is trained by the gradient descent method, and then the physical mean of each channel is verified by sensitivity analysis [27]. Fig. 13(b) shows the prepared test set in the space of Gen 1B-4 and Gen 1B-6. A batch of samples is taken from the prepared test set near the three boundaries (as the heat map in Fig. 13(c-1)). These samples are roughly divided into three subsets  $S_1$ ,  $S_2$ , and  $S_3$ , as shown in Fig. 13(c-2), which correspond to neighborhoods of  $l_1$ ,  $l_2$ , and  $l_3$ , respectively (Fig. 13(b)). Then, we analyze second-order sensitivity of each FL channel on the three subsets [27], where second-order sensitivities of Channels 4, 5, and 6 are as shown in Fig. 13(d). Results reveal Channels 4, 5, and 6 have highest sensitivity on  $S_1$ ,  $S_2$ , and  $S_3$ , respectively, which indicates the physical meaning of Channels 4, 5, and 6 corresponds to surfaces  $l_1$ ,  $l_2$ , and  $l_3$ , respectively. That is, design principles of neural network structures proposed in Section II can reflect correlation between FL channels and DSR boundaries, which enhances interpretability of neural networks in power system security assessment.

Table IV records training results of the neural network when GL is considered or not in loss function. Training process is similar to what is described in Section V-A. Table IV counts number of times network accuracy is above 85% and 90%, respectively. It is found considering GL can significantly improve training effect.

According to Fig. 5, CI is established by (11) ( $k$  is set as 5). Table V records network accuracy within obtained DSR when considering CI or not. Results show considering CI can effectively guarantee network predictions are not misjudged within  $\Omega_{5-1}$ . Although completeness of the obtained DSR may be reduced after considering CI, it can be maintained at around 90% by setting an appropriate value of  $k$ .

## VI. CONCLUSION

This paper has introduced the concept of incorporating physics-prior within the neural network framework to achieve better security assessment for power systems. Through the-

oretical analysis and simulation verification, the following attractive features can be summarized:

1) The neural network can be constructed according to physical properties of DSR. FL channels correspond to hyper-plane of DSR boundaries (unstable modes), and then predicted DSR is obtained by Boolean operations to FL's decision. For transmission grids, only one BL is required. This specific network structure enables us to better understand functions of networks while maintaining physical plausibility.

2) The proposed neural network design principles enable geometric characteristics of DSR to be incorporated into the neural network training process through loss function. Simulation results verify GL can improve performance of the neural network.

3) The credibility evaluation mechanism gives a CI for each network prediction while maintaining physical plausibility of CI. As we see, credibility evaluation effectively reduces misclassifications within obtained DSR to prevent drastic damage of power systems.

Although the proposed method in this paper focuses on DSR of transmission grids, there would not be any technical difficulties in extending this method to smart distribution grids [28] and integrated energy systems [29], as these systems have similar properties. Besides, the proposed methods and considerations in building a neural network with interpretable structures and credible predictions can provide guidance for machine intelligence applied in other industrial systems.

## APPENDIX A

Reachability distance allows sparse points to be further away from dense areas, thus making differentiation between areas of differing densities more pronounced. As shown in Fig. A1, the yellow area is the dense area (i.e., the neighborhood of  $p_2$ ), and the neighborhood of  $p_1$  and  $p_3$  are relatively sparse. When using Euclidean distance, distance between the two areas of  $p_1$  and  $p_2$  is  $D(p_1, p_2)$ . However, when using reachability distance (i.e.,  $RD_5(p_1, p_2)$ ), distance between the two areas is represented by the 5-order nearest neighbor distance of  $p_1$  (i.e.,  $D_5(p_1)$ ). It is clear that  $D_5(p_1)$  is larger than  $D(p_1, p_2)$ , indicating reachability distance allows for more separation between regions of differing densities. Similarly, reachability distance  $RD_5(p_2, p_3)$  is  $D_5(p_3)$ , and

TABLE IV  
TRAINING RESULTS OF NETWORK CONSIDERING GL OR NOT IN 300-BUS SYSTEM

| Minimum accuracy | Not considering GL | Considering GL |
|------------------|--------------------|----------------|
| 85%              | 101                | 122            |
| 90%              | 52                 | 96             |

TABLE V  
IMPACT OF CI ON THE OBTAINED DSR

| Training No. | Accuracy within the obtained DSR, (Not considering CI) | Accuracy within the obtained DSR, (Considering CI) | Completeness of $\Omega_{5-1}$ , $\zeta$ |
|--------------|--------------------------------------------------------|----------------------------------------------------|------------------------------------------|
| 1            | 96.5%                                                  | 100%                                               | 90.54%                                   |
| 2            | 98.1%                                                  | 100%                                               | 91.93%                                   |
| 3            | 93.8%                                                  | 100%                                               | 88.52%                                   |

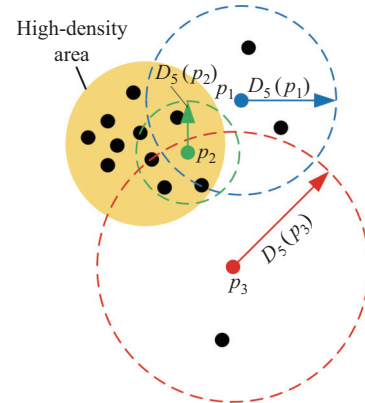


Fig. A1. Schematic diagram of reachability distance.

