A multi-task transient stability assessment method adapted to topology changes using multi-graph sample and aggregate-attention network

Lingxiang Huang[1], Kun Dong[2]*, Jianfeng Zhao[2], Kangli Liu[2], Cheng Jin[2] and Xirui Guo[2]

[1]College of Software Engineering, Southeast University, Nanjing, China, [2]School of Electrical Engineering, Southeast University, Nanjing, China

Transient stability assessment (TSA) plays a pivotal role in guiding power grid risk control strategies. However, it faces challenges when dealing with complex multi-graph inputs generated by pre-fault, fault occurrence, and post-fault states. Meanwhile, most previous research studies neglected the assessment of the transient stability level. To address this, we propose a multi-task transient stability assessment (MTTSA) approach. In MTTSA, we introduce a multi-graph sample and aggregate-attention network (GraphSAGE-A) designed to capture stability features even amidst topology changes. A multi-head attention mechanism and local normalization method are adopted for a better extraction of the global and contextual information. Additionally, we introduce a quantified transient stability risk index considering the transient stability boundary and incorporate a multi-task dense structure to enhance MTTSA's performance. Empirical tests, under changing operating conditions, conducted on the IEEE 39-bus system showcase a significant performance improvement with the proposed MTTSA method.

KEYWORDS

multi-task transient stability assessment, quantified transient stability risk index, transient stability boundary, topological change, multi-graph sample and aggregate-attention network

1 Introduction

With the increasing penetration of renewable energy and widespread utilization of power electronic devices, the modern power system becomes less resistant to disturbances and, thus, requires more reliable and effective transient stability assessment (TSA) methods. On one hand, the accuracy and speed of traditional analysis methods have been challenged. On the other hand, most previous studies concentrated on the binary stability category, which rarely concerned a quantified stability risk assessment under the influence of TSB. In such contexts, it is essential to adopt a multi-task transient stability assessment (MTTSA) method in order to predict the risk levels and activate emergency control on time.

The practical implementation of TSA is facilitated by the rapid proliferation of phasor measurement units (PMUs) (Gurusinghe and Rajapakse, 2016) and the development of feature extraction technology in power grids. PMUs provide real-time load monitoring data,

TABLE 1 Comparison of previous studies.

Method	Advantage	Disadvantage
Traditional TSA	High accuracy with appropriate simulation	Large calculating time costs
Machine learning models	Short-time response and high reliability	Sensitive to the noise of data
Deep learning models	Short-time response; high reliability; and robustness	Fail to recognize topology features
GCN	Short-time response; high reliability; robustness; and topology extraction	Lack of topological generalization ability
GraphSAGE and GAT	Short-time response; high reliability; robustness; and adaptability to topology	Lack of quantified risk assessment

and modern artificial intelligence algorithms can effectively extract features of large amounts of data.

As is shown in Table 1, traditional TSA is previously combined with the direct method of transient stability analysis and the time-domain simulation technique (Maria et al., 1990). To improve the accuracy of prediction, researchers proposed a fully time-domain simulation (TDS)-based method (Diao et al., 2017), which solves high-dimensional nonlinear differential-algebraic equations (DAEs) of the power system using numerical integration algorithms. However, the results of this numerical differentiation greatly depend on the parameters of power grid models, such as synchronous generators, loads, transmission lines, and emergency control devices. In addition, the increasing time costs brought by the growing scale of the power system limits the speed of the TDS-based method (Senyuk et al., 2023).

The goal of online TSA is realized due to the development of machine learning (ML) and deep learning methods. Liu et al. (2020) and Liu et al. (2021) developed the transient stability margin with the concept of critical clearing time (CCT), which is tested to easily change through multiple time-domain simulations. The qualitative methods based on the analysis of the kinetic energy of a generator rotor before, during, and after a contingency gradually develop stability criteria and form the trajectory-based transient stability index (TSI) (Glavic et al., 2007; Xu et al., 2015). The deployment of high-precision PMUs (Gurusinghe and Rajapakse, 2016) has led to an increase in the rate of power system parameter sampling, which, in turn, supports the utilization of large data in ML-based methods. The model-free ML methods, such as artificial neural network (ANN) (Siddiqui et al., 2018), support vector machine (SVM) (Gomez et al., 2011), and random forest (Mukherjee and De, 2020), can solve the problem of TSA. ANNs with a parallel structure were first applied to TSA (Siddiqui et al., 2018). Nonetheless, ANN-based TSA algorithms have not been widely adopted due to computational limitations. SVMs were used to process the binary classification task of TSA with short-time response and high reliability, but it is sensitive to the inevitable noise of input data. There are decision tree-based methods including fuzzy decision tree (Kamwa et al., 2012) and random forest (Mukherjee and De, 2020) with drawbacks of the proneness to re-training. Overall, the mentioned TSA algorithms are willing to overfit restrictions by generalization ability.

Different from traditional ML methods, deep learning models are capable of handling significantly larger-scale data by automatically generating features from raw input. Yu et al. (2018) developed a TSA system based on the long short-term memory network (LSTM) to capture the long-term dependencies along the time steps of time series. Furthermore, an LSTM-based gated recurrent unit is added to a two-stage TSA method for the analysis of uncertain samples after the first stage (Zhan et al., 2022). Convolutional neural network (CNN) is also widely used in many outstanding TSA models (Gupta et al., 2019; Shi et al., 2020). The time series sampled by PMUs are mapped into heatmap representation (Kamwa et al., 2012) or processed as a multi-channel matrix (Shi et al., 2020) to leverage the strengths of CNN in learning representations from images. Additionally, the gated recurrent unit model is developed to process the time-adaptive TSA (Chen and Wang, 2021). Moreover, an attention mechanism is introduced and combined with bidirectional (Bi)-LSTM to extract more robust features (Mahato et al., 2021). Nevertheless, a power grid inherently exhibits a structured graph representation, where buses are modeled as nodes and transmission lines are modeled as edges (Ishizaki et al., 2018). CNN, LSTM, or Bi-LSTM attention mechanisms failed to effectively recognize the system's topological features, which has an impact on TSA tasks.

Fortunately, the introduction of graph neural networks (GNNs) provides a promising approach for addressing the aforementioned issues. Although bus-modeled nodes possess numerous temporal features, the effectiveness of combining the TSA method based on random walk-based GNN is not satisfactory. The GNN models, including graph convolutional network (GCN) (Huang et al., 2020), graph sample and aggregate network (GraphSAGE) (Lin et al., 2022), and graph attention network (GAT) (Huang et al., 2021), effectively handled both the temporal and system topological features of nodes and are widely utilized in TSA tasks. Huang et al. (2020) proposed the online TSA comprising GCN, which explicitly integrates the bus (node) states with the topological characteristics. The GCN model is also capable of power grid fault location prediction and load shedding decision (Kim et al., 2019; Chen et al., 2020). As variants of GCN, GraphSAGE and GAT models enable the automatic generation of small-scale topological changes from raw samples.

The trajectory-based TSI observed in power systems is widely applied for the qualitative assessment of transient stability, but the fixed TSB affects the overall generalization ability of TSA methods for diverse samples (Yan et al., 2021). Zheng et al. (2017) transformed the TSA problem into a relationship between state space using time series approximation and the region of attraction, while Du and Hu (2021) and Li et al. (2021) quantified the transient stability risk of uncertain samples using the TSB associated with sample orientation in a two-stage LSTM-based approach. A trajectory-based piece-wise stability index (PSI) was developed by Huang et al. (2021) to describe the transient stability level instead of CCT.

