Electrical Power and Energy Systems 149 (2023) 109020

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

A data-driven method for online transient stability monitoring with vision-transformer networks

Fang Jiashu, Liu Chongru *, Zheng Le, Su Chenbo

State Key Lab. for Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, 102206, China

ARTICLE INFO

Keywords:
Transient stability assessment
Unstable generators identification
Vision-transformer

ABSTRACT

Online transient stability assessment (TSA) plays an important role in power system planning and operation. The massive integration of renewable energy sources into the grids increases the risks of relay malfunction; however, the existing TSA approaches only apply to scenarios where the fault could be successfully cleared, and are not applicable under relay failure scenarios. Besides, the post-fault TSA methods normally rely on the accurate fault clearance information, which would be affected by the clock synchronization errors in relays. Thus, this paper proposes an online monitoring system that assesses transient stability based on the current operating condition, which is independent of the accurate moments of fault occurrence and clearance provided by fault indicators or relays signals; the transient stability prediction during the fault duration provides the system operators with an early warning of system instability in the event of relay failures. Moreover, an adaptively adjusted criterion for instability is proposed to balance the false alarm and misclassification rates. Furthermore, novel Vision-Transformer-based models for both TSA and unstable generators identification are built in this paper. Compared with other networks, the self-attention structure in the Transformer leads to a global receptive field and higher resolution for capturing time-series information in each sample. Our experiments on the IEEE-39 system show that the proposed method can assess transient stability at each moment in different operation states. It outperforms existing machine learning-based models in terms of accuracy and robustness in TSA and unstable generators identification tasks.

1. Introduction

Real-time transient stability assessment (TSA) has been an important research area in power system planning and operation. In recent years, the proliferation of renewables and electronic devices in power systems has led to a dramatic increase in computational complexity [1], which brings challenges to transient stability analysis. Traditional transient assessment techniques based on time-domain simulations [2] demand accurate network configurations and suffer from high computational burdens. Direct methods [3], such as the transient-energy-function and equal-area criterion methods, provide approximate results with limited modeling volumes.

The rapid development of artificial intelligence technology has yielded novel solutions for real-time and accurate transient stability monitoring [4,5]. The measurements collected from the Wide Area Measurement System (WAMS) [6] via phasor measurement units (PMUs) enable the use of data-driven methods for online transient assessment.

The stability analysis of modern power systems could be adopted in pre- or post-disturbance environments [7]. The approaches analyzing transient stability in [8–10] use steady-state characteristics to evaluate the security status concerning potentially incoming contingencies, and preventive actions should be taken to withstand the probable contingencies. [8] proposed a gated recurrent unit (GRU)-based model that predicted the transient stability of the system after faults lasting 0.25 s using the current operating point information. [9,10] realized pre-fault dynamic security assessment by an ensemble learning method, and [9] allowed assigning critical instances with larger weights. It is noted that in such works, contingencies are assumed to be successfully cleared after a fixed period; model performance would degrade with longer fault duration (if faults are removed by backup protection) or under relay failure scenarios. The support vector machine (SVM) was extended to perform transient analysis for power systems and analyze unstable generators in [11], and [12] determined unstable attributes as security indicators. However, the accuracy and robustness of TSA under shallow models have proven highly reliant on feature selection [13], and the features extraction from raw data depends on professional experience, which limits the practicality of TSA in actual power systems. With the development of data-driven techniques, deep

* Corresponding author.
E-mail address: chongru.liu@ncepu.edu.cn (Liu C.).

https://doi.org/10.1016/j.ijepes.2023.109020
Received 29 April 2022; Received in revised form 18 January 2023; Accepted 3 February 2023
Available online 21 February 2023
0142-0615/

learning methods [14] that do not rely on feature extraction have been adopted. In [15], data within 0–5 s after the fault clearing were utilized to perform TSA, and generative adversarial networks were adopted to solve the problem of training data imbalance. For fast TSA, [16] proposed a deep belief network (DBN), whose structure can automatically learn the data distribution by unsupervised feature extraction process, giving rise to high robustness. [17] emphasized the aim of fast TSA after fault clearance, and the convolutional neural network (CNN) algorithm was trained to achieve early prediction w.r.t. the onset of potential transient instability, with fault-on PMU measurements as well as two adjacent sampling points in pre-/post-fault stages being utilized as model samples. However, the approaches presented in [11–13,15–21] were normally activated by the fault clearance signal collected by relays [22,23], which would suffer from clock desynchronization between electrical devices. The method in [24] is free from the information on the accurate fault clearance moment by using sliding windows, however, it only applies to scenarios where the fault could be successfully cleared, and cannot give instability prediction in the event of relay failures. Since the risk of relay failures increases due to the expansion of the capacity of renewable energy sources integrated into the grids [25,26], the availability of this method is limited occasionally.

Focusing on the aforementioned issues, a two-stage online monitoring system is built in this paper. The first stage involves predicting the transient stability, and the second stage identifies unstable generators under unstable scenarios (as in [11,27,28]). This online monitoring system is not only independent of the fault occurrence and clearance instant knowledge but also provides transient stability prediction based on the current operation point per cycle. Furthermore, as the trajectories for cases where the faults are not correctly cleared are added to the dataset, the proposed system is also applicable to relay failure scenarios so that operators can initiate subsequent emergency control strategies to maintain synchronism.

In this paper, we propose a Transformer structure for TSA and unstable generators identification tasks. The motivation to utilize Transformer for TSA includes its global receptive field and high-resolution maintaining [29] at every stage. These Transformer characteristics make it excel at extracting high-resolution electrical trajectories collected via PMUs, thus achieving better TSA and unstable generators identification performance. In passing, it is noted that unlike CNNs that capture global dependencies with downsampling operations [24,30–33], Transformer draws global dependencies [34,35] and maintains the representation with constant dimensionality throughout the processing stages. Specifically, the Transformer model is a novel deep architecture [29] that has been successfully applied to natural language processing [29], speech recognition [36], and computer vision [37], especially to the field of machine translation, and Transformer-based models have achieved state-of-the-art results [38].

To the best of our knowledge, this is the first time an online data-driven monitoring system, which predicts the transient stability based on the current operating point, considers relay failure scenarios; besides, its performance is independent of the accurate fault moment, making the proposed method unaffected by the desynchronization between electrical devices. This is also the first time a Transformer-based network has been used in the TSA task. The main contributions of this paper can be summarized as follows.

1. Vision-Transformer (ViT) models are built to perform transient assessment and identify unstable generators. The attention computation introduced by the proposed method improves the model robustness against noise, in particular, an obvious model advance can be achieved in the multi-label unstable generators identification task compared with the existing methods.
2. A generic continuous monitoring system for transient stability prediction is proposed based on the current operating condition, which is independent of the accurate fault occurrence and clearance moments, therefore it would not be influenced by the clock desynchronization from relay signals or fault indicators. Moreover, it is applicable under relay failure scenarios, and provides the system operators with an early warning of system instability in the event of relay failures, making it practical in modern power grids where the risk of relay failures rises due to the massive renewable energy integration.
3. An adaptively adjusted instability criterion is proposed to make a proper trade-off between misclassification and false alarm samples.

