

ORIGINAL RESEARCH

Open Access



Power system transient stability assessment based on the multiple paralleled convolutional neural network and gated recurrent unit

Shan Cheng^{*} , Zihao Yu, Ye Liu and Xianwang Zuo

Abstract

In order to accurately evaluate power system stability in a timely manner after faults, and further improve the feature extraction ability of the model, this paper presents an improved transient stability assessment (TSA) method of CNN + GRU. This comprises a convolutional neural network (CNN) and gated recurrent unit (GRU). CNN has the feature extraction capability for a micro short-term time sequence, while GRU can extract characteristics contained in a macro long-term time sequence. The two are integrated to comprehensively extract the high-order features that are contained in a transient process. To overcome the difficulty of sample misclassification, a multiple parallel (MP) CNN + GRU, with multiple CNN + GRU connected in parallel, is created. Additionally, an improved focal loss (FL) function which can implement self-adaptive adjustment according to the neural network training is introduced to guide model training. Finally, the proposed methods are verified on the IEEE 39 and 145-bus systems. The simulation results indicate that the proposed methods have better TSA performance than other existing methods.

Keywords: Transient stability assessment, MP CNN + GRU, Sample misclassification, Improved focal loss function

1 Introduction

Given the development of power systems and the integrated access of large intermittent renewable energy resources [1, 2], the existing power system is faced with various challenges and is more prone to various faults [3]. The three-phase short circuit fault, as the strongest destructive fault, may lead to power system transient instability. Thus, it is important to have an applicable transient stability assessment (TSA) method.

General TSA methods the time domain simulation (TDS) [4], the transient energy function (TEF) [5], and the extended equal area criterion (EEAC) methods [6]. The TDS method is computationally intensive and time-consuming, the TEF method has some state variables unavailable, while the EEAC method merely works on a modern power system with limited range of analysis.

Recently the progress of machine learning (ML) theory, such as with artificial neural networks (ANNs) [7, 8], support vector machines (SVMs) [9], decision trees (DT) [10], GRU [11] and their application in power system have made TSA methods more diverse. Although TSA law can be acquired from the data handled by ML, ML has insufficient ability to extract features from multi-dimensional data, and is prone to under-fitting.

In recent years, deep learning (DL) has contributed to TSA. Starting from the data itself, DL is expert in capturing the internal laws of large amounts of data, and has robust generalization performance. This overcomes the shortcoming of ML methods. In addition, TSA using a DL algorithm can effectively bypass the procedure of modeling and solving high-order nonlinear equations, and directly obtains the mapping relationship between input features and stability labels. Some have applied DL algorithms, such as long short-term memory (LSTM) [12, 13], stacked autoencoder (SAE) [14, 15], deep belief network (DBN) [16–18] and CNN

*Correspondence: hpucquyzu@ctgu.edu.cn

College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, Hubei, China

[19–22] to TSA. Reference [12] uses LSTM to obtain a temporal self-adaptive TSA scheme, aiming to balance the trade-off between TSA accuracy and rapidity, and mine the temporal data dependencies. An LSTM-based model which can decrease the predictive value with an invariant time step is proposed in [13], while [14] proposes an innovative algorithm by clustering a multi-branch stacked denoising autoencoder (MSDAE), combined with one-fusion layer and one logistic regression (LR), which together contribute to the distinctive capability of the mining feature. Reference [15] is based on SAE and proposes an ensemble classifier in which SAE is combined with a fusion layer to classify the state of the power system. Su et al. [16] attempts to integrate DBN with the reference-point-based non-dominated sorting genetic algorithm to develop a novel preventive control scheme, whereas [17] presents an advanced DBN which takes the structural features of power system during loss function construction into consideration to better perform TSA. In Liu et al. [18], the number of nodes in each layer of DBN is decided by a particle swarm optimization algorithm and the integrated algorithm has higher TSA accuracy. However, the input of DBN is limited by one-dimensional data, the training process is very slow, and the parameter selection is very difficult. Therefore, it is easy for DBN to fall into a local optimal solution and TSA application of DBN has significant limitations.

In contrast, CNN can adapt to the input data of various dimensions and improve the data-fitting degree by parameter sharing and weight reduction. In Gao et al. [19], a one dimensional-convolutional neural network (1D-CNN) with four convolutional-pooling layers is applied to TSA, and the demand of end-to-end time sequence extraction and TSA classification are fulfilled. However, [19] merely operates an individual CNN to extract the features and there is no in-depth study of the misclassification problem. To effectively address the misclassification problem, MP CNN is proposed in [20], and the classification results provided by several CNNs are synthesized according to the synthesis principle. In Shi et al. [21], a classification model with respect to the difference between two types of instability modes (aperiodic or oscillatory instability) based on CNN is created. This selects the bus voltage phasor provided by phasor measurement units (PMUs) as the original input, and outputs three types of classification results: stable, aperiodic unstable or oscillatory unstable. However, the parameter adjustment of the FL function needs repeated experiments. In Zhao and Shi [22] a creative CNN is designed which applies multi-size convolutional kernels instead of a single size convolutional kernel in order to extract the abstract features from multi-size time scales.

Although the above studies using CNN achieve good results, there exist the following problems:

- (1) A single CNN cannot effectively extract the high-order features contained in a macro long-term time sequence and does not, considered from the aspect of algorithm fusion, is a deficiency of CNN, one which severely restrains its TSA capability.
- (2) The existing studies have not deeply probed the misclassification problem of difficulty in a classified sample because when a single neural network is applied to TSA, the difficult to classify samples around the classification threshold exist as an intrinsic misclassification problem, resulting in restricted classification accuracy.
- (3) The previous FL function needs a large number of time-consuming experiments and a parameter adjustment procedure, which together result in low operational efficiency and will not be applicable in practical engineering.

Given the above problems, three solutions are proposed. The main contributions of the paper are:

- (1) To overcome the shortcomings of a single CNN network in macro feature extraction, an integrated network, called CNN+GRU, composed of CNN and GRU is proposed. CNN is to extract the high-order features contained in a local short-term time sequence while, most importantly, GRU can fully mine the abstract characteristics hidden in a macro long-term time sequence. They complement each other to extract features more comprehensively.
- (2) In order to solve the misclassification problem of difficult to classify samples and improve classification accuracy, multiple CNN+GRU are connected in parallel to form the MP CNN+GRU. This can synchronously output multiple TSA results and improve the classification ability of samples around the classification threshold.
- (3) To effectively avoid unnecessary parameter adjustment, an improved FL function is proposed. This can implement self-adaptive adjustment according to the neural network training, and has stronger engineering applicability.

The rest of the paper is organized as follows: Sect. 2 offers the introduction and basic structure of MP CNN+GRU, and CNN, GRU and CNN+GRU are introduced. Section 3 introduces the MP CNN+GRU-based TSA where feature selection, normalization, stability criterion etc. are involved. Simulation verification is discussed in Sect. 4, and conclusions are drawn in Sect. 5.

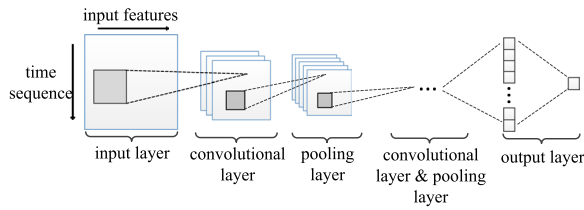


Fig. 1 The structure of CNN

2 The structure of the multiple paralleled CNN + GRU

MP CNN + GRU, as an innovative algorithm proposed in this paper, can not only effectively extract the macro long-term and local short-term features of input information, but also address the misclassification problem regarding difficult to classify samples to some extent. The network is illustrated in the following sub-sections.

2.1 The structure of CNN

As illustrated in Fig. 1, CNN [19, 23] is composed of input, hidden and output layers, while the hidden layer consists of convolutional layers and pooling layers. The convolutional kernel in the convolutional layer performs a convolutional operation to complete feature extraction regarding local information, whereas the pooling layers refine the most representative features from the convolutional layer and implement redundant information elimination. The convolutional and the pooling layers are stacked in turn to extract the high-order features.