In our study, we design a multi-graph sample and aggregate-attention network (GraphSAGE-A) for MTTSA feature aggregation. A multi-head attention mechanism is utilized to replace the global information aggregation from the neighbors, while the extraction for contextual features is addressed by the local normalization for edge weights. A TSI-based transient stability prediction task (TSP) and a quantified transient stability risk prediction task (QRP) considering the uncertain samples of TSP work in parallel and share the features abstracted by GraphSAGE-A. A transient stability risk index (RI) is developed to instruct the subsequent risk control. The main contributions of this paper are summarized as follows:

1) A multi-graph GraphSAGE-A model requires the system dynamics before, during, and after a fault (t_{0-}, t_{0+}, and t_{c+}) as input and realizes TSA with good precision even under topological changes.
2) A multi-head attention mechanism is adopted to extract the global and contextual information better, while the local normalization method for edge weights is proposed to improve the generalization ability for the subgraph.
3) A multi-task method concurrently considers the transient stability qualitative and quantitative analyses. An RI, considering the variability across different samples, is proposed to quantify the extent of the deviation from TSB.

The rest of this paper is organized as follows. Section 2 proposes the topology adaptive MTTSA method, while Section 3 presents the designs of GraphSAGE-A in detail. Experiments are conducted on the IEEE 39-bus system and discussed in Section 4. Finally, the conclusion is provided in Section 5.

2 Topology adaptive multi-task transient stability assessment method

2.1 MTTSA problem modeling

From a mathematical perspective, the power system TSA can be modeled by high-dimensional nonlinear DAEs with initial conditions at three snapshots as follows:

$$\begin{cases} \dfrac{dx}{dt} = f(x, y) \\ 0 = g_G(x, y) \\ (x, y)\big|_{t_{0-}} = (x_0, y_0), \\ G\big|_{t_{0-}} = G_{0-} \\ G\big|_{t_{0+}} = G_{0+} \\ G\big|_{t_{c+}} = G_{c+} \end{cases} \quad (1)$$

where vector x represents the state variables including dynamic generator rotor angles and bus voltages, while vector y denotes the algebraic variables (e.g., the static impedance of transmission lines). $G\big|_{t_j}$, $j = 0-, 0+, c+$ represents the parameterized network topologies of the graph type (Ishizaki et al., 2018) at the jth sampling shot before, during, and after a fault.

The mentioned data-driven TSA methods, using ML (Kamwa et al., 2012; Siddiqui et al., 2018) or deep learning models (Ishizaki et al., 2018; Yu et al., 2018), calculate the full-time series features of the period $\{(x, y)\big|_t\}$, $t \in t_{0-} \cup [t_{0+}, t_{c+} + k\Delta t]$. Here, Δt stands for a time step with $k > 0$. The pre-fault one, which evaluates the transient stability for a pre-defined disturbance setting, relies on numerical analysis to solve DAEs. In contrast, the post-fault one focuses on evaluating the stability trends during the occurrence of a contingency, which leads to increased demands for dynamics from PMUs after the fault. Both pre- and post-fault methods explored the correlations among the state variables, particularly the relationship between generator rotor angles and other state variables. Nevertheless, they failed to provide fast predictions because of the time-consuming TDS or a significant number of PMU sampling.

Three snapshots (t_{0-}, t_{0+}, and t_{c+}) utilized instead of the period $\{(x, y)\big|_t\}$, $t \in t_{0-} \cup [t_{0+}, t_{c+} + k\Delta t]$ realized online TSA by avoiding the vast majority of TDS or PMU sampling. However, we need to adopt the MTTSA method with a better ability for feature extraction, especially the topological structure.

2.2 Design of the MTTSA method

Regarding the significant requirement for topology representation learning, GraphSAGE-A becomes the critical point of the proposed MTTSA method, as shown in Figure 1. The aggregate-attention layers in the multi-graph node embedding module generate features for TSP and QRP modules. Considering the ensemble learning, we develop two predictor tasks, where TSP refers to the binary classification problem and QRP constructs the regression model. Both TSP and QRP belong to the full-connected network. Then, the MTTSA method is divided into the joint-training stage and the applying stage. In the joint-training stage, the sample loader structure allows mini-batch training for GraphSAGE-A, while TSP and QRP share the output from the multi-graph node-embedding module.

Meanwhile, the design with the labels of various tasks guarantees stability concerning coarse- and fine-grained embedding captured. As a result, richer inter-task information leads to the better performance of both TSP and QRP. During the online applying stage, state variables from TDS or PMU sampling are input into the trained GraphSAGE-A to generate the stable possibility $\hat{p}$ from TSP and the predicted RI $\hat{v}$ from QRP.

3 Multi-graph sample and aggregate-attention network

3.1 Model input and data preprocessing

Considering the dynamic embedding from TDS or PMU sampling, the amplitude and phase angle of voltage and the active and reactive power of both generators and loads had a total of four dimensional features for the $i-th$ bus at the $j-th$ snapshot. The raw input of the $K-th$ sample can be described as the following matrix $X_K \in R^{m \times 4n}$:

FIGURE 1
MTTSA method.
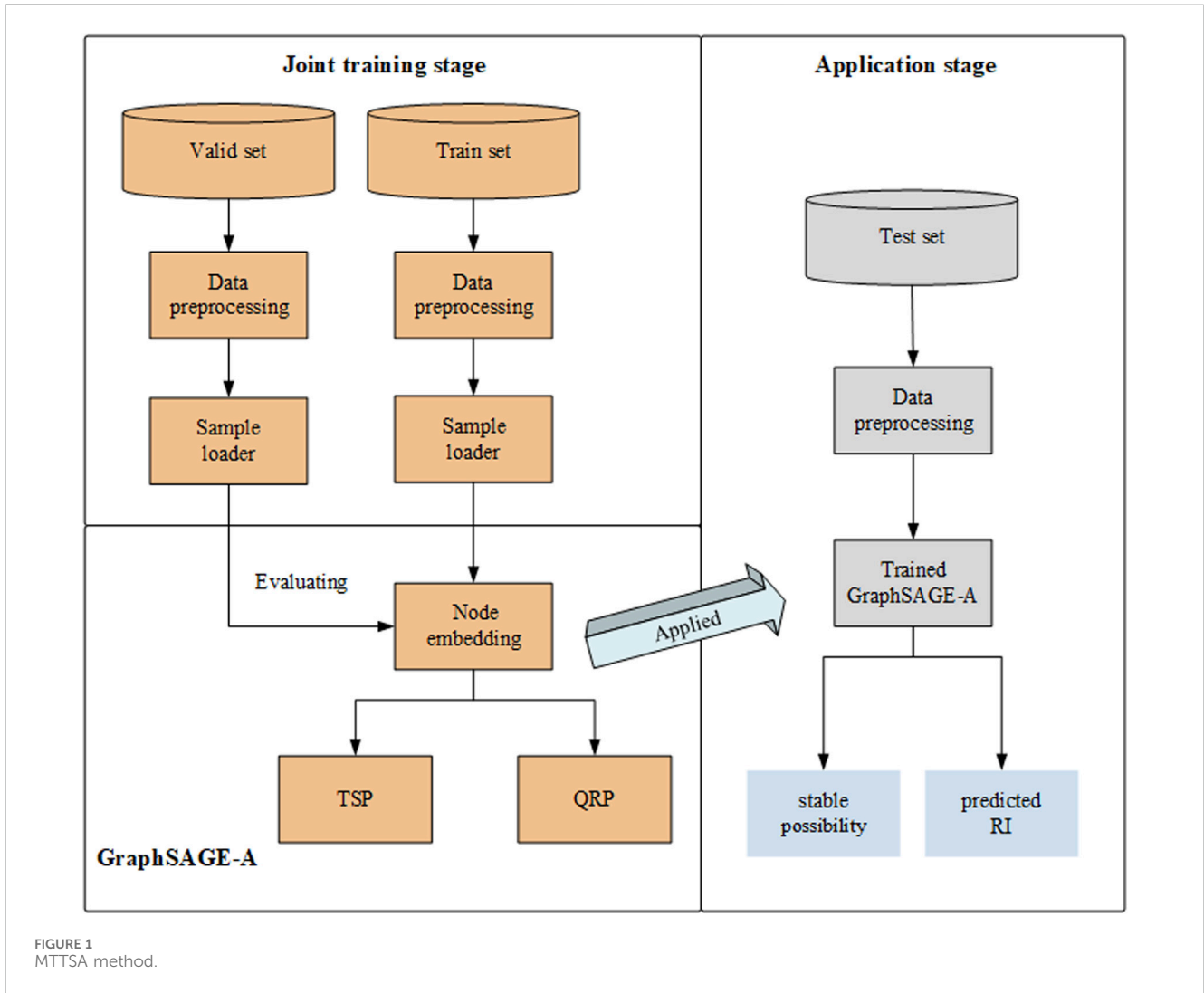
$$X_K = \begin{pmatrix} U_{11}, \theta_{11}, P_{11}, Q_{11}, & \cdots, U_{1i}, \theta_{1i}, P_{1i}, Q_{1i}, & \cdots, U_{1n}, \theta_{1n}, P_{1n}, Q_{1n} \\ \vdots & \ddots & \vdots \\ U_{j1}, \theta_{j1}, P_{j1}, Q_{j1}, & \cdots, U_{ji}, \theta_{ji}, P_{ji}, Q_{ji}, & \cdots, U_{jn}, \theta_{jn}, P_{jn}, Q_{jn} \\ \vdots & \ddots & \vdots \\ U_{m1}, \theta_{m1}, P_{m1}, Q_{m1}, & \cdots, U_{mi}, \theta_{mi}, P_{mi}, Q_{mi}, & \cdots, U_{mn}, \theta_{mn}, P_{mn}, Q_{mn} \end{pmatrix}, \quad (2)$$