The rest of this paper is organized as follows. Section 2 introduces the Transformer architecture built in this paper. Section 3 describes the details of the proposed online assessment method. A case is analyzed in Section 4 to demonstrate the performance of the constructed model. Conclusions are provided in Section 5.

2. Proposed ViT classifier for TSA and unstable generators identification

The ViT is a modified end-to-end architecture based on the standard Transformer. Unlike the standard Transformer, which receives a 1D sequence input, the ViT can reshape the spatial–temporal features into a sequence of flattened 2D patches. The use of the spatial–temporal matrix $x \in \mathbb{R}^{H\times W\times C}$ in a power system, where H and W respectively denote the numbers of generators and cycles at each sampling instant and C is the number of electrical features, motivates us to utilize the ViT, rather than the standard Transformer to perform transient stability prediction.

Fig. 1 depicts the proposed ViT classification model structure, which is composed of three main components: a linear embedding layer with positional embedding, a Transformer encoder, and a classifier layer.

The time-series information in a power system is uniformly subdivided into non-overlapping patches, and the sequence length of the patches is

$$n = HT/P^2 \tag{1}$$

where P is the resolution of each patch; generally, a smaller patch size leads to a longer sequence length, and vice versa.

Each patch is linearly projected into a token embedding layer; thus, the Transformer has a constant latent vector dimension D when using a learned embedding matrix E. The embedded patches are concatenated together with an additional learnable classification token, CLS, which is a unique design in Transformers that is considered a global feature representation summarizing all the patch tokens. Moreover, since self-attention in the Transformer encoder is permutation-invariant, an extra position embedding is added to each token to maintain the spatial arrangement in the original feature space.

All the tokens, including the CLS, are fed into the Transformer encoder, which is stacked with l identical Transformer blocks. Each Transformer block consists of a multi-head self-attention (MSA) layer and a feed-forward network (FFN) layer. The MSA and FFN layers are applied with layer normalization (LN) and residual connections to simplify the optimization process. The input of the ViT, z_0, and the processing step in the Transformer block can be expressed as:

$$\begin{aligned} z_0 &= \left[x_{\text{class}}\,; x_p^1 E; x_p^2 E; \cdots ; x_p^N E\right] + E_{\text{pos}}\,, E \in \mathbb{R}^{(P^2\cdot C)\times D} \\ z'_\ell &= MSA\left(LN\left(z_{\ell-1}\right)\right) + z_{\ell-1}, \ell = 1 \dots L \\ z_\ell &= FFN\left(\mathrm{LN}\left(z'_\ell\right)\right) + z'_\ell, \ell = 1 \dots L \end{aligned} \tag{2}$$

where z'_ℓ and z_ℓ denote the output features of the MSA module and the FFN module, respectively, x_p^i $(i = 1, \dots, N)$ denotes each patch, and E_{pos} is the positional embedding.

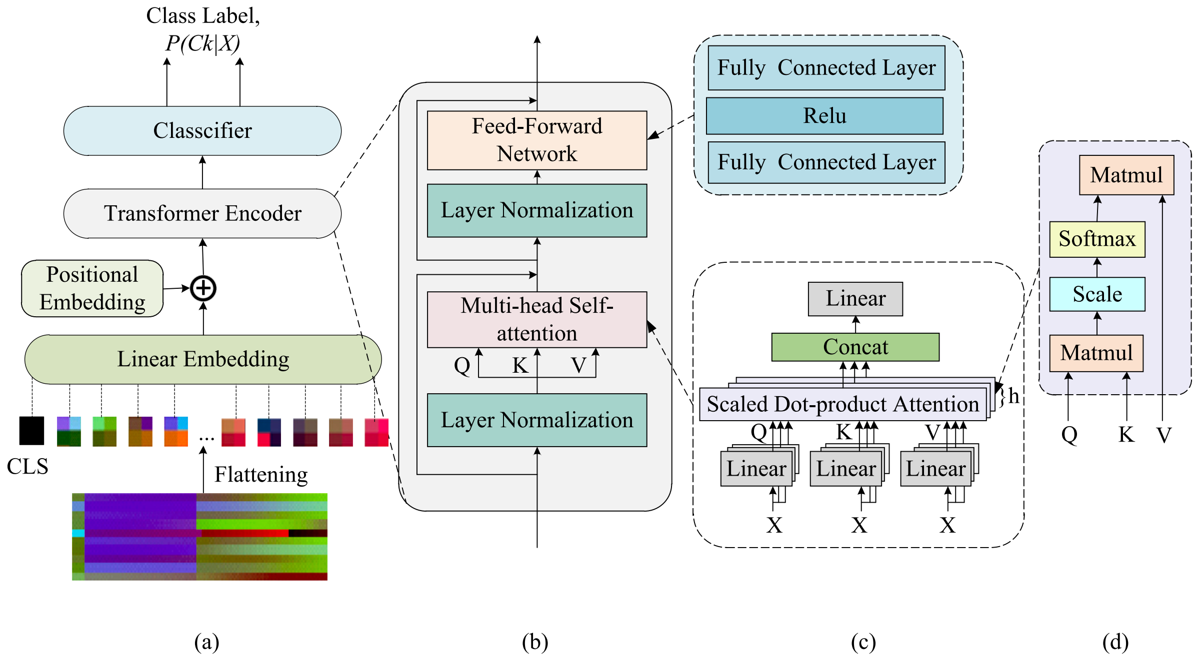

Fig. 1. The vision transformer architecture. (a) Main structure of the model, (b) the Transformer encoder architecture, (3) the multi-head self-attention layer, and (d) self-attention mechanism.

2.1. Multi-head self-attention mechanism

Scaled dot-product attention is the core component of the Transformer. The inputs of the MSA layer include a query (Q) and a key (K)-value (V) pair, which are linearly transformed from the input sequence X via three different learnable matrices as

$$\begin{aligned} Q &= W^Q X \\ K &= W^K X \\ V &= W^V X \end{aligned} \tag{3}$$

The scaled dot-product attention is calculated as

$$Attention(Q,K,V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{4}$$

where $Q,K,V \in \mathbb{R}^{\mathrm{d}}$, $\sqrt{d_k}$ is a scaling factor, and the softmax function obtains the attention weight values. The attention distribution determines the relative importance of a single patch embedding in the whole patching sequence. The dot-product attention block allows the features at each pixel position in a feature map to interact with features at all pixels rather than information in a local region [39], therefore the ViT can flexibly adjust their receptive field to maintain high robustness against noise [40].

Multi-head attention is utilized to extract diverse information from different representation subspaces. The multi-head attention mechanism contains h parallel 'heads', and each head corresponds to an independent scaled dot-product attention processing operation. The multi-head self-attention function is

$$\begin{aligned} &head_i = Attention(\mathrm{Q}_i, \mathrm{K}_i, \mathrm{V}_i) \\ &MultiHead(Q,K,V) = \\ &Concat(head_1, head_2, \dots, head_i)W^O \end{aligned} \tag{5}$$

where W^O is the matrix that aggregates the information derived from different subspaces.