$D_5(p_3)$  is greater than  $D(p_2, p_3)$ , while reachability distance  $RD_5(p_1, p_3)$  is  $D(p_1, p_3)$ .

## REFERENCES

- [1] O. A. Alimi, K. Ouahada, and A. M. Abu-Mahfouz, "A review of machine learning approaches to power system security and stability," *IEEE Access*, vol. 8, pp. 113512–113531, Jun. 2020.
- [2] I. Ahmed, G. Jeon, and F. Piccialli, "From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022.
- [3] M. Y. Sun, I. Konstantelos, and G. Strbac, "A deep learning-based feature extraction framework for system security assessment," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5007–5020, Sep. 2019.
- [4] J. An, J. C. Yu, Z. H. Li, Y. B. Zhou, and G. Mu, "A data-driven method for transient stability margin prediction based on security region," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1060–1069, Nov. 2020.
- [5] T. J. Wang and Y. Tang, "Transient stability preventive control based on graph convolution neural network and transfer deep reinforcement learning," *CSEE Journal of Power and Energy Systems*, doi: 10.17775/CSEEJPES.2022.05030.
- [6] G. Z. Wang, J. B. Guo, S. C. Ma, X. Zhang, Q. L. Guo, S. X. Fan, and H. T. Xu, "Data-driven transient stability assessment using sparse PMU sampling and online self-check function," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 3, pp. 910–920, May 2023.
- [7] J. Lan, Q. L. Guo, Y. Z. Zhou, and H. B. Sun, "Generation of large-scale convergent power flow samples through a data-driven approach," in *Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration*, 2020, pp. 722–726.
- [8] 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 Transactions on Power Systems*, vol. 34, no. 6, pp. 5044–5052, Nov. 2019.
- [9] C. Ren, H. Yu, Y. Xu, and Z. Y. Dong, "Understanding Discrepancy of Power System Dynamic Security Assessment with Unknown Faults: A Reliable Transfer Learning-based Method," *CSEE Journal of Power and Energy Systems*, vol. 10, no. 1, pp. 427–431, Jan. 2024.
- [10] Q. Wang, F. Li, Y. Tang, and Y. Xu, "Integrating model-driven and data-driven methods for power system frequency stability assessment and control," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4557–4568, Nov. 2019.
- [11] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," in *Proceedings of the 5th International Conference on Learning Representations*, 2017.
- [12] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated Machine Learning: Methods, Systems, Challenges*, Cham: Springer, 2019.
- [13] M. Andrychowicz, M. Denil, S. G. Colmenarejo, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. de Freitas, "Learning to learn by gradient descent by gradient descent," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 3988–3996.
- [14] S. X. Fan, J. B. Guo, S. C. Ma, L. X. Li, G. Z. Wang, H. T. Xu, J. Yang, and Z. N. Zhao, "Framework and Key Technologies of Human-machine Hybrid-augmented Intelligence System for Large-scale Power Grid Dispatching and Control," *CSEE Journal of Power and Energy Systems*, vol. 10, no. 1, pp. 1–12, Jan. 2024.
- [15] M. Lutter, C. Ritter, and J. Peters, "Deep lagrangian networks: Using physics as model prior for deep learning," 2019, arXiv:1907.04490. [Online]. Available: <https://arxiv.org/abs/1907.04490>
- [16] L. N. Sun, H. Gao, S. W. Pan, and J. X. Wang, "Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data," *Computer Methods in Applied Mechanics and Engineering*, vol. 361, pp. 112732, Apr. 2020.
- [17] Y. X. Yu, Y. L. Liu, C. Qin, and T. K. Yang, "Theory and method of power system integrated security region irrelevant to operation states: an introduction," *Engineering*, vol. 6, no. 7, pp. 754–777, Jul. 2020.
- [18] Y. X. Yu and Y. J. Wang, "Security region based real and reactive power pricing of power system," *Science in China Series E: Technological Sciences*, vol. 51, no. 12, pp. 2095–2111, 2008.
- [19] Z. Y. Yue, Y. L. Liu, Y. X. Yu, and J. B. Zhao, "Probabilistic transient stability assessment of power system considering wind power uncertainties and correlations," *International Journal of Electrical Power & Energy Systems*, vol. 117, pp. 105649, May 2020.
- [20] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*, Cambridge: MIT Press, 2016.
- [21] Y. Xu, Z. Y. Dong, J. H. Zhao, P. Zhang, and K. P. Wong, "A reliable intelligent system for real-time dynamic security assessment of power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1253–1263, Aug. 2012.
- [22] A. Gammerman and V. Vovk, "Prediction algorithms and confidence measures based on algorithmic randomness theory," *Theoretical Computer Science*, vol. 287, no. 1, pp. 209–217, Sep. 2002.
- [23] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," 2018, arXiv:1802.06360. [Online]. Available: <http://arxiv.org/abs/1802.06360>
- [24] G. Z. Wang, Y. R. Jing, J. B. Guo, Q. L. Guo, S. C. Ma, and T. Z. Wang, "Concepts and characteristics of security margin based on dynamic security region," *International Journal of Electrical Power & Energy Systems*, vol. 143, pp. 108469, Dec. 2022.
- [25] G. Z. Wang, S. C. Ma, J. B. Guo, Y. R. Jing, H. T. Xu, and Z. H. Zhou, "Power system transient stability assessment and key bus set selection based on multiple-agent cooperation," in *Proceedings of 2021 Power System and Green Energy Conference*, 2021, pp. 40–44.
- [26] Z. H. Zhou, G. Q. Pu, S. C. Ma, G. Z. Wang, D. J. Shao, Y. P. Xu, and J. Dang, "Assessment and optimization of power system transient stability based on feature-separated neural networks," *Power System Technology*, vol. 45, no. 9, pp. 3658–3666, Sep. 2021.
- [27] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Computer Physics Communications*, vol. 181, no. 2, pp. 259–270, Feb. 2010.
- [28] J. Xiao, G. Q. Zu, X. X. Gong, and F. X. Li, "Observation of security region boundary for smart distribution grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1731–1738, Jul. 2017.
- [29] T. Jiang, R. F. Zhang, X. Li, H. H. Chen, and G. Q. Li, "Integrated energy system security region: concepts, methods, and implementations," *Applied Energy*, vol. 283, pp. 116124, Feb. 2021.