where U_{ji}, θ_{ji}, P_{ji}, and Q_{ji} represent the voltage amplitude, voltage phase angle, active power, and reactive power of the $i-$th node at the $j-$th shot, respectively. Following the z-score normalization, the data are performed as follows:

$$X^\star = \frac{X - \mu}{\sigma}. \quad (3)$$

$X^\star$ is the normalized value of X, and μ and σ are the mean and standard deviation of the features, respectively.

The number of input graph nodes refers to the total buses of the power grid, while the information of edges refers to the transmission lines. The multi-graph representing the single sample is defined as follows:

$$G_{Kj} = \left(V_{Kj}, A_{Kj} \right). \quad (4)$$

G_{Kj} represents the graph of the $K-$th sample at the $j-$th snapshot, while V_{kj} represents the $j-$th row of the normalized X_K. A_{Kj} is the adjacency matrix. The formula is

$$A_{Kj}[u][v] = \begin{cases} 0 & (u, V) \notin E_{kj} \\ |Z_{m(v,u)}| & (V, V) \in E_{kj}, \\ 1 & u = v \end{cases} \quad (5)$$

where $A_{Kj}[u][v]$ are the element of the matrix A_{Kj} following the location (u, v), while v and u are any two node indices of the graph G_{Kj}. $E_{Kj}[u][v]$ and $|Z_{Kj(v,u)}|$ denote the existence and static impedance of a transmission line from the $v-$th bus to the $u-$th bus at the $j-$th snapshot of the $K-$th sample, respectively.

3.2 GraphSAGE

For each node targeted of the given G_{Kj}, the GraphSAGE model samples the neighbors and aggregates their features to the neighbors' vector $h_{N(v)}^k$ depending on the representations of the previous stage:

$$h_{N(v)}^k = \text{AGG}_k\left(\text{WeightNorm}\left(A_{Kj}[u][v] \right) * h_u^{k-1}, \forall u \in N(v) \right), \quad (6)$$

where k refers to the k − th stage and N(v) denotes the neighbor indices of the target node index v. The 1-hop neighbors indices, whose corresponding element with v in the adjacency matrix is not 0 or 1, are usually utilized according to the size of the graph. AGG_k is defined as the method at the k − th stage to process the representations, including mean, pooling and LSTM. In addition, the initial representations of node indices are the normalized node features.

$$h_v^0 = \left(U_{jv}^*, \theta_{jv}^*, P_{jv}^*, Q_{jv}^* \right) \in V_{Kj}. \qquad (7)$$

U_{jv}^*, θ_{jv}^*, P_{jv}^*, and Q_{jv}^* are the elements of node features V_{Kj}. Here, we adopt the local normalization of the adjacency matrix to meet the requirements of local aggregation as needed by the GraphSAGE algorithm. The method is

$$\text{WeightNorm}\left(A_{Kj}[u][v] \right) = \frac{A_{Kj}[u][v]}{\sum_{h \in N(v)} A_{Kj}[h][v]}, \forall u \in N(v), \qquad (8)$$

where $\text{WeightNorm}(A_{Kj}[u][v])$ refers to the result of local normalization for the edge from the u − th node to the v − th node and $A_{Kj}[u][v]$ stands for the element of A_{Kj} according to the location (u, v).

Then, the node index v tends to form its own vector h_v^k at the k − th stage for the next step such that

$$h_v^k = \sigma\left(W^k \cdot \text{CONCAT}\left(h_v^{k-1}, h_{N(v)}^k \right) \right), \qquad (9)$$

where W^k is the weight matrix for the training of the k − th stage. σ is usually the ReLU function for activation, and $CONCAT$ denotes the concatenating method for the input given.

3.3 Graph attention

The association degree differs between different nodes (Ding et al., 2022). It means the differences between global and contextual information cannot be performed in the weight set by the aggregation functions of the GraphSAGE model. A graph attention mechanism can compute the relationship between any two nodes regardless of their adjacency to extract the global view in the graph. Hence, the graph attention mechanism is used as the aggregation method to weigh the significance of the nodes among the neighbors. Furthermore, a multi-head graph attention mechanism allows multiple graph attention blocks to work in parallel to aggregate more contextual information at the same time.

First, a multi-head function is defined as follows:

$$\text{Multihead}(m)_{(v,u)}^k = \begin{pmatrix} \text{CONCAT}\left(Wh_v^{k-1,1}, Wh_u^{k-1,1}, \right) \\ ... \\ \text{CONCAT}\left(Wh_v^{k-1,m}, Wh_u^{k-1,m}, \right) \end{pmatrix}, \qquad (10)$$

where $\text{Multihead}(m)_{(v,u)}^k$ stands for the totally m heads attention processing the matrix at the k − th stage mapped by all the single typical attention processing matrixes for any two node indices from the 1st step to the m − th step. Now, we can calculate the weight α_{vu}^k between the node index u and v via a softmax function.

$$\alpha_{vu}^k = \frac{\exp\left(\text{LeakyReLU}\left(a^T \text{Multihead}(m)_{(v,u)}^k \right) \right)}{\sum_{j \in N(v)} \exp\left(\text{LeakyReLU}\left(a^T \text{Multihead}(m)_{(v,j)}^k \right) \right)}. \qquad (11)$$

In the formula, a^T denotes the parameter vector of the network, and LeakyReLU is a nonlinear activation function. Finally, the multi-head graph attention output for every node index can be expressed as follows:

$$h_v^k = \sigma\left(\sum_{j \in N(v)} \text{WeightNorm}\left(A_{Kj}[j][v] \right) \cdot \alpha_{vj}^k \cdot Wh_v^{k-1} \right). \qquad (12)$$

Figure 2 shows the aggregation for the target node given in GraphSAGE or GraphSAGE-A. The local normalization strategy allows the reshaping of edge weights for the subgraph.

Figure 3 shows the design of the node-embedding module as well as the dense layers for TSP and QRP. The multi-head attention mechanism considers the global and local structure by weighing the more important nodes. The three-layer structure, input-size, and out-size for each layer are designed according to the node numbers per graph and the dimension of node features.