Defining $X = [x_1, x_2, x_3, \dots, x_t, \dots, x_s]$, the original input of the input layer can be abbreviated as $X \in R^{s \times d}$, where s and d are the length of time sequence and the feature dimension, respectively. After convolutional operation, X enters the convolutional layer. Convolutional formulas are shown as (1), while in the pooling layer, a max-pooling operation expressed by (2), is applied.

$$a_{c,k} = \text{ReLU}(W_{c,k} * X + b_{c,k}) \quad (1)$$

$$a_p = \max(a_{c,k}) \quad i, j = 1, 2, \dots, n \quad (2)$$

The output layers have an identical structure to the common neural network, expressed by (3). We note that, for the TSA problem, neuron units on the last output layer are set to 1 and the output function is as shown in (4).

$$a_{fc} = \text{ReLU}(a_p W_{fc} + b_{fc}) \quad (3)$$

$$\tilde{y} = \sigma(a_{fc} W_{fc} + b_{fc}) \quad (4)$$

where $\sigma(\cdot)$ is a sigmoid activation function and \tilde{y} , the classification result, indicates the probability of different categories.

2.2 The structure of GRU

GRU [11, 24] realizes the memory and forgetting function of long-term features from input data through its unique update gate and reset gate structure. Compared with an LSTM network, GRU has simpler structure and shorter training time. In addition, compared with recurrent neural network (RNN), it can overcome the difficulty of gradient explosion. GRU's structure is shown in Fig. 2 and the data handling procedures are shown as:

$$R_t = \sigma(W_r \cdot [X_t, H_{t-1}]) \quad (5)$$

$$Z_t = \sigma(W_z \cdot [X_t, H_{t-1}]) \quad (6)$$

$$\tilde{H}_t = \tanh(W_h \cdot [X_t, (R_t \otimes H_{t-1})]) \quad (7)$$

$$H_t = Z_t \otimes \tilde{H}_t + (1 - Z_t) \otimes H_{t-1} \quad (8)$$

where X_t and H_{t-1} denote the original input and the hidden state in the previous period, respectively. W_r , W_z and W_h are the matrices to be trained, and R_t and Z_t are the calculation results of the reset gate and the update gate, respectively. $\sigma(\cdot)$ is used to control the outputs of R_t and Z_t between 0 and 1. The candidate hidden state (\tilde{H}_t) can be obtained through R_t . It should be noted that, in the extreme cases, $R_t = 0$ means discarding the whole previous processing results and $R_t = 1$ means retaining all the results. H_t means the candidate state, and Z_t is used to weigh the proportion of \tilde{H}_t and H_{t-1} about H_t .

2.3 The basic structure of MP CNN + GRU

In order to clearly illustrate MP CNN + GRU, CNN + GRU is clarified first. CNN + GRU, whose structure is depicted in Fig. 3a, is an innovative dual branch network, which can comprehensively acquire the features of original input from local short-term and macro long-term perspectives.

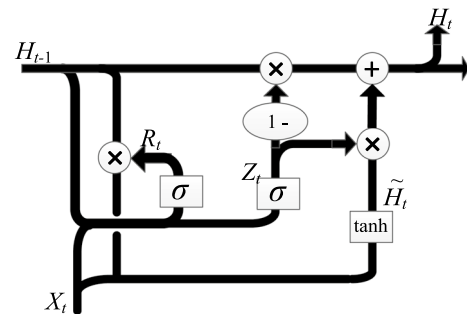


Fig. 2 The structure of GRU

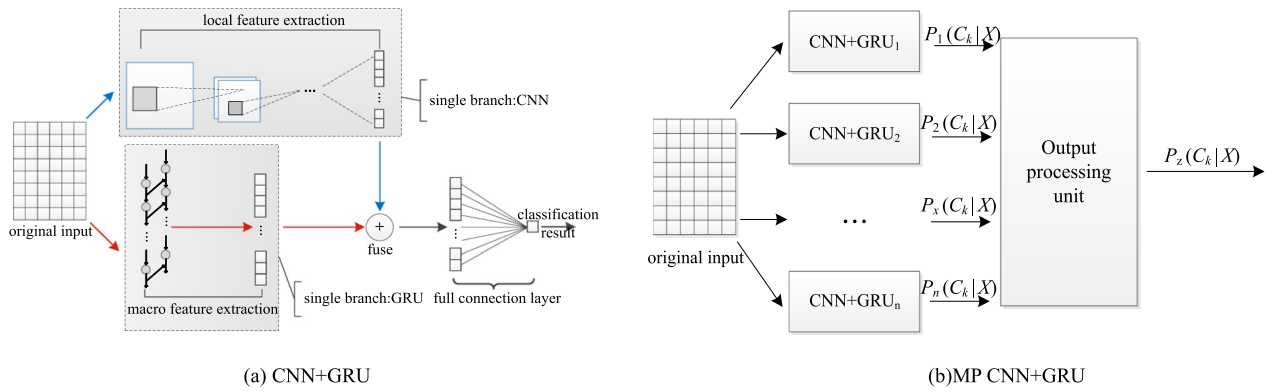


Fig. 3 The structure of MP CNN + GRU

After entering the input layer, the original input simultaneously and spontaneously flows into the CNN branch and the GRU branch. It is worth noting that the convolutional kernels in CNN mainly focus on the local short-term information operation within the corresponding scanning range. The previous convolutional operation has no correlation with the operation in the next step. Therefore, CNN has higher sensitivity to the abstract features contained in voltage magnitude and phase angle with a relatively short variation period. By contrast, GRU has an outstanding long-term forgetting feature and memory function from the macro long-term perspective. It reviews all the previously scanned inputs according to the next input, so as to realize the front hanging and back connection of information. So GRU is adopted to extract the high-order features implied in active and reactive power. These have a long variation period. Thus, both networks work together to complement each other. After the division of labor, the representative information handled by the dual branches is appropriately fused through the full connection layer in order to obtain the classification result.

An individual neural network applied in classification inevitably has a misclassification problem. Basically, the reason is that the neural network has intrinsic deviation when classifying the difficult to classify samples around the classification threshold. In the case of large deviation, the classification result will cross the classification threshold, for example, from the stable to the unstable assessment region, ultimately resulting in misclassification. This can cause enormous damage to the operation of the power system.

To resolve the latent issue, MP CNN + GRU is proposed in this paper. It is composed of CNN + GRU parallel connection and its structure is shown in Fig. 3b. The methodology is rooted in the randomness of neural

network training. The classification results regarding each CNN + GRU sub-model are synthesized to fundamentally solve the misclassification problem originating in intrinsic deviation. Thus, the output processing unit is adopted so as to obtain the final classification result. The methodology regarding output processing unit is shown as:

$$P_Z(C_k|X) = \frac{\sum_{i=1}^n P_i(C_k|X)}{n}, \quad k = 0, 1 \quad (9)$$

where $k=0$ represents a stable sample and $k=1$ means an unstable sample. $P_i(C_k|X)$ represents the probability that sample X is identified as category C_k in the i th CNN + GRU. $P_Z(C_k|X)$ is the final classification result, and is the average number value of category probability output by each CNN + GRU which denotes the probability that X is eventually identified as category C_k after comprehensive analysis.

Considering the real-time and effectiveness requirements of TSA, the number of the CNN + GRU sub-model is set to 3, i.e., n equals 3.

3 TSA based on the MP CNN + GRU

3.1 Feature selection and arrangement of original input

Selecting proper features makes a real difference to the TSA performance of the model, so three dominant factors are taken into consideration:

- (1) Human subjectivity should be significantly reduced.
- (2) Selected features are supposed to reliably summarize the transient fault information of the system [19, 22].
- (3) The arrangement mode of the original input should be appropriately selected because it can affect the readability of the model to the data during training.

Therefore, as mathematically expressed in (10), four kinds of representative and objective features are determined and arranged in the order of bus voltage magnitude, bus phase angle, and active and reactive power of the transmission line, as:

$$X = [V_1, V_2, \dots, V_m, \theta_1, \theta_2, \dots, \theta_m, P_1, P_2, \dots, P_t, Q_1, Q_2, \dots, Q_t] \quad (10)$$

where m represents the number of nodes in the network and t represents the number of transmission lines. All elements of X are vectors with the dimension of d which means the number of sampling points.