With reference to Fig. 1(b) and (c), the FFN layer takes the input from the Multi-Head Self-Attention layer and further transforms it by using two linear layers (Fully Connected Layers) with ReLU activation and dropout in between, formulated as

$$FFN(x) = FC(Dropout(ReLU(FC(x)))) \tag{6}$$

2.2. Classifier layer

The classifier layer utilizes the softmax function as

$$P\left(\mathrm{y}^{(i)} = j \mid \mathbf{X}^{(i)}\right) = \frac{e^{\theta_j^T \mathbf{X}^{(i)}}}{\sum_{l=1}^{k} e^{\theta_j^T \mathbf{X}^{(i)}}} \tag{7}$$

where $\mathrm{y}^{(i)}$ is the label of input $\mathbf{X}^{(i)}$ and θ_j is the parameter vector for class j, $P\left(\mathrm{y}^{(i)} = j \mid \mathbf{X}^{(i)}\right)$ is the probability of an input $\mathbf{X}^{(i)}$ being identified as class j, and k is the number of classes. Thus, $\mathbf{X}^{(i)}$ is assigned to the class with the highest probability.

3. Proposed online assessment method

The framework in this paper is composed of two processes: offline training and online identification. As seen in Fig. 2, during the offline training process, two tasks are separately trained for TSA and unstable generators identification by using a massive electrical database with relative labels. In the online identification stage, TSA and unstable generators prediction are performed by the proposed approach.

3.1. Offline training process

3.1.1. Time-domain stage

The operation data and fault information of the power system are obtained by simulation or historical measurements, and the time-domain transient stability simulation is performed by considering different operation modes. The transient stability index (TSI) is used as the stability criterion, yielding

$$TSI = \frac{360^\circ - |\Delta\delta_{\max}|}{360^\circ + |\Delta\delta_{\max}|} \tag{8}$$

where $|\Delta\delta_{\max}|$ is the absolute value of the maximum separation angle between any two generators without any control intervention. When $TSI > 0$ the system is considered stable and vice versa.

Sliding observation windows covering T cycles are set as the model samples, which are labeled y for the TSA task, yielding

$$y = \begin{cases} 1 \text{ if } (t_E < t_F) \text{ or } (t_F \le t_E < t_C \text{ and } TSI_F \ge 0) \text{ or } (t_E \ge t_C \text{ and } TSI_C \ge 0) \\ 0 \text{ if } (t_F \le t_E < t_C \text{ and } TSI_F < 0) \text{ or } (t_E \ge t_C \text{ and } TSI_C < 0) \end{cases} \tag{9}$$

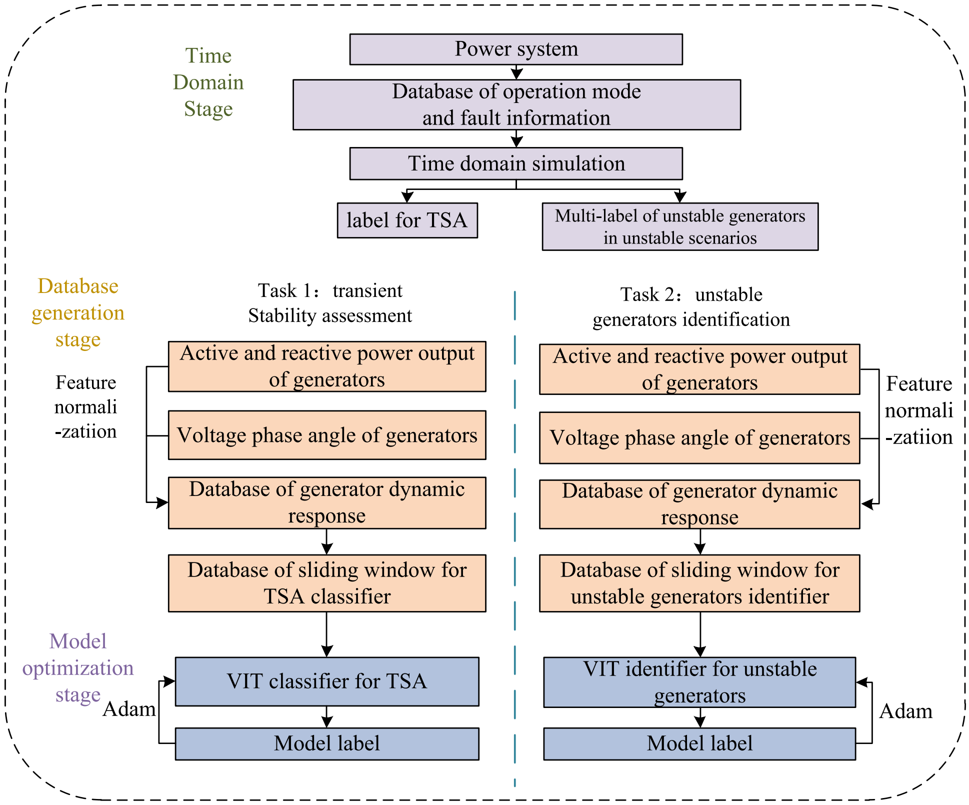

Fig. 2. The offline training process.

where t_E is the end moment of the observation window (i.e., the current monitoring moment), t_F is the fault occurrence moment, and t_C is the fault clearance moment. The label in (9) indicates the transient stability prediction under the current condition without any control intervention. TSI_F represents the transient stability prediction obtained by assuming that the fault was not cleared, and TSI_C represents the transient stability prediction obtained after the fault was cleared.

Transient stability predictions were given in [21,24] after fault clearance, in which period the data-driven methods assessed transient stability based on the value of transient instability probability $P_{t=T}$ (y = 0|X) given in (7). The data-driven TSA method in [21] needs to be activated when receiving the relay signal and is influenced by the clock synchronization errors between electrical devices. The approach in [24] is independent of the moments of fault occurrence and clearance, while the transient stability assessment was only available when the start moment of the observation window was after the fault clearance moment. In addition, relay failure scenarios were not considered in [21,24] and the samples obtained under relay failures were not added to the databases in [21,24].

However, as seen in Fig. 3(c), the online monitoring system in this paper is not only independent of the fault occurrence and clearance instant knowledge but can also provide transient stability predictions based on the current electrical measurements. In addition, this technique is applicable to relay failure scenarios as TSA is not limited to failure clearance period but covers each cycle in power system operation.

Moreover, a multi-class multi-label classifier is built to identify unstable generators under different operating patterns. An unstable generator is a generator whose rotor angle difference w.r.t. the slack generator is greater than 360° [11]. The label of an unstable generator that loses transient stability after a disturbance is set to 1, and the stable generators are labeled as 0.