3.4 TSP and QRP in dense

The output of the node-embedding module is concatenated by the outputs H_m^L of the three graphs for the K − th sample as follows:

$$\bar{H}_K = \text{CONCAT}(H_m^L), m = 1, 2, 3. \qquad (13)$$

Finally, the embedding matrix is formed as a vector, which is processed by full-connect multi-layer perceptron (MLP) modules: TSP and QRP. TSP refers to the classification task, whereas QRP refers to the regression task. We adopt the softmax and softsign functions for the output of TSP and QRP as (14) and (15), respectively.

$$\hat{p}_i = \frac{\exp\left(z_i^{TSP} \right)}{\sum_i \exp\left(z_i^{TSP} \right)}, \qquad (14)$$

$$\hat{v} = \frac{z^{QRP}}{1 + |z^{QRP}|}. \qquad (15)$$

The confidence $\hat{p}_i$ refers to the i − th category, while $\hat{v}$ denotes the predicted stability risk index.

3.5 A transient stability risk index

There is a three-dimensional label for each sample input in the GraphSAGE-A. The first and second dimensions of one-hot label refer to the TSP task via a trajectory-wise TSI as follows (Rahmatian et al., 2017):

$$TSI_K = \frac{360° - |\Delta\delta_{max}|_K}{360° + |\Delta\delta_{max}|_K}, \qquad (16)$$

where $|\Delta\delta_{max}|_K$ refers to the maximum rotor angle difference between any two generators during the sampling or simulation of the K − th sample. When $TSI_K < 0$, the K − th sample is unstable. Otherwise, the system is stable. It appears that the TSI defines a stability margin using a fixed value, while the TSB shifts slightly. Therefore, a transient stability RI is proposed to measure the stability deviation so that the special subsequent control schemes can be considered.

FIGURE 2
Sampling and aggregation process.

FIGURE 3
Structure of GraphSAGE-A.
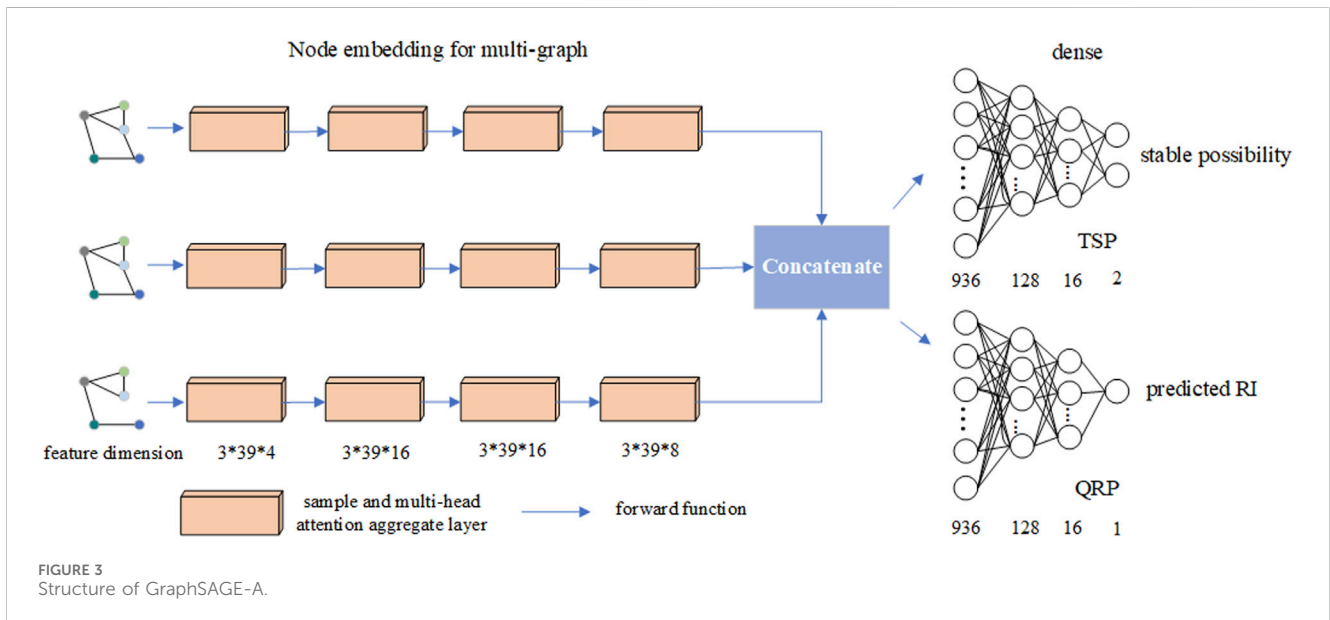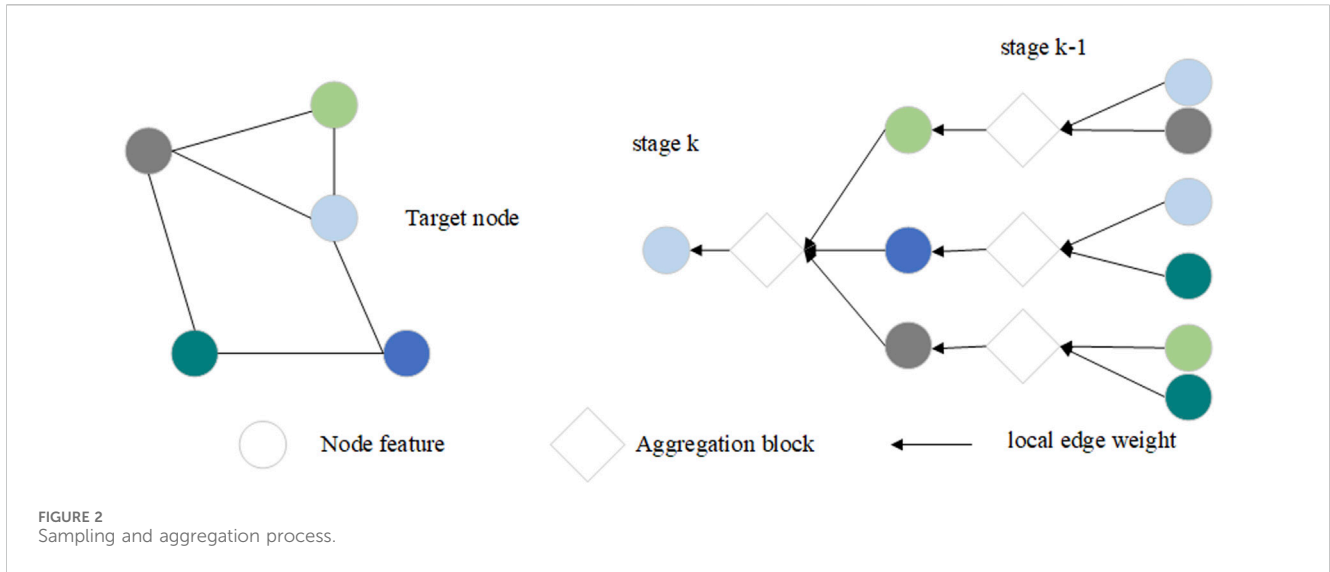
A fixed deviation Δ_K of the $K-th$ sample is defined as follows:

$$\Delta_K = |(360° - |\Delta\delta_{max}|_K)|. \qquad (17)$$

A normalized CCT is also considered as follows:

$$t_K^* = \frac{t_K - \mu_{CCT}}{\sigma_{CCT}}, \qquad (18)$$

where t_K^* and t_K^* refer to the normalized and original CTT of the $K-th$ sample, respectively. μ_{CCT} and σ_{CCT} denote the mean and standard deviation of the total dataset. Then, we developed an RI for the $K-th$ sample as follows:

$$RI_K = \begin{cases} TSI_K & \Delta_K > \delta_{ts} \\ \min(sigmoid(t_K^* + \xi) - 1, TSI_K) & \Delta_K \le \delta_{ts} \end{cases}, \qquad (19)$$

where sigmoid refers to the nonlinear activation function matching the strong nonlinearity and non-autonomy of the power system and

δ_{ts} denotes a discrimination threshold for the TSB shift among the samples of the training set. ξ considers the extreme cases, especially the multi-swing unstable samples. min selects the minimum value of inputs. The RI value is limited within the range of $[-1, 1]$ when proper δ_{ts} approaches 0.