**Guozheng Wang** received the B.S. (2015) and M.S. (2018) degrees from Central South University, Changsha, China, and Shandong University, Jinan, China, respectively. He received the Ph.D. degree from the Department of Electrical and Electronic Engineering, Tsinghua University, Beijing, China, in 2023. Now he is working for Beijing Huairou Laboratory. His research interests include power system analysis and control, and intelligent control applications.



**Jianbo Guo** received the B.S. (1982) and M.S. (1984) degrees from Huazhong University of Science and Technology, Wuhan, China, and China Electric Power Research Institute, Beijing, China, respectively. Since 2013, he has been an elected Academician of the Chinese Academy of Engineering, Beijing, China. He has long been engaged in power system analysis and control.



**Shicong Ma** received the B.S. (2002) and Ph.D. degree (2008) from Shandong University. He is currently employed by China Electric Power Research Institute. His research interests include power system stability and security, data mining for power system operation and control, and autonomous power system.



and control.

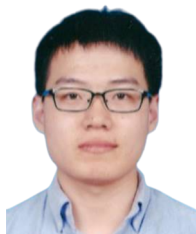
**Kui Luo** received the B.S. degree and M.S. degree in Electrical Engineering from Hefei University of Technology, Anhui, China, in 2010 and 2013, respectively. From 2014 to 2015, he was funded by the Chinese Scholarship Council (CSC) as a visiting scholar with the Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark. He is currently with the Power System Department, China Electric Power Research Institute as senior engineer. His research interests include renewable energy grid integration analysis



**Shixiong Fan** received the Ph.D. degree from the Department of Electrical and Electronic Engineering, Strathclyde University, Glasgow, U.K., in 2012. Now He is working for China Electric Power Research Institute as professorate senior engineer. His current research interests include power network dispatching automation and application of artificial intelligence and big data analysis in power grid.



**Xi Zhang** received B.Eng. degree from Beijing Jiaotong University, Beijing, China, in 2013 and Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, China, in 2017, both in Electrical Engineering. He worked as a postdoc in China Electric Power Research Institute, Beijing, China from 2018 to 2019. Currently, he is an Assistant Professor in Beijing Institute of Technology, Beijing, China since 2020. His research interests include complexity, resilience and stability analysis of the power system with high penetrations of power electronic devices.



**Tiezhu Wang** received the B.Eng. degree from Tsinghua University, Beijing, China, in 2012 and the M.Phil. Degree from China Electric Power Research Institute, Beijing, China, in 2015, both in Electrical Engineering. He is currently pursuing the Ph.D. degree with the Power System Department, China Electric Power Research Institute. His current research interests include power system simulation, and stability analysis for the power system with high penetrations of power electronics.



on “Voltage Control for Smart Grid”. His research interests include smart grids, cyber-physical systems and electrical power control center applications.

**Qinglai Guo** was graduated from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2000 with B.S degree. He received his Ph.D. degree from Tsinghua University in 2005 where he is now a Professor. He is an IET Fellow, IEEE Fellow and a CIGRE Member. Now he is TCPC of Energy Internet Coordinating Committee of IEEE PES, the co-chair of IEEE PES Work Group on “Energy Internet”, IEEE PES Task Force on “Cyber-Physical Interdependence for Power System Operation and Control”, and IEEE PES Task Force



**Weilin Hou** received her Master’s degrees from University of Wisconsin-Milwaukee in 2019 and from North China Electric Power University in 2020. She is now working for China Electric Power Research Institute. Her research areas are stability analysis of high-proportion renewable energy power system, frequency and inertia analytical calculation and wide-area control of power system stability.