3.2 Normalization preprocessing of original input

Common normalization methods include maximum and minimum, mean standard deviation, and so on. In this paper, mean standard deviation normalization is adopted to preprocess the original input, formulated by:

$$x_{normal} = \frac{x - x_{mean}}{x_{std}} \quad (11)$$

The normalization object of (11) is each vector element in (10), such as V_1 , θ_1 , P_1 , and Q_1 . x represents each element in the vector, and x_{normal} is the normalized result of x . x_{mean} denotes the average value of all elements in the corresponding vector, while x_{std} represents the standard deviation of all elements in the vector. Taking P_1 as an example, P_1 is a vector with d sampling points, while x_{mean} and x_{std} are the average value and the standard deviation of active power regarding d elements, respectively. After normalization, all x from P_1 , i.e., active power at each time of the transmission line, are uniformly compressed to 0–1.

3.3 Stability criterion

TSA applying neural network is essentially a classification problem, which needs to label all samples in a huge data set. 0 and 1 are used to label the stable samples and unstable samples, respectively. For a system with large numbers of generators subjected to a large disturbance, the power angle of each generator in the post-disturbance period can be used to compute the transient stability index (TSI) [10, 16, 19]. The TSI formula and label methodology are given respectively as:

$$TSI = \frac{360^\circ - |\Delta\delta_{max}|}{360^\circ + |\Delta\delta_{max}|} \quad (12)$$

$$Label = \begin{cases} 1 & TSI < 0 \\ 0 & TSI > 0 \end{cases} \quad (13)$$

where $\Delta\delta_{max}$ represents the maximal power angle difference between any two generators. If $|\Delta\delta_{max}|$ is greater than 360° ($TSI < 0$), the power system loses stability, and the corresponding sample is marked as 1. In the converse case, the sample is marked with 0.

To ensure the correctness and reliability of sample labeling during simulation, the power angles of generators are selected at least 8 s after fault removal to compute TSI.

3.4 The improved FL function

To clearly explain the improved FL function, the FL function and the binary cross entropy (BCE) function are introduced. The BCE function is the foundation of the FL function and is expressed by:

$$L_{BCE} = \begin{cases} -\ln y' & y = 1 \\ -\ln(1 - y') & y = 0 \end{cases} \quad (14)$$

where y and y' are defined as the real label and the classification probability, respectively. 0 and 1 represent stable samples and unstable samples, respectively.

The FL function expressed by (15) advances the BCE function, and this reduces the weight of easily classified samples and improves the fitting degree of difficult to classify samples.

$$l_{fl} = \begin{cases} -\alpha(1 - y')^\gamma \ln y' & y = 1 \\ -y'^\gamma \ln(1 - y') & y = 0 \end{cases} \quad (15)$$

In (15), α and γ are introduced to l_{fl} . γ is used to address the problem that the classification difficulty regarding different samples is unequal. When y' is exceedingly close to the classification threshold, the corresponding sample is defined as difficult to classify samples and more prone to misclassification. When y' greatly differs from the classification threshold, the sample is defined as an easily classified sample. γ improves the weight of difficult to classify samples in the loss function and reduce the weight of easily classified samples, so as to better fit the difficult to classify samples.

α is to balance the number of stable and unstable samples. As for the TSA problem, the number of stable samples is larger than that of unstable samples. In order to balance samples, previous studies [19, 22] set a fixed α . This entails a large number of time-consuming experiments to obtain an appropriate α . Moreover, the generalization ability of the model is restricted and the engineering efficiency is low. To address the problem, an improved FL function is shown as:

$$L_{fl} = -\frac{1}{S} \sum_{i=1}^S [\alpha y(1-y')^\gamma \ln y' + y'^\gamma (1-y) \ln(1-y')] \quad (16)$$

$$\alpha = S_1/S_2 \quad (17)$$

$$S = S_2 + S_1 \quad (18)$$

Here, a mini-batch training method [25] is adopted and α can be adjusted adaptively in each mini-batch training. This is a kind of simplified stochastic gradient descent (SGD) algorithm [26]. Every time a constant training instance is completed, the parameters are updated. Each parameter update is related, and this can improve the fitting degree of the neural network to data. S_1 and S_2 represent the number of stable and unstable samples in each mini-batch data set, respectively. Therefore, α can be adjusted automatically according to the proportion of stable and unstable samples in each mini-batch training. This greatly reduces unnecessary parameter adjustment processes.

3.5 Model evaluation index

In this paper, four kinds of typical indices are taken as TSA performance indices, and the corresponding formulas are given by:

$$A_{cc} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \quad (19)$$

$$P_{re} = \frac{T_N}{T_N + F_N} \quad (20)$$

$$R_{ec} = \frac{T_N}{T_N + F_P} \quad (21)$$

$$F_1 = \frac{2P_{re}R_{ec}}{P_{re} + R_{ec}} \quad (22)$$

accuracy, denoted by A_{cc} is the most commonly used TSA index, which intuitively evaluates the classification ability of the model. Precision, P_{re} , measures the proportion

of the number of real unstable samples in the classified unstable samples, while recall rate R_{ec} measures the proportion of the number of unstable samples correctly classified in the data set. Because of the contradictory relationship between P_{re} and R_{ec} , F_1 is used to weigh the two indices.

T_P and T_N refer to true stable and true unstable samples, respectively, whereas F_P and F_N refer to false stable and false unstable samples. The relationship between them is shown in Table 1.

If an unstable sample is misclassified as a stable sample, there will be disastrous consequences. In contrast, if a stable sample is misclassified as an unstable sample, a false alarm will appear but it will not cause huge damage to power grid operation. Therefore, F_P is more important than F_N .

3.6 Classification threshold

The selection of the classification threshold is of high significance to the indices of P_{re} and R_{ec} . Because F_P is more important than F_N , it is crucial to reduce the number of F_P . Improving the classification threshold can effectively enhance the conservatism of the model and reduce the number of F_P , so as to improve R_{ec} of unstable samples. The threshold formula is shown as:

$$y = \begin{cases} 0 & P_Z(C_0|X) \leq \gamma, P_Z(C_1|X) > 1 - \gamma \\ 1 & P_Z(C_0|X) \geq \gamma, P_Z(C_1|X) < 1 - \gamma \end{cases} \quad (23)$$

The initial threshold value γ is 0.5. This paper improves the recall rate of unstable samples by manually adjusting γ .

4 Simulation verification

4.1 Data set acquisition

In order to verify the effectiveness of the proposed methods, simulation verification is carried out on the IEEE 39-bus system and IEEE 145-bus system. The IEEE 39-bus system is composed of 39 bus nodes and 46 branches, while the IEEE 145-bus system consists of 145 bus nodes and 453 branches.

To obtain the most representative data set with sufficient data, a degree of freedom in statistics realm is introduced and previous work [19, 22] is fully studied. First, a degree of freedom is adopted to guide the amount of data to be generated. Degree of freedom applied to the TSA problem refers to the required minimum sample quantity to obtain a model that fully fits the data set. Here, the training data set is defined as control points to control the adjustment trend about network parameters, and the model parameters are defined as observation points to observe and output the final TSA results. The ratio is 0.928 on the IEEE 39-bus system and 0.933 on the

Table 1 The confusion matrix

Prediction label (y')	Real label (y)	
	Stable	Unstable
Stable	T_P	F_P
Unstable	F_N	T_N