3.6 Loss function

Multi-task downstream networks, including the TSP task and the QRP task, share the output from node-embedding step as input. Evidently, the joint loss function of GraphSAGE-A consists of the loss from TSP and QRP. Additionally, a regularization item should be considered as follows:

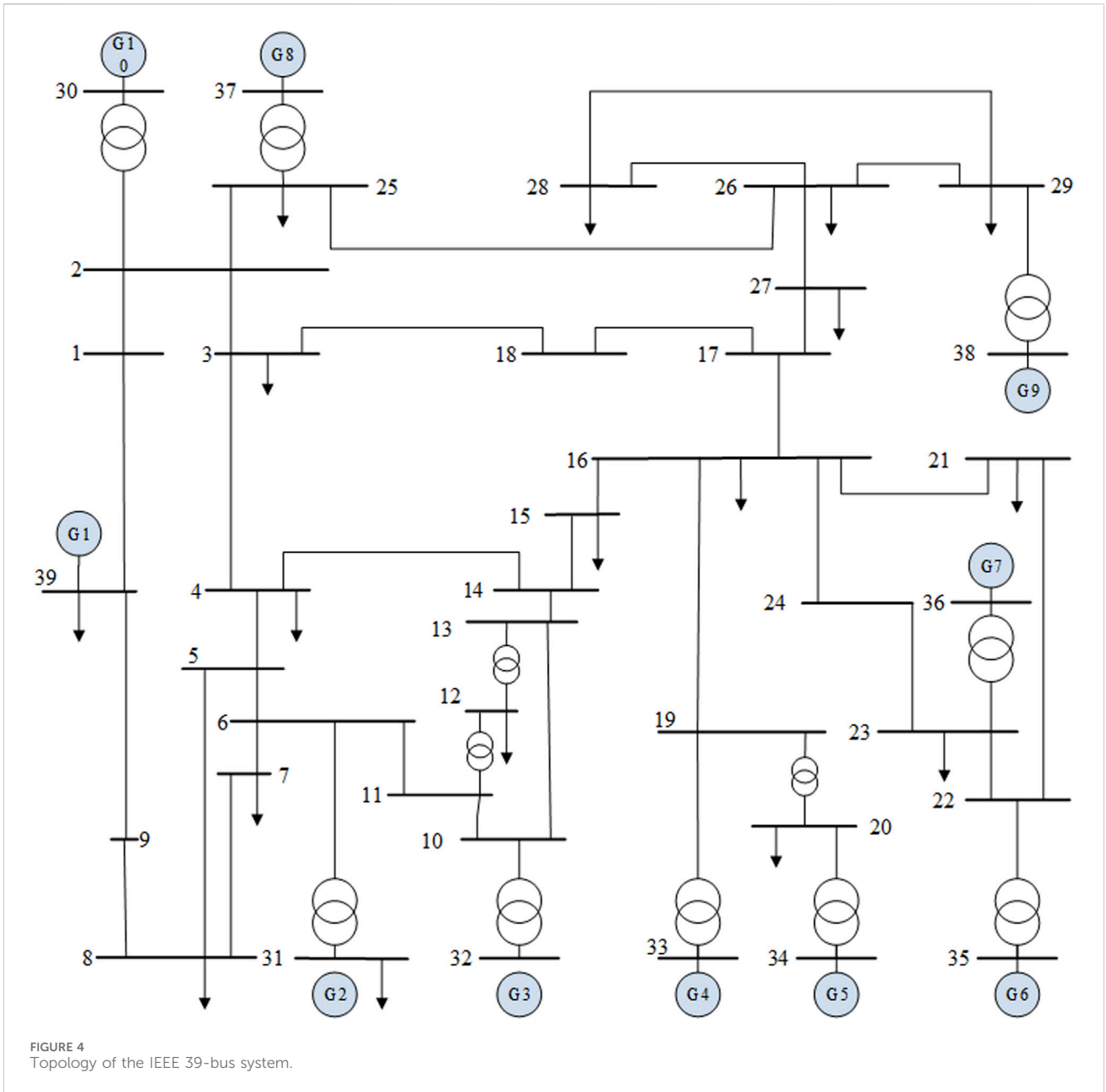$$LOSS = L_{TSP} + L_{QRP} + \beta L_R, \qquad (20)$$

FIGURE 4
Topology of the IEEE 39-bus system.

where the binary classification loss L_{TSP} is defined via the cross-entropy function with batch size B as follows:

$$L_{TSP} = -\frac{1}{B} \sum_b \left[o_1^b \cdot \log\left(\hat{p}_1^b\right) + \left(1 - o_1^b\right) \cdot \log\left(1 - \hat{p}_1^b\right) \right], \quad (21)$$

where (o_0^b, o_1^b) is the one-hot label for the TSP task and $\hat{p}_b = (\hat{p}_0^b, \hat{p}_1^b)$ denotes the stable possibility of the $b-th$ sample. As for the regression loss L_{QRP}, a mean squared error (MSE)-based function is utilized with batch size B as follows:

$$L_{QRP} = -\frac{1}{B} \sum_b \left(\hat{v}_b - RI_b\right)^2, \quad (22)$$

where $\hat{v}_b$ and RI_b stand for the predicted and real RI of the bth sample. We adopt the L2 function as the regularization item with weight β. The Adam optimization is used for training.

3.7 Evaluation metrics

To evaluate the performance of the TSP task, we introduce the accuracy and f1-score metrics given by the following equations:

$$ACC = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \times 100\%, \quad (23)$$

$$F1 = \frac{2T_P}{2T_P + F_P + F_N} \times 100\%, \quad (24)$$

where T_P, T_N, F_P, and F_N, respectively, stand for the number of true positive, false negative, false positive, and false negative results. The MSE can evaluate the performance of the regressor. The formula is

$$MSE = \frac{\left(\hat{v}_b - RI_b\right)^2}{T_P + T_N + F_P + F_N}. \quad (25)$$

4 Experiments

4.1 Test system and simulation settings

The proposed MTTSA method is tested on the IEEE 39-bus system. As is shown in Figure 4, the total system consists of 39 buses, 10 generators, 19 loads, 34 transmission lines, and 12 transformers.

All the samples are collected via TDS on PSD-BPA with sixth-order generator models. The loads are set as a proportional mix of the constant impedance and induction motor model. The settings of the simulation can be depicted as follows:

- Situations based on all transmission lines working are called the "Base" cases, while the "N-1" cases need one transmission line off by avoiding those cases with islands
- A total of 11 different load levels within 80%–120% are considered, while the generator outputs randomly match the power flow
- Three-phase grounding faults and two-phase grounding faults at the 0% of any transmission lines last for 0.1 s during every simulation lasting for 6 s
- Labels of each contingency for TSP and QRP can be collected and computed

Thus, 11,160 samples are selected with 5,929 stable ones and 5,231 unstable ones. The shuffled dataset is divided into the train set, validation set, and test set with a ratio of 3:1:1, respectively, to ensure the balance of distribution. All the experiments of the MTTSA method are conducted on a 64-bit computer with Intel(R) Core i7-12700 CPU@ 3.61 GHz CPU, 32 GB RAM, and NVIDIA GeForce RTX 3080Ti 12G GPU.

4.2 Tests on the multi-task training method

Figure 5 shows the effectiveness of the proposed multi-task training strategy with the same settings of the node-embedding module. The joint-training method clearly leads to a significant increase from three perspectives, including the ACC, F1-score, and MSE. The QRP mask promotes a clear instability risk from certain samples because enriched information was obtained from the multi-task training method.

4.3 Hyperparameter testing in the GraphSAGE-A

For the whole structure of GraphSAGE-A, the multi-head attention aggregation is developed as the key physical-enhanced part. Regarding the hyperparameters, the number of heads K and dropout rate $dropout$, both on the node features and the attention weights, respectively, are considered as the key parameters to be adjusted. During the tests of different parameter pairs, such as $K = (1, 2, 4, 6, 8)$ and $dropout = (0, 0.1, 0.2, 0.3, 0.4, 0.5)$ (Figure 6), we find that each new setting brings enhanced performance on TSP and QRP, where the ACC and the F1-score increase by approximately 0.39% and 0.002%, respectively, while the MSE drops to 0.0068%. The optimal hyperparameter is selected as $K = 6$ and $dropout = 0.4$. All the hyperparameters of the model are shown in Table 2.