3.1.2. Database generation stage

In this paper, the fault durations are set to 0.10 s, 0.15 s, and 0.20 s. Besides, the electrical trajectories in relay failure scenarios are also added to the database. Expanding the database to include relay failure cases is an essential step to make the proposed method functional under relay failure cases. This enables to construct a monitoring system applicable to a more general situation than the existing works, where only the fault clearance scenario is considered. Fig. 4 shows the comparison between the database in this paper and those in [21,24]; the database in [21] only covered post-fault data, so it can only assess transient stability after the relay operation, and the database in [24] allowed the proposed method to recognize the system operation state but is not applicable to the relay failure scenario. However, the trajectories for cases in which the faults are not correctly cleared are added to the dataset in this paper, so the proposed model extracts information covering pre- to post-disturbance period and relay failure samples. As a result, the proposed system is applicable to relay failure scenarios compared to existing methods.

The voltage phase angle and active and reactive power output in each generator are collected as time-series features. Meanwhile, z-score normalization is applied to eliminate the effect of feature scale differences, yielding

$$l' = \frac{l - \bar{L}}{\sigma_L} \tag{10}$$

where l and l' are the original and normalized value, respectively; $\bar{L}$ and σ_L are the mean value and standard deviation, respectively.

3.1.3. Model optimization stage

Cross-entropy loss is used as the loss function in TSA task with L2 regularization [41], and these operations can effectively prevent the model from overfitting high-dimensional electrical data. In addition,

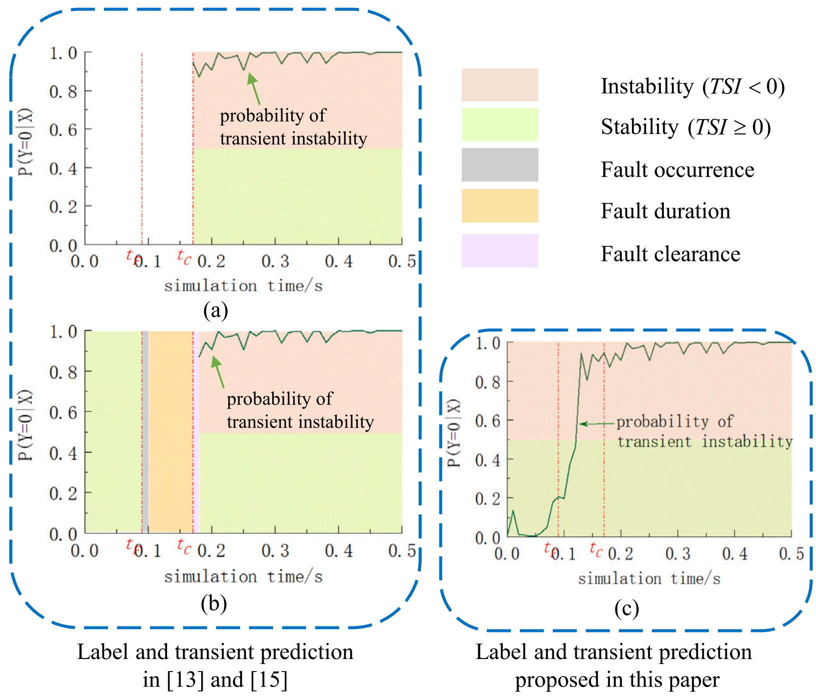

Fig. 3. Comparison between the sample labels and the periods indicating transient predictions.

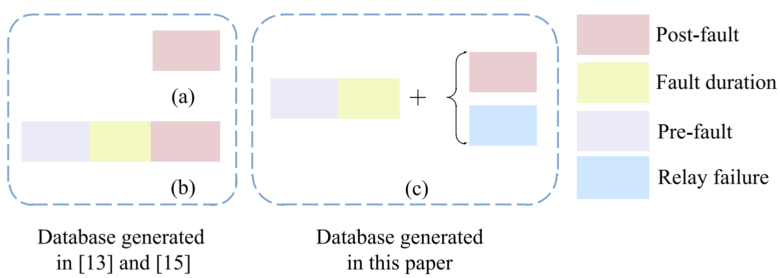

Fig. 4. Database comparison.

the Adam optimizer is applied to speed up the model training process, as this approach adaptively determines the step size during gradient descent [42].

Two ViT classifiers are separately trained for TSA and unstable generators identification, where the ViT unstable generator identifier is a multi-class classifier since more than one generator may be recognized as unstable generators.

3.2. Online identification stage

In this paper, the TSA task is performed continuously with sliding windows. The sliding observation window is composed of the current and previous (T-1) measurements, and the real-time measurements are collected via PMUs once per cycle.

The measurements from PMUs are stored as multi-dimensional arrays in the proposed neural network. Fig. 5 visualizes a classic unstable scenario by rearranging the multi-dimensional electrical features as R-G-B channels. This visualization shows the unstable scenario covering 0.5 s from pre-fault to post-fault period, the instability moment when the maximum separation angle in generators exceeds 360° is 3.96 s. As can be seen in Fig. 5, obvious changes can be noticed at t_F and t_C; the unstable generator presents different characteristics from the other generators, and its unstable characteristics become more noticeable over time.

As shown in Fig. 6, the online identification stage continuously assesses transient stability with real-time electrical measurements. And an additional criterion should be satisfied before the system is formally recognized as unstable. The moment when the model first formally recognizes transient instability is set as t_{alarm}; the measurements observed at t_{alarm} are sent to the second classifier to identify unstable generators.

The proposed criterion for instability recognition is

$$P_{t=T}(y = 0|X) > \tau \tag{11}$$

Clearly, the prediction probability of each classifier in (7) shows how confident the classifier is in supporting the prediction [43]. The prediction is more convincing if the output for one class is much higher than those of the other classes; for instance, $P_{t=T}(y = 1|X) = 0.99$ and $P_{t=T}(y = 0|X) = 0.01$ in the binary classifier represent a high probability for label 1.

Two factors need to be considered when designing the threshold τ in (11). The first concerns the overall performance of the classifier for imbalanced data. As a matter of fact, as the number of stable samples is prevailing over that of unstable samples, the threshold of classification can be changed to bias the predictive behavior of a classification model. Second, the false positive (FP) samples (i.e., unstable samples predicted as stable) normally lead to more severe detrimental effects than the false negative (FN) samples (i.e., stable samples predicted as unstable); however, it is not economical to only concentrate on

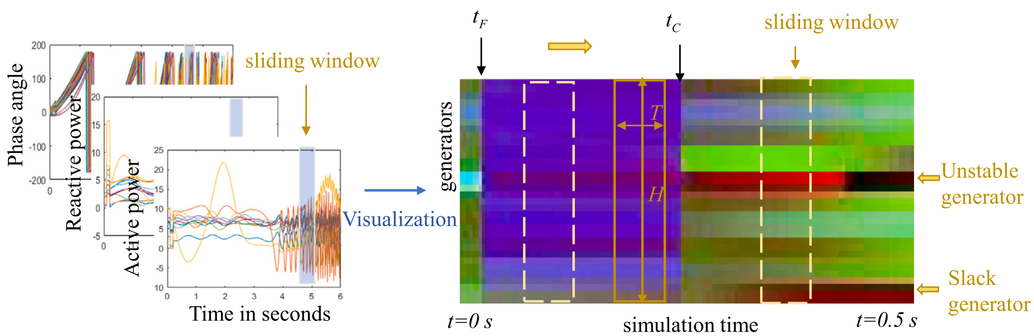

Fig. 5. Visualization of the dynamic information in an unstable scenario.