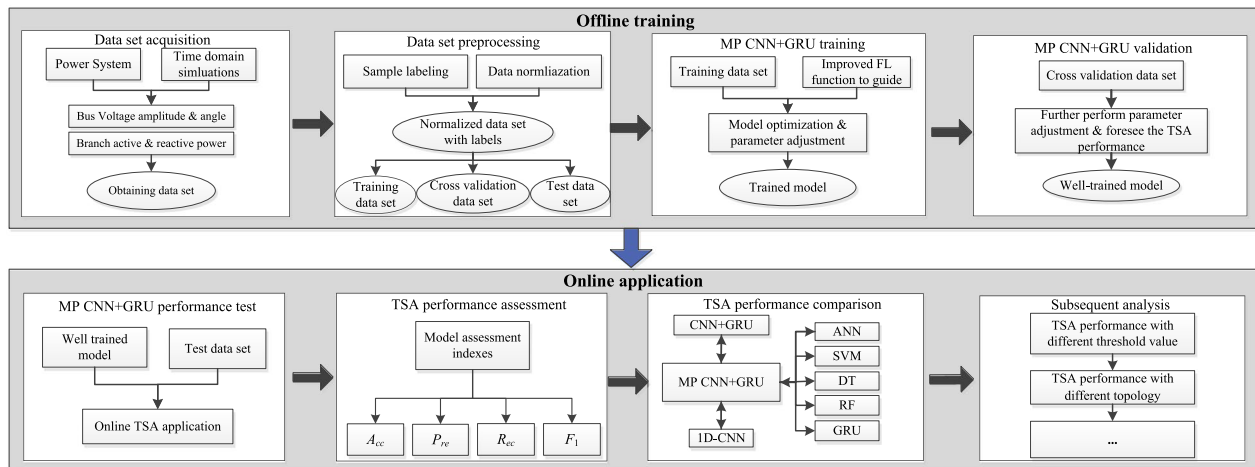


Fig. 4 The overall TSA process

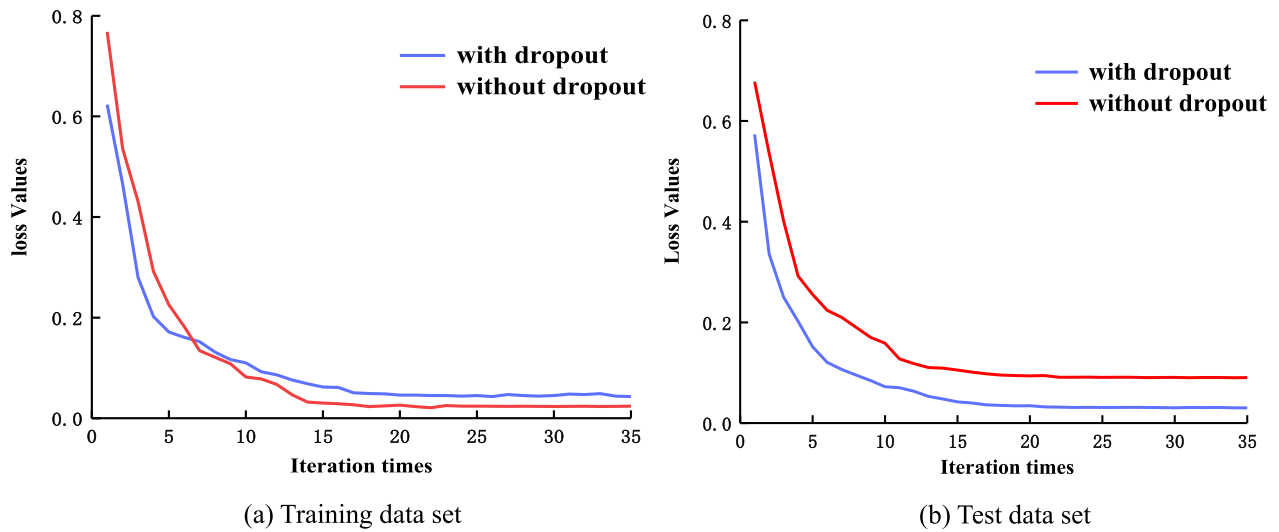


Fig. 5 Learning curve on the IEEE 39-bus system

IEEE 145-bus system, both of which are larger than 0.9 to ensure the quantity of sample data in the obtained data set. Secondly, based on [19, 22], four factors of branches, fault locations, fault durations and load burdens, are taken into consideration to ensure the representation of the obtained data set.

The TDS parameters settings of the IEEE 39-bus system for Power System Simulator/Engineering (PSS/E) are shown in Table 8 in the Appendix. The simulation settings of the IEEE 145-bus system are consistent except that the fault lines are different from Table 8. During the TDS, Python API of PSS/E is used to repeatedly call

PSS/E to implement batch transient simulation. The obtained data set on the IEEE 39-bus system contains 14,280 samples and the number of stable and unstable samples are 11,200 and 3080 respectively, so the ratio of stable samples to unstable samples is approximate 3.6:1. Similarly, the obtained data set on the IEEE 145-bus system consists of 169,260 samples, while the ratio of stable and unstable samples is 5:1. To ensure the correctness and reliability of sample labeling, the sample label is ultimately completed by calculating the TSI value at 10 s. All of the obtained data sets are divided into training, cross validation and test data sets according to the ratio of 3:1:1.

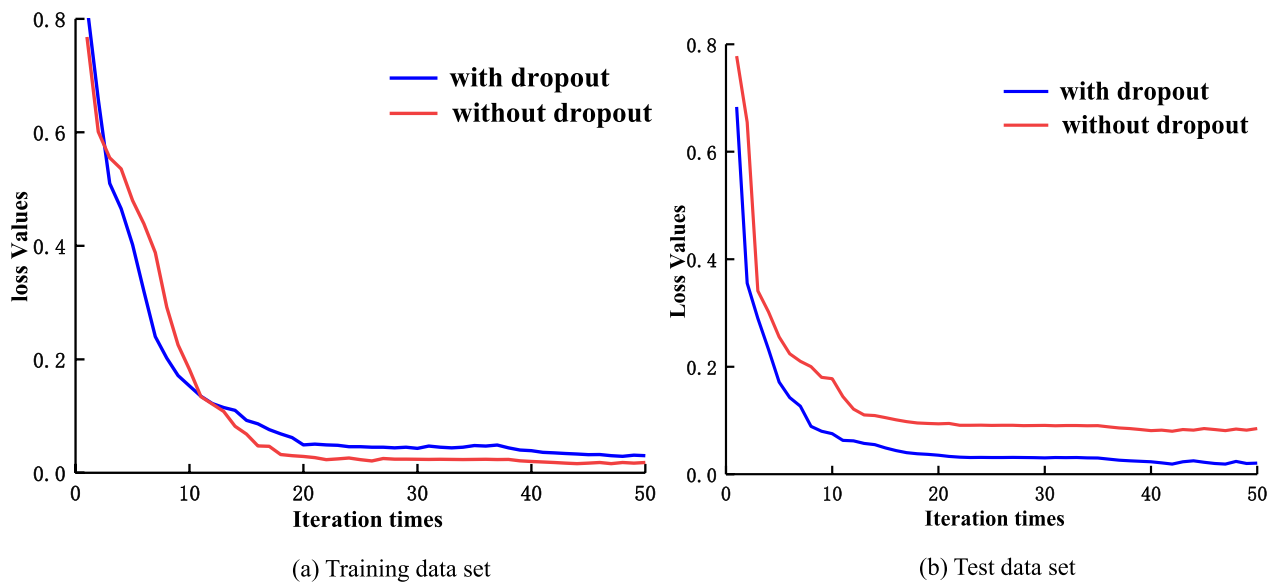


Fig. 6 Learning curve on the IEEE 145-bus system

Table 2 TSA performance on the IEEE 39-bus system

Method	A_{cc} (%)	P_{re} (%)	R_{ec} (%)	F_1
ANN	96.95	90.28	92.29	0.9127
SVM	96.32	96.72	81.94	0.8872
DT	95.97	90.61	86.11	0.8830
RF	96.57	92.12	88.10	0.9006
GRU [11]	97.72	94.01	93.08	0.9354
1D-CNN [19]	98.74	96.25	96.63	0.9644
CNN + GRU	98.91	97.02	96.83	0.9692
MP CNN + GRU	99.40	98.40	98.21	0.9831

Table 3 TSA performance on the IEEE 145-bus system

Method	A_{cc} (%)	P_{re} (%)	R_{ec} (%)	F_1
ANN	96.83	91.32	93.11	0.9220
SVM	96.81	95.62	85.94	0.9052
DT	92.97	93.61	87.11	0.9024
RF	95.57	93.82	88.55	0.9110
GRU [11]	96.72	93.66	93.20	0.9343
1D-CNN [19]	98.66	95.85	95.97	0.9591
CNN + GRU	98.83	97.32	96.91	0.9711
MP CNN + GRU	99.32	98.11	97.88	0.9799

4.2 TSA procedure

The overall TSA procedure, expressed by Fig. 4, mainly consists of two stages, i.e., offline training and online application. At the offline training stage, the obtained

Table 4 TSA result of a single sample on the IEEE 39-bus system

Model	I	II
Prediction results		
$P_1(C_0 X)$	0.496	0.488
$P_2(C_0 X)$	x	0.524
$P_3(C_0 X)$	x	0.510
$P_z(C_0 X)$	0.496	0.507
Prediction label	0	1
Real label	1	1

Table 5 TSA result of a single sample on the IEEE 145-bus system

Model	I	II
Prediction results		
$P_1(C_0 X)$	0.511	0.508
$P_2(C_0 X)$	x	0.487
$P_3(C_0 X)$	x	0.498
$P_z(C_0 X)$	0.511	0.497
Prediction label	1	0
Real label	0	0

data set is divided into training, cross validation and test data sets, and is used to optimize the MP CNN + GRU. The training data set is used to conduct parameter fitting and adjustment of the original model, while the cross validation data set is to foresee the TSA performance of the

