Robust Representation Learning for Power System Short-Term Voltage Stability Assessment under Diverse Data Loss Conditions

Lipeng Zhu, *Member, IEEE,* Weijia Wen, Yinpeng Qu, Feifan Shen, Jiayong Li, *Member, IEEE,* Yue Song, *Member, IEEE,* and Tao Liu, *Member, IEEE*

***Abstract*—With the help of neural network-based representation learning, significant progress has been recently made in data-driven online dynamic stability assessment (DSA) of complex electric power systems. However, without sufficient attention to diverse data loss conditions in practice, the existing data-driven DSA solutions' performance could be largely degraded due to practical defective input data. To address this problem, this work develops a robust representation learning approach to enhance DSA performance against multiple input data loss conditions in practice. Specifically, focusing on the short-term voltage stability (SVS) issue, an ensemble representation learning scheme (ERLS) is carefully designed to achieve data loss-tolerant online SVS assessment: 1) based on an efficient data masking technique, various missing data conditions are handled and augmented in a unified manner for lossy learning dataset preparation; 2) the emerging spatial-temporal graph convolutional network (STGCN) is leveraged to derive multiple diversified base learners with strong capability in SVS feature learning and representation; 3) with massive SVS scenarios deeply grouped into a number of clusters, these STGCN-enabled base learners are distinctly assembled for each cluster via multi-linear regression to realize ensemble SVS assessment. Such a divide-and-conquer ensemble strategy results in highly robust SVS assessment performance when faced with various severe data loss conditions. Numerical tests on the benchmark Nordic test system illustrate the efficacy of the proposed approach.**

***Index Terms*—Deep representation learning, dynamic stability assessment, graph convolution, ensemble learning, missing data, short-term voltage stability.**

This work was supported in part by the National Natural Science Foundation of China under Grant 52207094, 52377095, and 62088101, in part by the Natural Science Foundation of Hunan Province under Grant 2022JJ30007, in part by Hunan Key Laboratory for Internet of Things in Electricity under Grant 2019TP1016, in part by the Science and Technology Innovation Program of Hunan Province under Grant 2023RC3114, in part by the Fundamental Research Funds for the Central Universities under Grant 22120230432, and in part by the Research Grants Council of the Hong Kong Special Administrative Region under the Early Career Scheme through Project No. 27206021. *(Corresponding Authors: Feifan Shen and Tao Liu.)*

Lipeng Zhu, Yinpeng Qu, Feifan Shen, and Jiayong Li are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: zhulpwhu@126.com; quyinpeng@whu.edu.cn; shenfeifan@hnu.edu.cn; j-y.li@connect.polyu.hk).

Weijia Wen is with the State Grid Hunan Information & Telecommunication Company, Changsha 410004, China (e-mail: cook_lex@foxmail.com).

Yue Song is with the Department of Control Science and Engineering, Tongji University, Shanghai 201804, China, also with the National Key Laboratory of Autonomous Intelligent Unmanned Systems, Shanghai 201210, China, and also with the Frontiers Science Center for Intelligent Autonomous Systems, Ministry of Education, Shanghai 200120, China (e-mail: ysong@tongji.edu.cn).

Tao Liu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong 999077, China (e-mail: taoliu@eee.hku.hk).

NOMENCLATURE

ARMA	Autoregressive moving average.
CNN	Convolutional neural network.
CSG	China Southern Power Grid.
CT	Current transformer.
DAE	Deep auto-encoder.
DL	Deep learning.
DSA	Dynamic stability assessment.
DT	Decision tree.
ELM	Extreme learning machine.
ERLS	Ensemble representation learning scheme.
FC	Fully connected.
GAN	Generative adversarial network.
GCN	Graph convolutional network.
IM	Induction motor.
KLD	Kullback-Leibler divergence.
kNN	k nearest neighbor.
LSTM	Long short-term memory.
ML	Machine learning.
MLR	Multi-linear regression.
1-D	One-dimensional.
OTW	Observation time window.
PMU	Phasor measurement unit.
PT	Potential transformer.
RF	Random forest.
RNN	Recurrent neural network.
RVFL	Random vector functional link.
SA	Situation awareness.
SGC	Spatial graph convolution.
STGC	Spatial-temporal graph convolution.
STGCN	Spatial-temporal graph convolutional network.
SVM	Support vector machine.
SVS	Short-term voltage stability.
TD	Time-domain.
TGC	Temporal gated convolution.
TS	Time series.
t-SNE	t-distributed stochastic neighbor embedding.
WAMS	Wide-area measurement system.

I. INTRODUCTION

A. Research Background

AS the most complex man-made nonlinear systems on the earth, electric power systems act as critical infrastructure in human society [1], [2]. Despite the remarkable development

of modern power systems in recent decades, how to maintain stable and secure operations [3], [4] to prevent system collapse and even blackouts for such highly complex systems still remains a significant challenge. In this context, it is imperative to enhance a specific power system's capability in risky situation awareness (SA), so as to take prompt countermeasures to stabilize the system and prevent the occurrence of catastrophic failures. To this end, one of the most crucial tasks is to implement timely dynamic stability assessment (DSA) during system online monitoring. By doing so, proper control actions can be quickly taken to credibly save the system from potential instability.