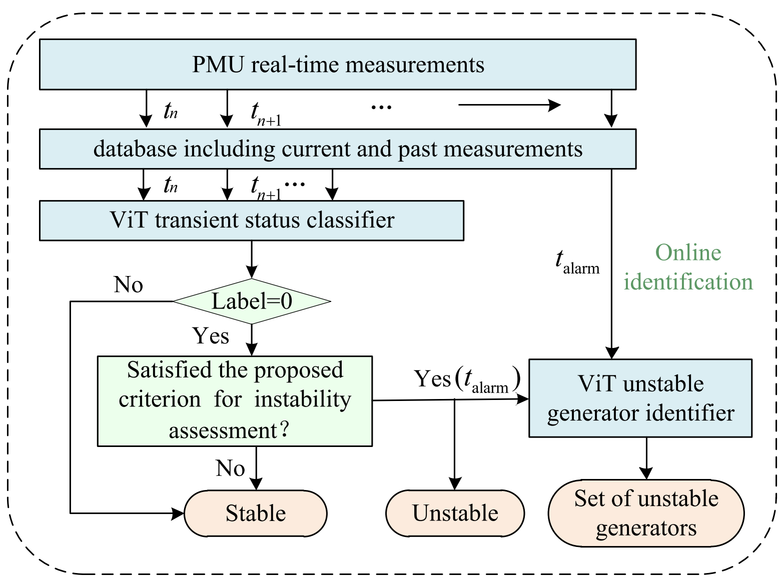

Fig. 6. The online identification process.

reducing the number of FP samples. Therefore, we built a threshold algorithm to determine the best value of τ considering the trade-off between precision and recall.

Specifically, precision evaluates the ratio of correctly predicted stable (TP) samples to the total predicted stable samples, conversely recall evaluates the ratio of TP samples to actual stable samples, yielding

$$Precision(\tau) = \frac{TP(\tau)}{TP(\tau) + FP(\tau)} \tag{12}$$

$$Recall(\tau) = \frac{TP(\tau)}{TP(\tau) + FN(\tau)} \tag{13}$$

In order to optimize the economical effect considering FP and FN samples, we built an algorithm to determine the proper threshold τ in (11) by pursuing the highest weighted sum of precision and recall. To this end, the objective of the algorithm writes

$$\max_{\tau} score(\tau) = \alpha \times Precision(\tau) + Recall(\tau) \tag{14}$$

where $\alpha(>0)$ is a weight set by power system operators considering security and economical effects. The threshold based on this approach can be adaptively adjusted with different databases of power systems. An illustrative example of practical design will be provided in the following.

4. Case studies

4.1. Database generation

The case study is based on the 50 Hz IEEE-39 power system. Massive databases are generated by considering different operation conditions and contingencies. The generations and loads are randomly changed around 80%–120% of the basic operating level, and the bus voltage is maintained within 0.95–1.05 p.u. Three phase-to-ground short-circuit faults are applied to three locations (25%, 50%, and 70% of the length) [44] on the transmission line. The fault durations are set to 0.10 s, 0.15 s, and 0.20 s, and simulations of scenarios in which the faults are not successfully cleared are also conducted. The database covers the simulation of 10 s. Transient stability simulations are conducted with Power System Toolbox 3.0 [45]. PMUs are employed to measure the electrical information of generators once per cycle covering pre-fault, fault duration, and fault clearance periods. A total of 9200 scenarios are utilized in the TSA task, and the databases for the unstable generators identification task are generated in 3892 unstable scenarios. The training, validation, and testing sets are randomly divided at an 8:1:1 ratio. PyTorch, which is a Python-based scientific computing package, is used as the machine learning framework.

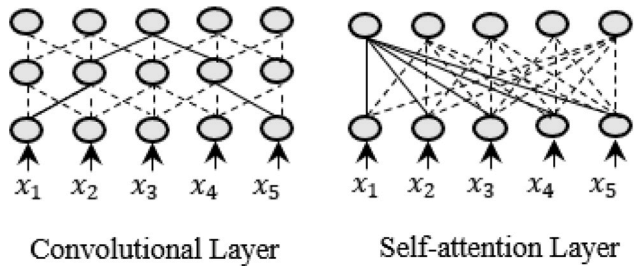

Fig. 7. Architectures of different sequential extraction neural networks.

Table 1
Performance comparison.

Model	Acc. of TSA (%)	Acc. of unstable generators (%)
MLP [20]	94.29	83.13
SVM [11]	95.51	84.50
DBN [16]	97.23	89.21
GRU [8]	97.12	91.29
CNN [24]	98.63	91.01
Transformer	98.92	94.21

4.2. Performance of classifier

Two separate classifiers are trained in different structures. The first classifier is built for the TSA task, and the second classifier is built for unstable generators identification. The classifiers are trained for up to 300 epochs. The observation length T is set to 5 [24]. The ViT classifier for TSA has 2 Transformer blocks stacked with 3 attention heads, and the ViT classifier for unstable generators has 4 Transformer blocks with 3 attention heads. Different comparison models, including the deep structure models (CNN, GRU, and DBN) and shallow models (SVM and MLP) are trained separately for the two tasks. Specifically, two CNN classifiers are built based on [24], with kernel size being set to 10×3 to capture global features. Four layers MLPs are built based on [20], with the kernel sizes being set to 100, 50, 30, and 5 for TSA model, and 100, 50, 30, and 10 for generators identification. Note the transient information is reshaped as 1-D vectors in SVM and MLP, and principal component analysis (PCA) is utilized to enhance the performance. The hyper-parameters of models utilized in this paper are tuned using the grid-search process [46] in order to achieve the best performance.

Fig. 7 shows the architectures of different deep neural networks. The CNN extracts local features through the convolutional kernel to ultimately obtain global information, while the local features in the Transformer directly interact with each other to draw global dependencies. The maximum path length between any two input patches is 1 in the Transformer and logk(n) in the CNN, where n is the sequence length and k is the convolution kernel size [29].

As seen in Table 1, the Transformer and CNN methods achieve high accuracy (Acc) in the TSA task, which suggests that both the CNN and Transformer have superior performance in extracting the features that indicate the transient stability. However, the ViT structure achieves significant improvements in the unstable generators task over the other methods. In addition, the performances of the SVM and MLP degrade significantly in multi-label unstable generators identification task relative to the TSA task, indicating the limited ability of shallow models to realize complicated multi-label unstable generators identification tasks. It should be noted that these performance results indicate the accuracy rates obtained with different samples but not in different scenarios, with 490 samples generated in each scenario as sliding observation windows.