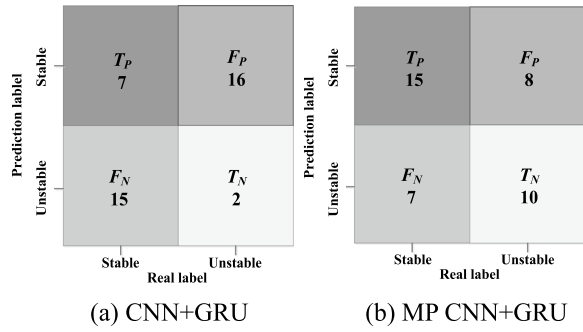


Fig. 7 TSA analysis results on the IEEE 39-bus system

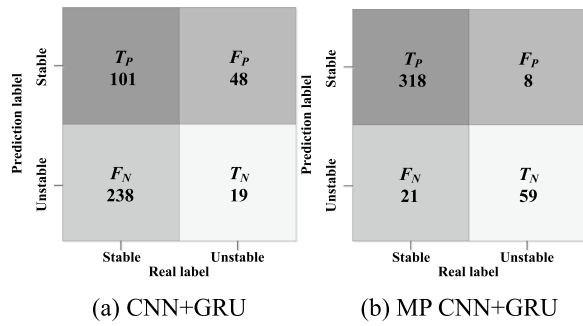


Fig. 8 TSA analysis results on the IEEE 145-bus system

trained model, and more importantly, further complete parameter adjustment. At the online application stage, the trained model is put into practical application on the test data set.

4.3 Model training process

Figures 5 and 6 show the learning curves on the test and training data sets of the IEEE 39-bus system and IEEE 145-bus system. They indicate that, in the last stage of training iteration, the value of the blue line is higher than that of the red line, while the loss value of the red line is very low and approximately 0. However, the loss value concerning the red line is conversely higher than that of the blue line on the test data set. Therefore, the over-fitting phenomenon occurs but it can be greatly alleviated by the dropout method. After a large number of experiments, both dropout rates are determined as 0.2 on the IEEE 39-bus system and IEEE 145-bus system. Most significantly, the overfitting case also indicates that the quantity of the generated sample data is enough for both the IEEE 39-bus system and IEEE 145-bus system.

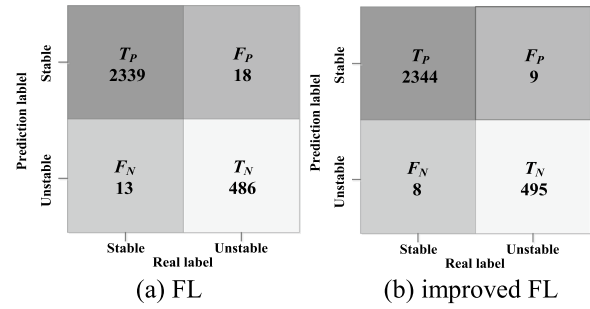


Fig. 9 TSA results on the test data set of the IEEE 39-bus system

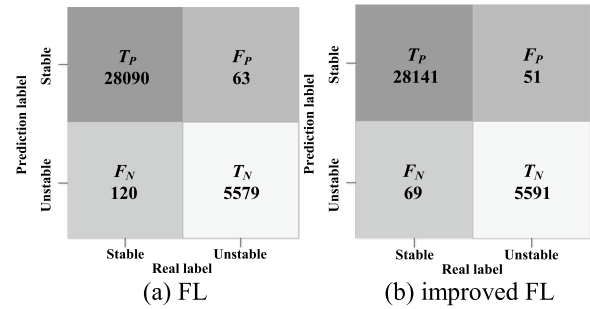


Fig. 10 TSA results on the test data set of the IEEE 145-bus system

The parameter settings of the MP CNN + GRU on the IEEE 39-bus system and IEEE 145-bus system are shown in Tables 9 and 10 in the Appendix, respectively.

4.4 Classification performance comparison

In order to demonstrate the superior TSA capability of MP CNN + GRU, the proposed model is compared with other methods in the ML realm, including ANN, SVM, DT, random forest (RF), GRU [11], 1D-CNN [19], CNN + GRU. For the model fed with one-dimensional vector, such as SVM, the two-dimensional original input is transformed. The data sets used for training and testing of each model are kept consistent with MP CNN + GRU. The dropout rate of ANN is also 0.2, and the activation function adopts *ReLU*, to be consistent with MP CNN + GRU. Eventually, the ANN structure with the best data-fitting performance is 300–300–150–100–1. Because a computation divergence phenomenon emerges from SVM, DT and RF during the simulation process, the principal component analysis (PCA) method is adopted to perform dimension reduction. TSA results of the models are shown in Tables 2 and 3.

The following observations can be drawn from the simulation results:

- (1) Compared with the other four types of methods, ANN, SVM, DT and RF have defective TSA performance. ANN has the highest A_{cc} , up to 96.95% and 96.83%, respectively. However, P_{re} of ANN, being 90.28% and 91.32%, are unacceptable, which indicates that ANN is not accurate enough to classify unstable samples. Although the DT-based RF algorithm has slightly better TSA performance than DT, its R_{ec} of 88.10% on the IEEE 39-bus system and 88.55% on the IEEE 145-bus system are too low.
- (2) CNN + GRU has outstanding TSA performance. As for the IEEE 39-bus system, compared with GRU in [11] and 1D-CNN in [19], R_{ec} is the index with the greatest improvement of 3.75%. As for the IEEE 145-bus system, R_{ec} also has the largest improvement index, rising from 93.20% to 96.91%.
- (3) MP CNN + GRU performs well in both test power systems and further improves the TSA performance of CNN + GRU. The accuracy rates, 99.40% and 99.32%, are maintained above 99%. Thus, MP CNN + GRU has excellent TSA performance.

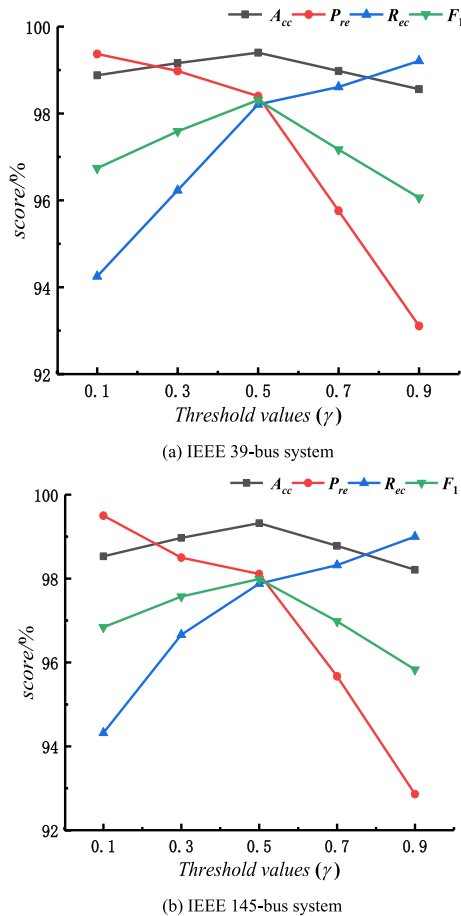


Fig. 11 TSA performance with different classification thresholds

4.5 MP CNN + GRU's TSA performance

To probe the reason for the TSA performance improvement of MP CNN + GRU compared with CNN + GRU, we perform an in-depth study about the classification ability of the proposed model on the IEEE 39-bus system and IEEE 145-bus system from two perspectives: TSA result of an individual sample and TSA results on the test data set.

4.5.1 TSA result analysis of a single difficult to classify sample

Tables 4 and 5 demonstrate the TSA result of a single difficult to classify sample on the IEEE 39-bus system and IEEE 145-bus system, where I and II represent CNN + GRU and MP CNN + GRU, respectively.

The real label of the researched sample on the IEEE 39-bus system is 1. However, the TSA result of the single CNN + GRU is 0 and differs from the real label. Thus, there is a misclassification phenomenon, which can cause serious consequences for the operation of the power system. By comparison, because the distinct multi-parallel structure of MP CNN + GRU can simultaneously output three TSA results of 0.488, 0.524 and 0.510, the final analysis result is 1. Thus, MP CNN + GRU effectively avoids the misclassification of unstable samples into stable samples, and ensures the reliability and correctness of TSA.