FIGURE 5
Performance of different strategies.
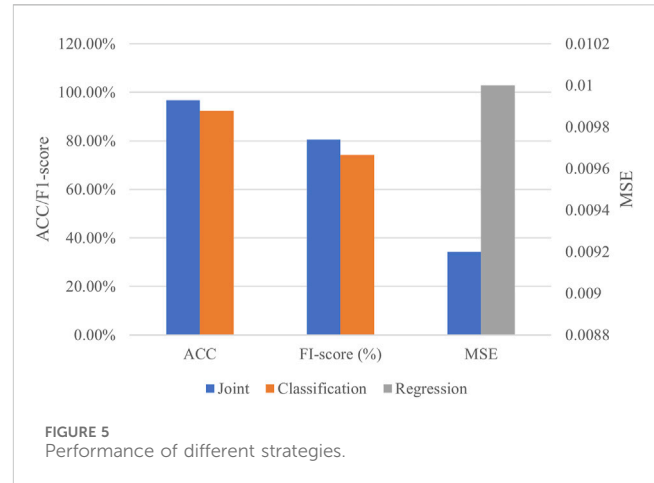
4.4 Verification of the multi-head attention structure

The local normalized edge weight method and the multi-head attention structure allow for more efficient adaptability to topology. In order to verify the improvement, the GraphSAGE-mean and each method detached GraphSAGE-A are compared to the proposed model with the same hyperparameter. The results are shown in Table 3. Essentially, adding both the single and simultaneous one as the developed model has shown significant profits compared to the baseline model.

4.5 Verification of RI

Figure 7 shows the distribution of CCT and RI of the uncertain samples when shifting TSB considered ($\Delta_K \leq \delta_{ts}$) with settings of $\delta_{ts} = 15$, $\xi = 2$ in (16). Apart from the stable samples, the nonlinear function can map samples located on the stable boundary to the appropriate range. The subsequent control strategies can be adopted for the power grid when the RI with suitable parameters is sufficiently small.

4.6 Comparison with baseline methods

We compare five baseline models with the proposed GraphSAGE-A method. Among them, CNN considers the feature-shaped 2D image as inputs, while LSTM adopts the raw inputs. The GNN-based models analyzed the same graph dataset with the three-layer node-embedding structure. Table 4 provides the results.

All the metrics of GraphSAGE-A performed better than the others, where ACC exceeds 96.66% and MSE decreases to 0.0092. The multi-head attention mechanism has been verified as a key factor for performance improvement, owing to its proper combination of global and local information.

Among the baseline methods, graph-deep structures such as GCN, GraphSAGE, and GAT stacked with topological processing show better performance than the traditional deep networks.
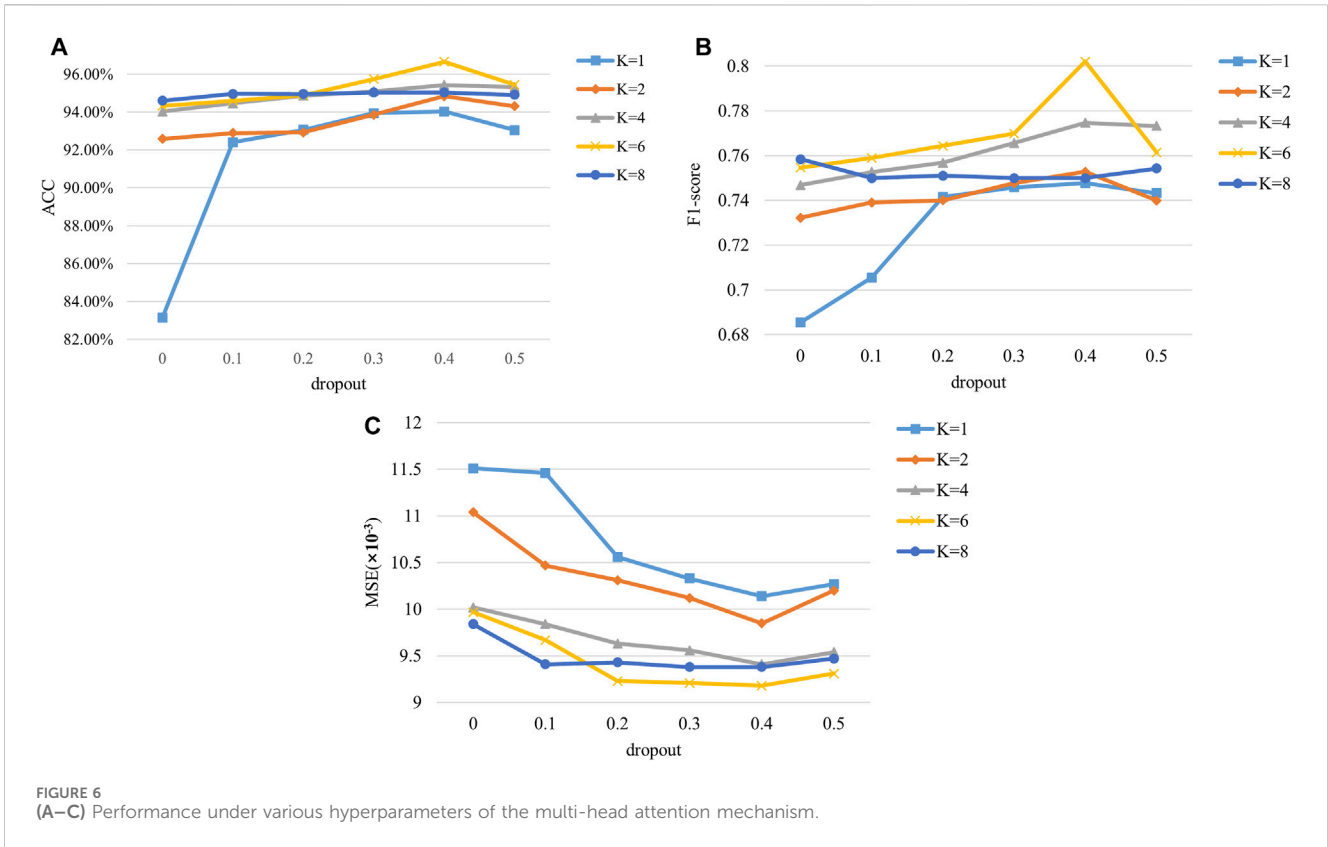
FIGURE 6
(A–C) Performance under various hyperparameters of the multi-head attention mechanism.

TABLE 2 Hyperparameters of GraphSAGE-A.

Batch size	171	Learning rate	0.001
Number of heads K	6	Regularizer weight β	0.0005
Dropout rate	0.4	Node-embedding layer size	(4, 16, 16, and 8)
TSP layer size	(936, 128, 16, and 2)	QRP layer size	(936, 128, 16, and 1)

TABLE 3 Tests on the model structure verification.

Model structure	ACC	F1-score	MSE ($\times 10^{-3}$)
Proposed	96.66	0.8021	9.18
GraphSAGE-A with the global normalized edge weight graph	94.21	0.7324	9.76
GraphSAGE-A with the typical attention mechanism	94.03	0.7477	10.14
GraphSAGE-mean	92.56	0.6988	12.04

Compared to GCN, the approximate structure of GraphSAGE-mean exhibits slight improvements across three metrics. Similarly, GraphSAGE-A also demonstrates slight improvements compared to GAT, which shares the multi-head attention structure.

4.7 Effectiveness analysis of the topological generalization ability

We further use the trained GraphSAGE-A in more samples from even worse conditions of the IEEE 39-bus system to test the generalization ability for the topology. The total 2,000 samples from "N-2" and "N-3" cases are selected with 960 stable and 1,040 unstable ones. The results are shown in Table 5.