Unlike the object detection task, which involves a sparse input feature map, each pixel in the TSA feature map indicates the actual temporal electrical measurements. As an example, Fig. 5 visualizes the feature maps constructed with different electrical features. Besides, the unstable generators are varied under different fault scenarios, as shown in Fig. 8, where rotor angle instability occurs in single or multiple generators. The unstable characteristics can also be observed in Fig. 5. This motivates the use of the algorithm with a higher resolution (1) and direct interaction (4) when extracting the feature map in the unstable generators identification task, as the unstable trajectories are easier to be distinguished from those of stable generators when using a higher resolution. The previous work that solved the TSA task with CNN backbone captures global information [24] using a larger kernel size, which led to an efficiency reduction due to the significantly larger number of parameters and computations [30]. Conversely, the Transformer constantly retains the feature dimension with high resolution throughout all processing stages, foregoing the explicit downsampling operation in CNN. Moreover, the attention mechanism in the Transformer allows the measurements at each generator and each time position in a feature map to refer to global information rather than local information [29]. Since the higher resolution and the direct interactions among global features in ViT structure naturally lead to fine-grained predictions [47], the proposed model achieves better performance in TSA and unstable generators identification tasks, as has been proven by the results.

Table 2
Computational efficiency of different methods.

Model	Offline processing time/s		Online test time per sample/s	
	TSA	Generators	TSA	Generators
MLP [20]	62.930	133.094	0.012	0.030
SVM [11]	42.069	101.039	0.012	0.032
DBN [16]	564.102	2600.425	0.129	0.172
GRU [8]	583.294	2784.065	0.137	0.167
CNN [24]	491.996	2419.827	0.101	0.124
Transformer	610.183	3091.249	0.121	0.142

The offline training and online testing times of different methods are summarized in Table 2. All computations were executed on a server with a CPU 6240R 2.4G, 24C/48T. In general, the SVM and MLP models consume less time with their shallow structures, at the expense of their achieved accuracy (as discussed above). In contrast to the progressive behavior of recurrent and convolutional operations, (4) indicates that the attention computation captures long-range dependencies directly by computing interactions between any two position regardless of their positional distance [39], leading to an increased time consumption during the training process compared with CNN. The training time of the proposed model is highly influenced by the pixel size as shown in (1), so the pixel size is set to 2 to achieve a compromise between the information precision in feature maps and computational efficiency.

4.3. Performance of classifiers under noise pollution

The model performance in noise free environment is shown in Table 1. As the electrical data collected from PMUs suffer from wide-area noise, it is of great significance to ensure that the proposed method can withstand noise. According to [48], the total PMU vector error should be less than 1%, which corresponds to the signal-to-noise ratio (SNR) of 40 dB; besides that case, the SNR values of 10, 20 are also tested as in [49]. The model is trained by the original simulation data, while different levels of Gaussian noise are added to the test set to evaluate the model's robustness.

As shown in Tables 3 and 4, the accuracy of each method degrades to a certain degree. Larger accuracy decreases are observed in the unstable generators identification task than in the TSA task, since the multi-label prediction of dynamic responses in generators is more complicated. While the models with deep layers are quite robust against noise compared with shallow models, the proposed model has the best performance in both tasks, especially in the unstable generator identification task, which can be ascribed to the introduction of the

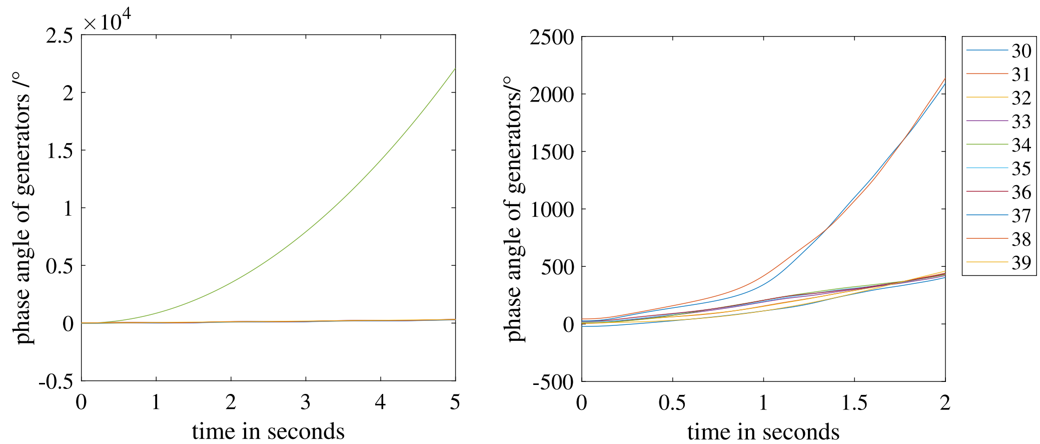

Fig. 8. Phase angle curves showing the generator stability in scenarios with (a) one unstable generator, and (b) two unstable generators. Both scenarios are effectively addressed by the proposed method.

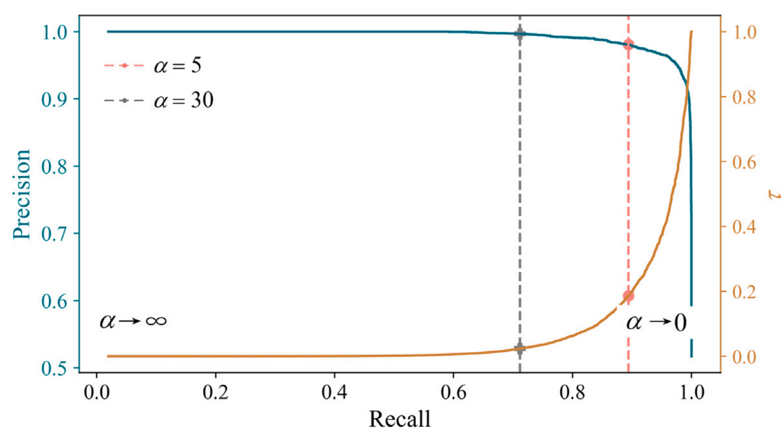

Fig. 9. Recall–precision curve and threshold-precision curve yielding maximum score in (14) with different weights.

Table 3
Accuracy (in %) of different models under various SNR levels for TSA.

Model	SNR (dB)		
	40	20	10
MLP [20]	93.01	92.21	88.21
SVM [11]	94.19	93.08	90.69
DBN [16]	96.18	95.58	93.33
GRU [8]	95.62	94.44	92.98
CNN [24]	97.43	97.19	95.47
Transformer	97.89	97.42	96.78

Table 4
Accuracy (in %) of different models under various SNR levels for unstable generators identification.