Similarly, MP CNN + GRU on the IEEE 145-bus system avoids the sample misclassified into an unstable sample. In summary, for TSA of a single sample, the superiority of MP CNN + GRU is mainly reflected on its strong classification capability regarding difficult to classify samples around the classification threshold.

4.5.2 TSA result analysis of difficult to classify samples on the test data set

After a series of final screening, it is found that there are 40 difficult to classify samples, accounting for about 1.40% of the total samples on the test data set for the IEEE 39-bus system and 406 difficult to classify samples for the IEEE 145-bus system. Figures 7 and 8 show the corresponding TSA results of MP CNN + GRU and CNN + GRU. The results indicate that MP CNN + GRU greatly reduces the number of misclassified samples and improves TSA accuracy. The numbers of F_N and F_P are reduced from 15 to 7 and from 16 to 8 for the IEEE 39-bus system, while for the IEEE 145-bus system they are reduced from 238 to 21 and from 48 to 8. Thus, it proves that MP CNN + GRU has distinctive TSA classification capability in practical application.

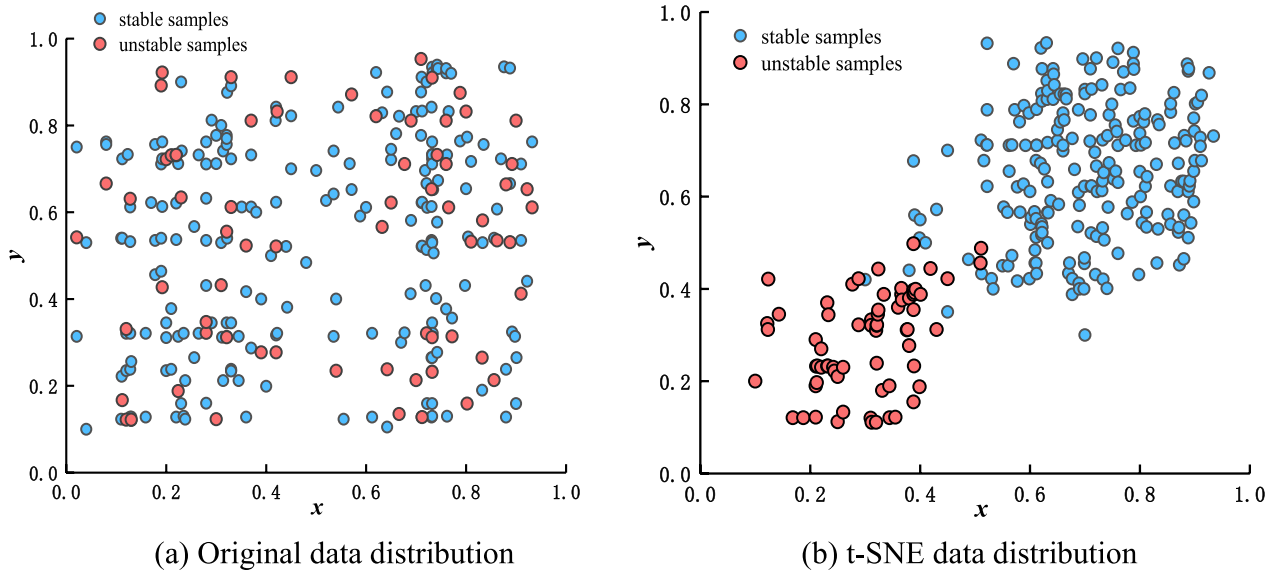


Fig. 12 t-SNE result on the IEEE 39-bus system

4.6 TSA results analysis of the improved FL function

To verify the effectiveness of the improved FL, FL and improved FL are adopted as loss functions to train MP CNN+GRU, respectively, whereas the data sets applied in training and testing remain unchanged. Figures 9 and 10 show the confusion matrices of the two models on the test data set. By contrast, α in the improved FL has no parameter adjustment process and only γ needs to be adjusted continuously.

Clearly, improved FL can balance samples and enhance the fitting degree of unstable samples. Reducing F_p makes a real difference to alleviating damage to the power system. Additionally, the simulation demonstrates that the improved FL can simplify the tedious experimental process and meaningfully build up engineering efficiency.

4.7 TSA results with different classification thresholds

For a power system with highly nonlinear characteristics, the classification information that TSA model needs to acquire has great complexity and the trained model cannot achieve 100% classification accuracy. The MP CNN+GRU proposed in this paper can make the TSA accuracy reach a very high level by virtue of its distinct multi-synchronization, short-term and long-term feature extraction structure. After training, the classification threshold γ can be manually modified to reduce the misclassification phenomenon and improve the recall rate.

Figure 11 illustrates the TSA results with different thresholds on the IEEE 39-bus system and IEEE 145-bus system.

- (1) With the increasing value of γ , P_{re} decreases and R_{ec} increases. Both A_{cc} and F_1 show a tendency of slightly increasing and then decreasing. When γ

Table 6 Comparison of computation speed on the IEEE 39-bus system

Simulation comparison Model	III	IV (s)	V (ms)	A_{cc} (%)
GRU [11]	2856	2.18	0.763	97.72
1D-CNN [19]	2856	2.31	0.809	98.74
CNN+GRU	2856	2.70	0.945	98.91
MP CNN+GRU	2856	3.22	1.127	99.40

Table 7 Comparison of computation speed on the IEEE 145-bus system

Simulation comparison Model	III	IV/s	V/ms	A_{cc} %
GRU [11]	33,852	40.622	1.200	96.72
1D-CNN [19]	33,852	44.006	1.301	98.66
CNN+GRU	33,852	49.381	1.458	98.83
MP CNN+GRU	33,852	80.312	2.372	99.32

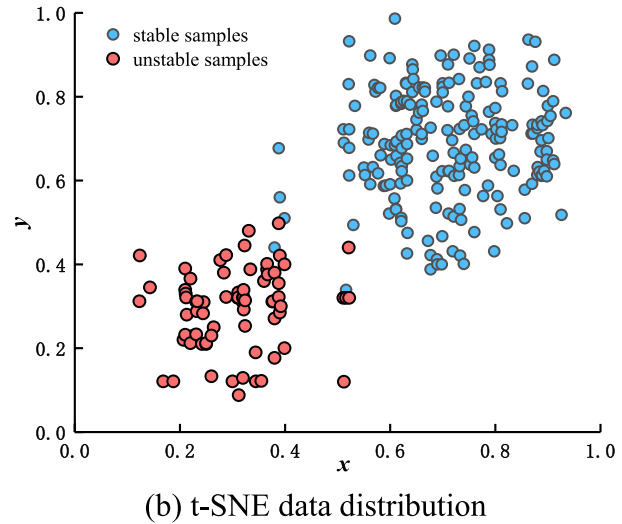
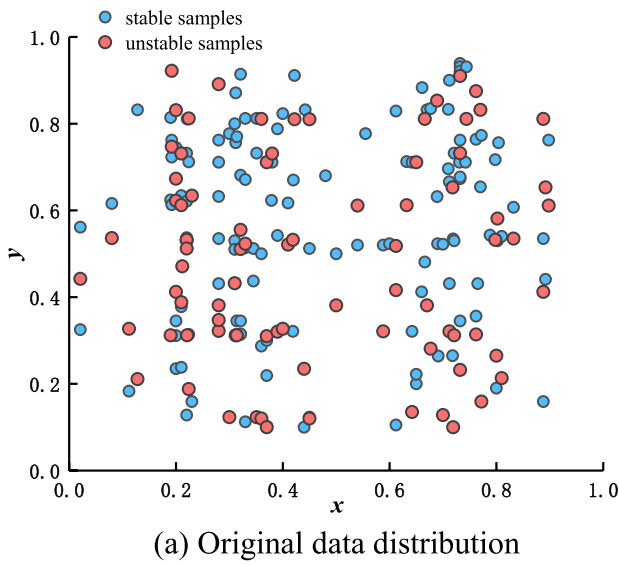


Fig. 13 t-SNE result on the IEEE 145-bus system

equals to 0.5, A_{cc} and F_1 synchronously reach the maximum values of 99.40% and 0.9831 on the IEEE 39-bus system, and 99.32% and 0.9799 on the IEEE 145-bus system, respectively.