Among the baseline models, traditional deep networks such as CNN and LSTM experience significant performance degradation when facing topological changes, where the ACC and F1-score decline by approximately 10% and 0.002% while the MSE drops to 0.017%. Due to the limitations in its inductive learning, GCN slightly decreases when facing topological changes. GraphSAGE, which considers reasoning and learning, GAT with introduced attention mechanism, and GraphSAGE-A, which simultaneously considers both, maintain stable performance. The results indicate that the method proposed in this paper outperforms other models in terms of adaptability to topology.

4.8 Analysis of unbalance distribution

A total of 500 samples from "Base," "N-1," "N-2," and "N-3" cases are selected with 450 stable ones and 50 unstable ones to verify the extreme imbalance. A focal loss function is developed to rewrite the loss of TSP (Lin et al., 2017) as follows:
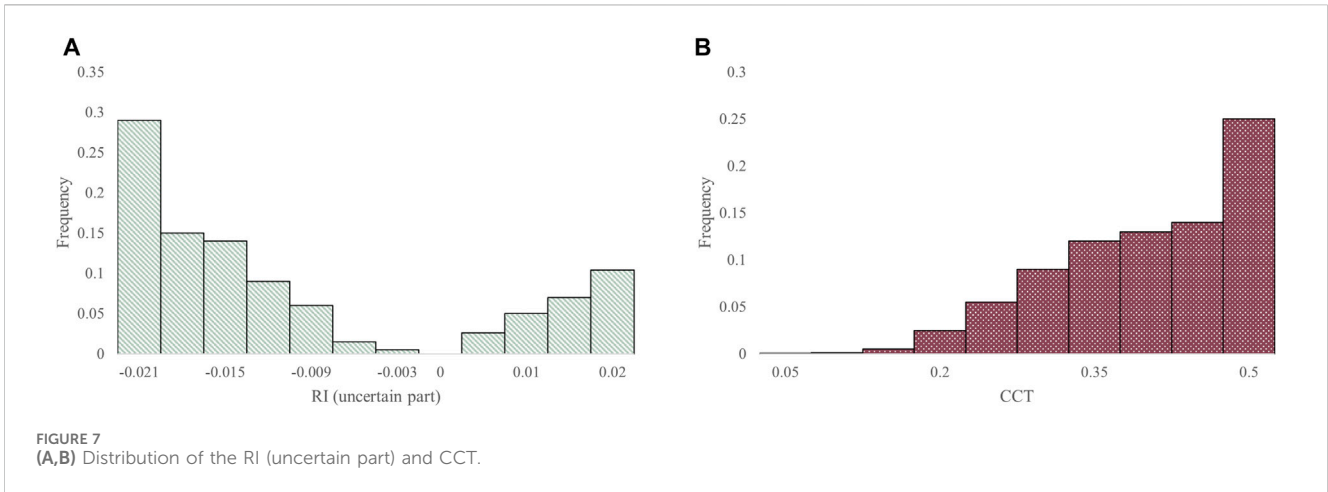
FIGURE 7
(A,B) Distribution of the RI (uncertain part) and CCT.

TABLE 4 Comparison with baseline methods on "Base" and "N-1."

Model	ACC	F1-score	MSE ($\times 10^{-3}$)
CNN	91.96	0.6749	13.44
LSTM	90.84	0.6704	13.5
GCN	92.48	0.6987	12.1
GraphSAGE-mean	92.56	0.6988	12.04
GAT	96.6	0.7986	9.25
Proposed	96.66	0.8021	9.18

TABLE 5 Comparison with baseline methods on "N-2" and "N-3."

Model	ACC	F1-score	MSE ($\times 10^{-3}$)
CNN	82.48	0.5467	16.79
LSTM	80.21	0.5231	17.57
GCN	91.68	0.671	12.46
GraphSAGE-mean	92.56	0.6986	12.05
GAT	96.54	0.7985	9.46
Proposed	96.65	0.8021	9.18

$$L_{TSP} = -\frac{1}{B} \sum_b \left[\alpha \cdot \left(o_1^b\right)^\gamma \cdot \log\left(\hat{p}_1^b\right) + (1-\alpha) \cdot \left(1 - o_1^b\right)^\gamma \cdot \log\left(1 - \hat{p}_1^b\right) \right], \quad (26)$$

where (o_0^b, o_1^b) is the one-hot label for the TSP task and $\hat{p}_b = (\hat{p}_0^b, \hat{p}_1^b)$ denotes the stable possibility of the $b-th$ sample. $\alpha \in (0, 1)$ is a weighting factor for class 1 and $1 - \alpha$ is for class 0. $\gamma \in [0, 2]$ is a value setting.

During the tests of different parameter pairs, such as $\alpha = (0.09, 0.1, 0.12, 0.14, 0.16)$ and $\gamma = (0, 0.5, 1, 1.5, 2)$ (Figure 8), the optimal hyperparameter is selected as $\alpha = 0.12$ and $\gamma = 1.5$. The performance under imbalance is maintained with a proper setting.
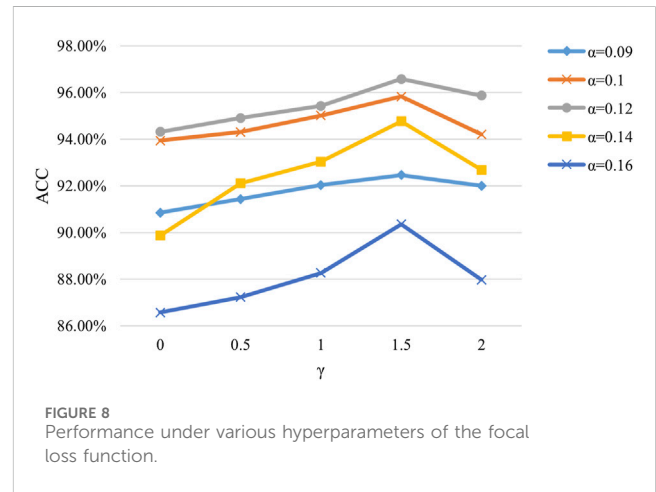

FIGURE 8
Performance under various hyperparameters of the focal loss function.

5 Conclusion

In this paper, an MTTSA method is adopted to detect the accurate TSP and QRP of power systems after faults. The system features from pre-fault, fault occurrence, and post-fault snapshots (t_{0^-}, t_{0^+}, and t_{c^+}) are transformed into multi-graph inputs. In addition to the transient stability binary analysis, the multi-task strategy allows the transient stability risks to be quantified and labeled by the RI, which considers the stability margin variability across various samples. During the joint-training stage, a topology adaptive network GraphSAGE-A is developed to learn the transient features, where a multi-head attention mechanism efficiently aggregates global information from the neighbors via proper weights. The extraction for contextual features is addressed by the local normalized strategy.

Therefore, tests based on the IEEE 39-bus system illustrate the obvious improvement of the proposed MTTSA method compared to baseline models. The adopted RI has been proved to quantify the transient stability risk considering the uncertain samples caused by TSB. Combined with the joint downstream, the risk of the power grid is rapidly calculated to guide the risk control strategies. Through the proper setting of credibility thresholds, the multi-head attention structure and the local normalization for edge weights are verified to

be effective for model performance. Particularly, under severe topology variations, the MTTSA retains a steady precision and topology adaptive ability.