Model	SNR (dB)		
	40	20	10
MLP [20]	81.21	80.51	79.33
SVM [11]	82.19	81.30	80.99
DBN [16]	87.80	86.80	85.03
GRU [8]	90.24	88.74	85.63
CNN [24]	89.92	88.81	87.25
Transformer	93.12	92.49	90.88

attention mechanism. Indeed, the Gaussian noise in the power system is similar to the texture noise in image classification tasks, and the self-attention structure is proven more robust against such noise than the Convs in CNN as discussed in [40], the ViT can maintain the representation of dynamic responses in power system with long-range interaction modeling capabilities owing to the attention computation using (4), whereas convolutions are content-independent as the same filter weights are applied to all inputs [40]. Besides, the attention operations can be addressed as trainable spatial smoothing of feature maps [50], leading to high robustness of the proposed model against Gaussian noise.

4.4. Design of optimal threshold

This paper proposes a method [see (14)] to adaptively optimize the threshold value. Such a process is demonstrated in this section, by utilizing data from 10 different line faults. During this process, the model requires no additional retraining or modification.

The recall–precision curve is illustrated in Fig. 9, showing the classifier's performance across different thresholds and thus determining the best threshold for a specific task. The threshold value that optimizes (14) changes by varying the weight α, and the pertinent recall and precision values change according to (12) and (13). Specifically, two weights, i.e., 5 and 30, are highlighted in Fig. 9, and their corresponding thresholds are 0.186 and 0.024, respectively.

Two different scenarios are considered in the following. When the power system operator adopts a more conservative control strategy to limit FP samples, the weighted ratio in (14) is set to 30 with the corresponding threshold adjusted to 0.024, therefore the misclassified unstable samples are extremely rare in this case with precision reaching 0.987. This allows the stable probability $P_{t=T}(y = 1|X)$ in samples to reach an incredibly high value when they are determined as stable, presenting a very high confidence level of the stable prediction that minimizes the FP samples. Conversely, as long as reducing economic losses due to false alarms is considered, a certain level of FP samples can be tolerated with the threshold being set to a higher value. In the latter scenario, the weight α is set to 5, recall increases accordingly with the reduction of FN samples, and the proposed method foresees power system instability when $P_{t=T}(y=0|X)$ is higher than 0.186.

5. Conclusion

This paper presents a two-stage online monitoring system for transient stability assessment and unstable generators identification based on the ViT network. The proposed TSA scheme provides the transient status based on the current operation point without the knowledge of fault occurrence or clearance signals, and is applicable to relay failure scenarios. Moreover, the proposed method outperforms the existing approaches in terms of accuracy and robustness, and the improvement of model performance is particularly noticeable for the unstable generators identification task owing to the introduction of the attention mechanism. Finally, an adaptively adjusted criterion for instability assessment is proposed considering the imbalance of database and economical effects.

CRediT authorship contribution statement

Fang Jiashu: Conceptualization, Methodology, Writing – original draft. **Liu Chongru:** Supervision. **Zheng Le:** Co-supervision, Writing – review & editing. **Su Chenbo:** Investigation, Validation.

Declaration of competing interest

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

Data availability

The data that has been used is confidential.

References

[1] Liao X, Liu K, Niu H, Luo J, Li Y, Qin L. An interval Taylor-based method for transient stability assessment of power systems with uncertainties. Int J Electr Power Energy Syst 2018;98:108–17. http://dx.doi.org/10.1016/j.ijepes.2017.11.043.
[2] Stott B. Power system dynamic response calculations. Proc IEEE 1979;67(2):219–41. http://dx.doi.org/10.1109/PROC.1979.11233.
[3] Pai M. Energy function analysis for power system stability. Power electronics and power systems, Springer US; 2012.
[4] Lashgari M, Shahrtash SM. Fast online decision tree-based scheme for predicting transient and short-term voltage stability status and determining driving force of instability. Int J Electr Power Energy Syst 2022;137:107738. http://dx.doi.org/10.1016/j.ijepes.2021.107738.
[5] Li X, Liu C, Guo P, Liu S, Ning J. Deep learning-based transient stability assessment framework for large-scale modern power system. Int J Electr Power Energy Syst 2022;139:108010. http://dx.doi.org/10.1016/j.ijepes.2022.108010.
[6] Samantaray SR, Kamwa I, Joos G. Phasor measurement unit based wide-area monitoring and information sharing between micro-grids. IET Gener, Transm Distrib 2017;11(5):1293–302. http://dx.doi.org/10.1049/iet-gtd.2016.1419, arXiv:https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-gtd.2016.1419.
[7] Xu Y, Dong ZY, Zhang R, Po Wong K. A decision tree-based on-line preventive control strategy for power system transient instability prevention. Internat J Systems Sci 2014;45(2):176–86. http://dx.doi.org/10.1080/00207721.2011.626906.
[8] Ren C, Xu Y, Zhang R. An interpretable deep learning method for power system transient stability assessment via tree regularization. IEEE Trans Power Syst 2022;37(5):3359–69. http://dx.doi.org/10.1109/TPWRS.2021.3133611.
[9] Chen Z, Ren C, Xu Y. An improved AdaBoost-based ensemble learning method for data-driven dynamic security assessment of power systems. In: 2022 IEEE power & energy society general meeting (PESGM). 2022, p. 1–5. http://dx.doi.org/10.1109/PESGM48719.2022.9917151.
[10] Li Q, Xu Y, Ren C, Zhao J. A hybrid data-driven method for online power system dynamic security assessment with incomplete PMU measurements. In: 2019 IEEE power & energy society general meeting (PESGM). 2019, p. 1–5. http://dx.doi.org/10.1109/PESGM40551.2019.8973710.
[11] Guo T, Milanović JV. Online identification of power system dynamic signature using PMU measurements and data mining. IEEE Trans Power Syst 2016;31(3):1760–8. http://dx.doi.org/10.1109/TPWRS.2015.2453424.
[12] Sun K, Likhate S, Vittal V, Kolluri VS, Mandal S. An online dynamic security assessment scheme using phasor measurements and decision trees. IEEE Trans Power Syst 2007;22(4):1935–43. http://dx.doi.org/10.1109/TPWRS.2007.908476.
[13] Baltas NG, Mazidi P, Ma J, de Asis Fernandez F, Rodriguez P. A comparative analysis of decision trees, support vector machines and artificial neural networks for on-line transient stability assessment. In: 2018 International conference on smart energy systems and technologies (SEST). 2018, p. 1–6. http://dx.doi.org/10.1109/SEST.2018.8495872.
[14] Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw 2015;61:85–117. http://dx.doi.org/10.1016/j.neunet.2014.09.003.
[15] Zhou X, Guan X, Sun D, Jiang H, Peng J, Jin Y, Zhang Y. Transient stability assessment based on gated graph neural network with imbalanced data in internet of energy. IEEE Internet Things J 2022;9(12):9320–31. http://dx.doi.org/10.1109/JIOT.2021.3127895.
[16] Wu S, Zheng L, Hu W, Yu R, Liu B. Improved deep belief network and model interpretation method for power system transient stability assessment. J Mod Power Syst Clean Energy 2020;8(1):27–37. http://dx.doi.org/10.35833/MPCE.2019.000058.
[17] Zhu L, Hill DJ, Lu C. Hierarchical deep learning machine for power system online transient stability prediction. IEEE Trans Power Syst 2020;35(3):2399–411. http://dx.doi.org/10.1109/TPWRS.2019.2957377.
[18] Kamwa I, Samantaray SR, Joos G. Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices. IEEE Trans Smart Grid 2010;1(2):144–58. http://dx.doi.org/10.1109/TSG.2010.2052935.
[19] Balasubramaniam K, Venayagamoorthy GK, Watson N. CNN based power system transient stability margin and voltage stability index prediction. In: 2013 IEEE computational intelligence applications in smart grid (CIASG). 2013, p. 13–20. http://dx.doi.org/10.1109/CIASG.2013.6611493.
[20] Zhou Y, Guo Q, Sun H, Yu Z, Wu J, Hao L. A novel data-driven approach for transient stability prediction of power systems considering the operational variability. Int J Electr Power Energy Syst 2019;107:379–94. http://dx.doi.org/10.1016/j.ijepes.2018.11.031.
[21] Shi Z, Yao W, Zeng L, Wen J, Fang J, Ai X, Wen J. Convolutional neural network-based power system transient stability assessment and instability mode prediction. Appl Energy 2020;263:114586. http://dx.doi.org/10.1016/j.apenergy.2020.114586.
[22] Mazhari SM, Khorramdel B, Chung CY, Kamwa I, Novosel D. A simulation-based classification approach for online prediction of generator dynamic behavior under multiple large disturbances. IEEE Trans Power Syst 2021;36(2):1217–28. http://dx.doi.org/10.1109/TPWRS.2020.3021137.
[23] Gomez FR, Rajapakse AD, Annakkage UD, Fernando IT. Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. IEEE Trans Power Syst 2011;26(3):1474–83. http://dx.doi.org/10.1109/TPWRS.2010.2082575.
[24] Gupta A, Gurrala G, Sastry PS. An online power system stability monitoring system using convolutional neural networks. IEEE Trans Power Syst 2019;34(2):864–72. http://dx.doi.org/10.1109/TPWRS.2018.2872505.
[25] Fang Y, Jia K, Yang Z, Li Y, Bi T. Impact of inverter-interfaced renewable energy generators on distance protection and an improved scheme. IEEE Trans Ind Electron 2019;66(9):7078–88. http://dx.doi.org/10.1109/TIE.2018.2873521.
[26] Fang Z. Adaptability analysis of directional element in photovoltaic power station. In: 2021 International conference on networking, communications and information technology (NetCIT). 2021, p. 38–45. http://dx.doi.org/10.1109/NetCIT54147.2021.00015.
[27] Soni BP, Saxena A, Gupta V, Surana S. Identification of generator criticality and transient instability by supervising real-time rotor angle trajectories employing RBFNN. ISA Trans 2018;83:66–88. http://dx.doi.org/10.1016/j.isatra.2018.08.008.
[28] Siddiqui SA, Verma K, Niazi KR, Fozdar M. Real-time monitoring of post-fault scenario for determining generator coherency and transient stability through ANN. IEEE Trans Ind Appl 2018;54(1):685–92. http://dx.doi.org/10.1109/TIA.2017.2753176.
[29] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in neural information processing systems, vol. 30. Curran Associates, Inc.; 2017.
[30] Tan M, Le QV. MixConv: Mixed depthwise convolutional kernels. 2019, http://dx.doi.org/10.48550/ARXIV.1907.09595.