- (2) γ with too high or too low value can incur the reduction of A_{cc} , resulting in frequent misclassification and the decline of classification ability, which is not suitable for online TSA. It is worth noting that when γ is 0.9, P_{re} declines to the lowest value. Unacceptable P_{re} leads to frequent false alarms in the power system. When γ equals 0.1, R_{ec} reaches the minimum. Too low a recall rate causes a series of disastrous consequences. It can be inferred from Fig. 11 that when γ is 0, P_{re} reaches 100%. On the other hand, when γ is 1, R_{ec} equals 100%. But selecting 0 or 1 as the classification threshold gives no contribution to the online application.
- (3) The varying γ exerts enormous impact on both P_{re} and R_{ec} , while it has small impact on A_{cc} and F_1 . The resulting variation ranges regarding P_{re} and R_{ec} are approximately 6.30% and 5.00% on the IEEE 39-bus system, and 6.6% and 4.7% on the IEEE 145-bus system, respectively. A_{cc} almost keeps unchanged with a tiny variation range of 0.9% on the IEEE 39-bus system and 1.1% on the IEEE 145-bus system. F_1 is approximately 0.023 on the IEEE 39-bus system and 0.020 on the IEEE 145-bus system.

Thus, after training MP CNN + GRU with high accuracy, manually modifying and adjusting the classification threshold can improve the recall rate concerning unstable samples, reduce the misclassification phenomenon and ensure the conservatism of the model in practical application.

4.8 Visualization analysis of MP CNN + GRU's classification ability

To enhance the TSA interpretability of the model performance, t-distributed stochastic neighbor embedding (t-SNE) [27, 28], as a kind of visualization algorithm, is introduced to show the sample-processing procedure. The distance for the points with great similarity in low-dimensional space is closer after t-SNE. On the contrary, the distance for the points with less similarity is large.

The t-SNE algorithm is to convert the Euclidean distance of high-dimensional data into conditional probability to express the similarity between each sample. The obtained conditional probability is:

$$P(j|i) = \frac{S(x_i, x_j)}{\sum_{k=1, k \neq i}^n S(x_i, x_k)} \quad j \neq i, \quad i = 1, 2, \dots, n \quad (24)$$

where $S(x_i, x_j)$ represents the similarity between i and j . Then, the PCA method is used to reduce the data dimension and to retain the most representative characteristics

of each sample. The conditional probability between each sample after dimension reduction is as follows:

$$Q(j|i) = \frac{S'(z_i, z_j)}{\sum_{k=1, k \neq i}^n S'(z_i, z_k)} \quad j \neq i, \quad i = 1, 2, \dots, n \quad (25)$$

where $S'(z_i, z_j)$ represents the similarity between i and j after dimensional reduction. The closer the distance is, the more similar the two samples are.

As for the TSA problem, stable samples have the same properties, and unstable and stable samples have different properties so the distance between stable samples is relatively short after t-SNE. By comparison, the distance between stable samples and unstable samples is long (Figs. 12 and 13).

4.9 Computational speed analysis about MP CNN + GRU

In order to further verify the superior TSA performance of MP CNN + GRU, its calculation speed is fully analysed in this section. MP CNN + GRU is compared with 1D-CNN in [19] and GRU in [11]. The simulation results of the calculation efficiency and TSA accuracy are shown in Tables 6 and 7.

Here, III represents the number of samples on the test data set. IV and V mean the TSA time of all the samples on the test data set and TSA time of each sample, respectively. It can be seen from Tables 6 and 7 that the computational speed with respect to all of the models can be fully satisfied with the TSA requirement. Although the speed of GRU and 1D-CNN is faster than that of CNN + GRU and MP CNN + GRU, the A_{cc} of GRU and 1D-CNN is lower than that of CNN + GRU and MP CNN + GRU. Additionally, the calculation speed of CNN + GRU is slightly slower than GRU and 1D-CNN because of its complicated dual construction. Finally, within the reasonable range, MP CNN + GRU obtains better TSA performance at the expense of computational speed and the calculation efficiency of MP CNN + GRU on the IEEE 145-bus system is half that of the IEEE 39-bus system, because the input dimension of the IEEE 145-bus system has much more complexity. This brings great difficulty for MP CNN + GRU to implement data analysis. Thus, MP CNN + GRU is a practical method in engineering application.

5 Conclusions

In this paper, a TSA method of CNN + GRU, which is based on CNN and GRU, is proposed. The MP CNN + GRU is then formed by parallel connection of

multiple CNN + GRU, so that the classification accuracy of the difficult to classify samples can be advanced. Finally, the improved FL function which can implement self-adaptive adjustment is proposed to guide model training. The proposed methods are verified by simulations on the IEEE 39-bus system and IEEE 145-bus system. The conclusions are as follows:

- (1) Compared with other AI algorithms and single CNN and GRU algorithms, CNN + GRU can fully extract the high-order features from the micro short-term and macro long-term perspectives, and build the mapping relationship between the original input and the system stability labels. This gives better TSA performance. Its TSA accuracy is up to 98.91% on the IEEE 39-bus system and 98.83% on the IEEE 145-bus system.
- (2) MP CNN + GRU can simultaneously provide multiple TSA through its unique multi-parallel structure. For the TSA result of a single sample, MP CNN + GRU has a certain error correction ability. For TSA on a large number of samples, it can significantly improve the precision and recall rate concerning unstable samples. These two indices are as high as 98.40% and 98.21% on the IEEE 39-bus system, and 98.11% and 97.88% on the IEEE 145-bus system. Thus, MP CNN + GRU has distinctive TSA performance.
- (3) The improved FL function can not only avoid the cumbersome parameter adjustment process and enhance the engineering efficiency in practical application, but also build up the TSA accuracy regarding unstable samples and relieve the disastrous impact of misclassification on the power system. The simulation also indicates that α has no parameter adjustment process.

However, this paper has not considered the TSA performance with single or multiple noisy original inputs. Future research will focus on how to ensure high TSA accuracy of the model with noisy original inputs and the TSA method for optimal PMU configuration of the power system considering economic project cost.

Appendix

See the Tables 8, 9 and 10.

Table 8 The parameter setting of PSS/E batch transient simulation on the IEEE 39-bus system

Types of parameters	Parameter values	Parameter quantity
System frequency (Hz)	60	1
Fault-applied transmission line	1–2, 1–39, 2–3, 2–25, 3–4, 3–18, 4–5, 4–14, 5–6, 5–8, 6–7, 6–11, 7–8, 8–9, 9–39, 10–11, 10–13, 13–14, 14–15, 15–16, 16–17, 16–19, 16–21, 16–24, 17–18, 17–27, 21–22, 22–23, 23–24, 25–26, 26–27, 26–28, 26–29, 28–29	34
Load burden	80%, 90%, 100%, 110%, 120%	5
Fault location	0.1, 0.25, 0.4, 0.5, 0.6, 0.75, 0.9	7
Fault duration time (second)	6/60, 8/60, 10/60, 12/60, 14/60, 16/60, 18/60, 20/60, 22/60, 24/60, 26/60, 28/60	12
Fault type	Three-phase short circuit	1

Table 9 The parameter setting of MP CNN + GRU on the IEEE 39-bus system

CNN branch	Quantity/function
Time sequence dimension	34
Input feature dimension	170
Convolution layer	3/ReLU
Convolution kernel size	5 × 5
Convolution kernel Quantity	64
Pooling layer	3
Pooling size	3 × 3
Pooling method	Max-pooling
Full connection layer units	128
Dropout rate	0.2
GRU branch	Quantity/function
Time dimension(s)	34
Input feature dimension(<i>d</i>)	170
GRUs	128
Full connection layer units	128
Dropout rate	0.2
Full connection layer	Quantity/function
Layers	3/ReLU
Units on the first layer	128
Units on the second layer	64
Units on the last layer	1
Loss function	Improved FL
Optimizer	Adam
Initial learning rate	4×10^{-3}
The first decay rate (β_1)	0.8
The second decay rate (β_2)	0.5
Epsilon (ϵ)	2×10^{-9}
Total parameters of the model	3078