Further works will focus on optimizing the parameters of the model and improving the generalization ability for addressing more complex data.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

LH: conceptualization, methodology, writing–original draft, and writing–review and editing. KD: conceptualization, methodology, supervision, and writing–review and editing. JZ: writing–review and editing. KL: writing–review and editing. CJ: writing–review and editing. XG: writing–review and editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chen, K., Hu, J., Zhang, Y., Yu, Z., and He, J. (2020). Fault location in power distribution systems via deep graph convolutional networks. *IEEE J. Sel. Areas Commun.* 38, 119–131. doi:10.1109/JSAC.2019.2951964

Chen, Q., and Wang, H. (2021). Time-adaptive transient stability assessment based on gated recurrent unit. *Int. J. Electr. Power Energy Syst.* 133, 107156. doi:10.1016/j.ijepes.2021.107156

Diao, R., Jin, S., Howell, F., Huang, Z., Wang, L., Wu, D., et al. (2017). On parallelizing single dynamic simulation using HPC techniques and APIs of commercial software. *IEEE Trans. Power Syst.* 32, 2225–2233. doi:10.1109/TPWRS.2016.2601024

Ding, Y., Zhao, X., Zhang, Z., Cai, W., and Yang, N. (2022). Graph sample and aggregate-attention network for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi:10.1109/LGRS.2021.3062944

Du, Y., and Hu, Z. (2021). Power system transient stability assessment based on snapshot ensemble LSTM network. *Sustainability* 13, 6953. doi:10.3390/su13126953

Glavic, M., Ernst, D., Ruiz-Vega, D., Wehenkel, L., and Pavella, M. (2007). "E-SIME - a method for transient stability closed-loop emergency control: achievements and prospects," in 2007 iREP Symposium - Bulk Power System Dynamics and Control - VII. Revitalizing Operational Reliability, Charleston, SC, USA, 19-24 August 2007 (IEEE), 1–10. doi:10.1109/IREP.2007.4410519

Gomez, F. R., Rajapakse, A. D., Annakkage, U. D., and Fernando, I. T. (2011). Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. *IEEE Trans. Power Syst.* 26, 1474–1483. doi:10.1109/TPWRS.2010.2082575

Gupta, A., Gurrala, G., and Sastry, P. S. (2019). An online power system stability monitoring system using convolutional neural networks. *IEEE Trans. Power Syst.* 34, 864–872. doi:10.1109/TPWRS.2018.2872505

Gurusinghe, D. R., and Rajapakse, A. D. (2016). Post-disturbance transient stability status prediction using synchrophasor measurements. *IEEE Trans. Power Syst.* 31, 3656–3664. doi:10.1109/TPWRS.2015.2496302

Huang, J., Guan, L., Su, Y., Yao, H., Guo, M., and Zhong, Z. (2020). Recurrent graph convolutional network-based multi-task transient stability assessment framework in power system. *IEEE Access* 8, 93283–93296. doi:10.1109/ACCESS.2020.2991263

Huang, J., Guan, L., Su, Y., Yao, H., Guo, M., and Zhong, Z. (2021). A topology adaptive high-speed transient stability assessment scheme based on multi-graph attention network with residual structure. *Int. J. Electr. Power Energy Syst.* 130, 106948. doi:10.1016/j.ijepes.2021.106948

Ishizaki, T., Chakrabortty, A., and Imura, J.-I. (2018). Graph-theoretic analysis of power systems. *Proc. IEEE* 106, 931–952. doi:10.1109/JPROC.2018.2812298

Kamwa, I., Samantaray, S. R., and Joos, G. (2012). On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based

catastrophe predictors. *IEEE Trans. Smart Grid* 3, 152–161. doi:10.1109/TSG.2011.2164948

Kim, C., Kim, K., Balaprakash, P., and Anitescu, M. (2019). "Graph convolutional neural networks for optimal load shedding under line contingency," in 2019 IEEE Power and Energy Society General Meeting (PESGM), Atlanta, GA, USA, 04-08 August 2019 (IEEE), 1–5. doi:10.1109/PESGM40551.2019.8973468

Li, X., Yang, Z., Guo, P., and Cheng, J. (2021). An intelligent transient stability assessment framework with continual learning ability. *IEEE Trans. Ind. Inf.* 17, 8131–8141. doi:10.1109/tii.2021.3064052

Lin, H., Chen, Z., Chen, J., and Chen, W. (2022). "Transient stability analysis of AC-DC hybrid power grid under topology changes based on deep learning," in 2022 12th International Conference on Power and Energy Systems (ICPES), Guangzhou, China, 23-25 December 2022 (IEEE), 514–519. doi:10.1109/ICPES56491.2022.10073461

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. *Proc. IEEE Int. Conf. Comput. Vis.* 42, 318–327. doi:10.1109/tpami.2018.2858826

Liu, S., Liu, L., Fan, Y., Zhang, L., Huang, Y., Zhang, T., et al. (2020). An integrated scheme for online dynamic security assessment based on partial mutual information and iterated random forest. *IEEE Trans. Smart Grid* 11, 3606–3619. doi:10.1109/TSG.2020.2991335

Liu, X., Min, Y., Chen, L., Zhang, X., and Feng, C. (2021). Data-driven transient stability assessment based on kernel regression and distance metric learning. *J. Mod. Power Syst. Clean. Energy* 9, 27–36. doi:10.35833/MPCE.2019.000581

Mahato, N. K., Dong, J., Song, C., Chen, Z., Wang, N., Ma, H., et al. (2021). "Electric power system transient stability assessment based on Bi-LSTM attention mechanism," in 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), Chongqing, China, 08-11 April 2021 (IEEE), 777–782. doi:10.1109/ACPEE51499.2021.9437089

Maria, G. A., Tang, C., and Kim, J. (1990). Hybrid transient stability analysis (power systems). *IEEE Trans. Power Syst.* 5, 384–393. doi:10.1109/59.54544

Mukherjee, R., and De, A. (2020). Development of an ensemble decision tree-based power system dynamic security state predictor. *IEEE Syst. J.* 14, 3836–3843. doi:10.1109/JSYST.2020.2978504

Rahmatian, M., Chen, Y. C., Palizban, A., Moshref, A., and Dunford, W. G. (2017). Transient stability assessment via decision trees and multivariate adaptive regression splines. *Electr. Power Syst. Res.* 142, 320–328. doi:10.1016/j.epsr.2016.09.030

Senyuk, M., Safaraliev, M., Kamalov, F., and Sulieman, H. (2023). Power system transient stability assessment based on machine learning algorithms and grid topology. *Mathematics* 11, 525. doi:10.3390/math11030525

Shi, Z., Yao, W., Zeng, L., Wen, J., Fang, J., Ai, X., et al. (2020). Convolutional neural network-based power system transient stability assessment and instability mode prediction. *Appl. Energy* 263, 114586. doi:10.1016/j.apenergy.2020.114586

Siddiqui, S. A., Verma, K., Niazi, K. R., and Fozdar, M. (2018). Real-time monitoring of post-fault scenario for determining generator coherency and transient stability through ANN. *IEEE Trans. Ind. Appl.* 54, 685–692. doi:10.1109/TIA.2017.2753176

Xu, Y., Dong, Z. Y., Zhang, R., Xue, Y., and Hill, D. J. (2015). A decomposition-based practical approach to transient stability-constrained unit commitment. *IEEE Trans. Power Syst.* 30, 1455–1464. doi:10.1109/TPWRS.2014.2350476

Yan, R., Geng, G., and Jiang, Q. (2021). Data-driven transient stability boundary generation for online security monitoring. *IEEE Trans. Power Syst.* 36, 3042–3052. doi:10.1109/TPWRS.2020.3042210

Yu, J. J. Q., Hill, D. J., Lam, A. Y. S., Gu, J., and Li, V. O. K. (2018). Intelligent time-adaptive transient stability assessment system. *IEEE Trans. Power Syst.* 33, 1049–1058. doi:10.1109/TPWRS.2017.2707501

Zhan, X., Han, S., Rong, N., Liu, P., and Ao, W. (2022). A Two-Stage transient stability prediction method using convolutional residual memory network and gated recurrent unit. *Int. J. Electr. Power Energy Syst.* 138, 107973. doi:10.1016/j.ijepes.2022.107973

Zheng, L., Hu, W., Hou, K., Xu, X., and Shao, G. (2017). "Real-time transient stability assessment based on deep recurrent neural network," in 2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia), Auckland, New Zealand, 04-07 December 2017 (IEEE), 1–5. doi:10.1109/ISGT-Asia.2017.8378321