[31] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. 2015, http://dx.doi.org/10.48550/ARXIV.1511.07122.

[32] Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, Liu W, Xiao B. Deep high-resolution representation learning for visual recognition. IEEE Trans Pattern Anal Mach Intell 2021;43(10):3349–64. http://dx.doi.org/10.1109/TPAMI.2020.2983686.

[33] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical image computing and computer-assisted intervention –MICCAI 2015. Cham: Springer International Publishing; 2015, p. 234–41.

[34] Ramachandran P, Parmar N, Vaswani A, Bello I, Levskaya A, Shlens J. Stand-alone self-attention in vision models. 2019, http://dx.doi.org/10.48550/ARXIV.1906.05909.

[35] Hu H, Zhang Z, Xie Z, Lin S. Local relation networks for image recognition. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV). 2019, p. 3464–73.

[36] Zhang Q, Lu H, Sak H, Tripathi A, McDermott E, Koo S, Kumar S. Transformer transducer: A streamable speech recognition model with transformer encoders and RNN-T loss. In: ICASSP 2020 - 2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). 2020, p. 7829–33. http://dx.doi.org/10.1109/ICASSP40776.2020.9053896.

[37] Ranftl R, Bochkovskiy A, Koltun V. Vision transformers for dense prediction. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV). 2021, p. 12179–88.

[38] Liu X, Duh K, Liu L, Gao J. Very deep transformers for neural machine translation. 2020, http://dx.doi.org/10.48550/ARXIV.2008.07772.

[39] Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. 2018, p. 7794–803. http://dx.doi.org/10.1109/CVPR.2018.00813.

[40] Naseer MM, Ranasinghe K, Khan SH, Hayat M, Shahbaz Khan F, Yang M-H. Intriguing properties of vision transformers. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, editors. Advances in neural information processing systems, vol. 34. Curran Associates, Inc.; 2021, p. 23296–308.

[41] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324. http://dx.doi.org/10.1109/5.726791.

[42] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014.

[43] Nguyen TT, Luong AV, Dang MT, Liew AW-C, McCall J. Ensemble selection based on classifier prediction confidence. Pattern Recognit 2020;100:107104. http://dx.doi.org/10.1016/j.patcog.2019.107104.

[44] Gupta A, Gurrala G, Sastry PS. Instability prediction in power systems using recurrent neural networks. In: Proceedings of the 26th international joint conference on artificial intelligence. 2017, p. 1795–801.

[45] Chow J, Cheung K. A toolbox for power system dynamics and control engineering education and research. IEEE Trans Power Syst 1992;7(4):1559–64. http://dx.doi.org/10.1109/59.207380.

[46] Florea A-C, Andonie R. Weighted random search for hyperparameter optimization. 2020, CoRR abs/2004.01628, arXiv:2004.01628.

[47] Mao Y, Zhang J, Wan Z, Dai Y, Li A, Lv Y, Tian X, Fan D-P, Barnes N. Generative transformer for accurate and reliable salient object detection. 2021, http://dx.doi.org/10.48550/ARXIV.2104.10127.

[48] IEEE standard for synchrophasor data transfer for power systems, IEEE std c37.118.2-2011 (revision of IEEE std c37.118-2005). 2011, p. 1–53. http://dx.doi.org/10.1109/IEEESTD.2011.6111222.

[49] Chen K, Hu J, He J. Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. IEEE Trans Smart Grid 2018;9(3):1748–58. http://dx.doi.org/10.1109/TSG.2016.2598881.

[50] Park N, Kim S. How do vision transformers work?. 2022, http://dx.doi.org/10.48550/ARXIV.2202.06709.