Table 10 The parameter setting of MP CNN + GRU on the IEEE 145-bus system

CNN branch	Quantity/function
Time sequence dimension	34
Input feature dimension	1096
Convolution layer	5/ReLU
Convolution kernel size	7×7
Convolution kernel Quantity	128
Pooling layer	5
Pooling size	5×5
Pooling method	Max-pooling
Full connection layer units	128
Dropout rate	0.2
GRU branch	Quantity/function
Time dimension(s)	34
Input feature dimension(d)	1096
GRUs	256
Full connection layer units	256
Dropout rate	0.2
Full connection layer	Quantity/function
Layers	3/ReLU
Units on the first layer	128
Units on the second layer	64
Units on the last layer	1
Loss function	Improved FL
Optimizer	Adam
Initial learning rate	4×10^{-3}
The first decay rate (β_1)	0.8
The second decay rate (β_2)	0.5
Epsilon (ϵ)	2×10^{-9}
Total parameters of the model	36,284

Abbreviations

TSA: Transient stability assessment; CNN: Convolutional neural network; GRU: Gated recurrent unit; MP: Multiple paralleled; FL: Focal loss; TDS: Time domain simulation; TEF: Transient energy function; EEAC: Extended equal area criterion; ML: Machine learning; ANN: Artificial neural network; SVM: Support vector machine; DT: Decision tree; DL: Deep learning; LSTM: Long short-term memory; SAE: Stacked autoencoder; DBN: Deep belief network; MSDAE: Multi-branch stacked denoising autoencoder; LR: Logistic regression; 1D-CNN: One dimensional-convolutional neural network; PMUs: Phasor measurement units; RNN: Recurrent neural network; TSI: Transient stability index; BCE: Binary cross entropy; SGD: Stochastic gradient descent; PSS/E: Power system simulator/engineering; RF: Random forest; PCA: Principal component analysis; t-SNE: t-Distributed stochastic neighbor embedding.

Acknowledgements

The authors would like to thank the referees and editors of the journals for valuable and constructive comments.

Authors information

Not applicable.

Author contributions

The entire research work has been carried out by ZY, YL and XZ under guidance of SC. The individual contributions of the authors are specified as

follows: Methodology, SC and ZY; Data set acquisition and Validation, YL and XZ; Writing-Original Draft Preparation, ZY; Writing-Review and Editing, SC and XZ; Funding Acquisition, SC and YL. All authors read and approved the final manuscript.

Funding

This research was funded by the National Natural Science Foundation of China under Grant No. 51607105.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations**Competing interests**

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.

Received: 15 March 2022 Accepted: 25 September 2022

Published online: 17 October 2022

References

- O'Shaughnessy, E., Heeter, J., Shah, C., & Koebrich, S. (2021). Corporate acceleration of the renewable energy transition and implications for electric grids. *Renewable and Sustainable Energy Reviews*, 146, 111160.
- Erdiwansyah, M., Husin, H., Nasaruddin, M. Z., & Muhibbuddin, A. (2021). A critical review of the integration of renewable energy sources with various technologies. *Protection and Control of Modern Power Systems*, 6(1), 34–57.
- Telukunta, V., Pradhan, J., Agrawal, A., Singh, M., & Srivani, S. G. (2017). Protection challenges under bulk penetration of renewable energy resources in power systems: A review. *CSEE Journal of Power and Energy Systems*, 3(4), 365–379.
- Cecati, C., & Latafat, H. (2012). Time domain approach compared with direct method of Lyapunov for transient stability analysis of controlled power system. In: *International symposium on power electronics power electronics, electrical drives, automation and motion* (pp. 695–699).
- Chiang, H. (2010). *Direct methods for stability analysis of electric power system theoretical foundation, BCU methodologies, and applications* (pp. 6–8). New Jersey: Wiley.
- Xue, Y., Wehenkel, L., Belhomme, R., Rousseaux, P., Pavella, M., Euxibie, E., Heilbronn, B., & Lesigne, J.-F. (1992). Extended equal area criterion revisited (EHV power systems). *IEEE Transaction on Power System*, 7(3), 1012–1022.
- Zhou, Z., Pu, G., Ma, S., Wang, G., Shao, D., Xu, Y., & Dang, J. (2021). Assessment and optimization of power system transient stability based on feature-separated neural networks. *Power System Technology*, 45(9), 3658–3667.
- Desai, J. P., & Makwana, V. H. (2021). A novel out of step relaying algorithm based on wavelet transform and a deep learning machine model. *Protection and Control of Modern Power Systems*, 6(4), 500–511.
- You, D., Wang, K., Ye, L., Wu, J., & Huang, R. (2013). Transient stability assessment of power system using support vector machine with generator combinatorial trajectories inputs. *International Journal of Electrical Power and Energy Systems*, 44(1), 318–325.
- Matin, R., Yu, C. C., Atefeh, P., Ali, M., & Willian, G. D. (2017). Transient stability assessment via decision trees and multivariate adaptive regression splines. *Electric Power Systems Research*, 142, 320–328.
- Chen, Q., & Wang, H. (2021). Time-adaptive transient stability assessment based on gated recurrent unit. *International Journal of Electrical Power and Energy Systems*, 133, 107156.
- Yu, J. J. Q., Hill, D. J., Lam, A. Y. S., Gu, J., & Li, V. O. K. (2018). Intelligent time-adaptive transient stability assessment system. *IEEE Transactions on Power Systems*, 33(1), 1049–1058.
- Chen, Q., Wang, H., & Lin, N. (2021). Imbalance correction method based on ratio of loss function values for transient stability assessment. *CSEE Journal of Power and Energy Systems*. <https://doi.org/10.17775/CSEEJPES.2021.00290>
- Zhu, Q., Chen, J., Zhu, L., Shi, D., Bai, X., Duan, X., & Liu, Y. (2018). A deep end-to-end model for transient stability assessment with PMU data. *IEEE Access*, 6, 65474–65487.
- Tan, B., Yang, J., Tang, Y., Jiang, S., Xie, P., & Yuan, W. (2019). A deep imbalanced learning framework for transient stability assessment of power system. *IEEE Access*, 7, 81759–81769.
- Su, T., Liu, Y., Zhao, J., & Liu, J. (2022). Deep belief network enabled surrogate modeling for fast preventive control of power system transient stability. *IEEE Transactions on Industrial Informatics*, 18(1), 315–326.
- Wu, S., Zheng, L., Hu, W., Yu, R., & Liu, B. (2020). Improved deep belief network and model interpretation method for power system transient stability assessment. *Journal of Modern Power Systems and Clean Energy*, 8(1), 27–37.
- Liu, W., Hao, D., Zhang, S., & Zhang, Y. (2021). Power system transient stability assessment based on PSO-DBN. In: *2021 6th international conference on power and renewable energy (ICPRE)* (pp. 333–337).
- Gao, K., Yang, S., Liu, S., & Li, X. (2019). Transient stability assessment for power system based on one-dimensional convolutional neural network. *Automation of Electric Power Systems*, 43(12), 18–26.
- Tian, F., Zhou, X., Shi, D., Chen, Y., Huang, Y., & Yu, Z. (2019). Power system transient stability assessment based on comprehensive convolutional neural network model and steady-state feature. *Proceedings of the CSEE*, 39(14), 4025–4032.
- Shi, Z., Yao, W., Zeng, L., Wen, J., Fang, J., Ai, X., & Wen, J. (2020). Convolutional neural network-based power system transient stability assessment and instability mode prediction. *Applied Energy*, 263, 114586.
- Zhao, K., & Shi, L. (2021). Transient stability assessment of power system based on improved one-dimensional convolutional neural network. *Power System Technology*, 45(8), 2945–2957.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. In: *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1746–1751).
- Pan, E., Ma, Y., Dai, X., Fan, F., Huang, J., Mei, X., & Ma, J. (2019). GRU with spatial prior for hyperspectral image classification. In: *IGARSS 2019–2019 IEEE international geoscience and remote sensing symposium* (pp. 967–970).
- Gou, P., & Yu, J. (2018). A nonlinear ANN equalizer with mini-batch gradient descent in 40Gbaud PAM-8 IM/DD system. *Optical Fiber Technology*, 46, 113–117.
- Rios, D., & Jüttler, B. (2022). LSPIA, (stochastic) gradient descent, and parameter correction. *Journal of Computational and Applied Mathematics*, 406, 113921.
- Laurens, V. D. M., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2605), 2579–2605.
- Gisbrecht, A., Schulz, A., & Hammer, B. (2015). Parametric nonlinear dimensionality reduction using kernel t-SNE. *Neurocomputing*, 147(1), 71–82.