B. Literature Review

In terms of power system online DSA, many research efforts have been made in the research community. In particular, with the wide deployment of substantial smart sensing and communication devices in today's power grids, e.g., synchronized phasor measurement units (PMUs) recording wide-area system dynamics with a high resolution [5], a new paradigm called data-driven DSA has been successfully developed for power system real-time SA. With the inherent relationships between initial system dynamics and eventual stability status intelligently inferred offline based on data analytics and machine learning (ML) techniques, data-driven DSA can be reliably performed at a fast speed during online application. For instance, as for short-term voltage stability (SVS) assessment, one of the most challenging DSA problems in practical power systems, some promising ML-based data-driven solutions have been recently reported in the literature [6]–[12]. Since SVS is generally driven by various fast-acting dynamic loads like induction motors (IMs) and power electronically controlled devices [13], [14], it often presents complicated stability patterns and characteristics in practical complex power systems. Faced with this issue, the existing studies have leveraged various well-designed ML approaches, e.g., shapelet-assisted decision trees (DTs) [6], [7], random forests (RFs) [8], support vector machines (SVMs) [9], extreme learning machines (ELMs) [15], and random vector functional link (RVFL)-based neural networks [11], [12], to derive data-driven online SVS assessment models.

More recently, the emerging neural network-based deep learning (DL) methods [16]–[21] with overwhelming advantages in representation learning have been introduced to data-driven SVS assessment. By incorporating well-characterized networked spatial correlations into the recurrent neural network (RNN) algorithm that specializes in temporal representation learning, a spatial-temporal feature learning approach to online SVS assessment is developed in [16]. Similarly, with the combination of graph convolutional network (GCN) and RNN, an SVS assessment scheme comprehensively learning spatial and temporal features from SVS dynamics is proposed in [17]. In [18], a more efficient DL network called spatial-temporal GCN (STGCN) that replaces the RNN module with the parallel-enabled one-dimensional (1-D) convolutional neural network (CNN) is employed to perform online SVS assessment. With the introduction of the attention mechanism, a gated recurrent graph attention network is utilized in [19] to perform spatial-temporal correlation learning, which results in adaptive SVS assessment against topological changes. In [20], generative adversarial network (GAN)-based data augmentation is carried out to realize data-driven SVS assessment with a small SVS dataset. Considering possible cyber attacks to data-driven solutions, a comprehensive analysis, verification, and mitigation strategy is presented in [21] to address SVS assessment-oriented adversarial samples. This strategy can help enhance the robustness of existing ML-based SVS assessment solutions to potentially insecure cyber environments. Compared to conventional ML-based schemes with shallow learning structures [6]–[12], these DL-enabled alternatives [16]–[21] with much stronger capability in automatic feature learning and representation have great potential to further improve the reliability and efficiency of online SVS assessment.

Despite the progress described above, there exists a non-trivial research gap in the existing studies on data-driven SVS assessment. Concretely, the majority of them [6], [8]–[10], [12], [17], [18] simply assume that the PMU data constituting SVS assessment models' inputs are fully available in online monitoring contexts, without consideration of potential PMU data loss conditions. In fact, as device malfunctions and failures are always inevitable in PMU-enabled wide-area measurement systems (WAMS), PMU data losses are widely witnessed in practical power grids. Although a few studies [7], [16] have discussed the impacts of PMU data losses on the performance of online SVS assessment in numerical tests, they do not propose effective solutions to enhance their tolerance to such defective conditions. In [11], an active missing data-tolerant SVS assessment scheme is proposed by maximizing power network observability. Yet this scheme is mainly designed to address *dimensional data losses*, i.e., consecutive PMU measurements of certain channels are totally missing. When faced with other conditions, e.g., *temporary data losses* where only partial data points in certain PMU channels are lost, it cannot sufficiently exploit the valuable information hidden behind the remaining normal data points. This could affect the reliability of online SVS assessment.

In fact, in the broader field of data-driven DSA, some similar attempts [22], [23] have also been made to tackle the missing data issue. However, with the same focus on network observability, they cannot efficiently handle temporary data loss conditions as well. Different from these efforts, a data imputation-assisted DSA scheme that can uniformly correct pre-fault missing data with a GAN is introduced in [24]. Nonetheless, as the GAN merely produces single snapshots of system states before fault occurrence, it is difficult for this scheme to address system transient data with much more complicated spatial-temporal dynamics. Actually, similar to GAN, some classical data imputation techniques, e.g., the autoregressive moving average (ARMA) algorithm, the k nearest neighbor (kNN) method, and collaborative filtering [25], may also be used to correct missing values, based upon which the existing data-driven solutions [6], [8]–[10], [12], [17], [18] can compatibly work for online DSA. However, when faced with severe conditions involving high proportions of data losses

in practice, it is difficult for these imputation techniques to reliably fix missing data. Consequently, the wrongly imputed values could misguide subsequent representation learning, thereby degrading online DSA performance. Overall, how to effectively cope with diverse severe PMU data loss conditions to achieve reliable online DSA remains a challenging task in the research community.

C. Motivation and Contribution

Given the above research gap, this paper develops a robust representation learning approach for online SVS assessment under various complicated PMU data loss conditions. Unlike the existing studies relying on network observability optimization or data imputation [11], [22]–[24], this approach designs an SVS assessment-oriented ensemble representation learning scheme (ERLS) by robustly implementing representation learning with lossy PMU data. First, an efficient data masking technique is introduced to handle various missing data conditions and augment the diversity of SVS cases. Taking the promising STGCN as the basic representation learning tool [18], [26], multiple diversified base learners are trained for ensemble decision-making. Then, an array of ensemble SVS assessment models are built in a divide-and-conquer fashion: 1) grouping all the given SVS cases into a number of typical clusters via a deep auto-encoder (DAE); 2) constructing an ensemble SVS assessment model for each cluster by assembling the STGCN-enabled base learners with multi-linear regression (MLR). By doing so, the obtained ensemble models are able to achieve reliable SVS assessment performance under various PMU data loss conditions during online application. The major contributions and merits of this paper are three-fold:

1) This work develops a robust representation learning approach to power system online SVS assessment that can extensively tackle multiple PMU data loss conditions in a unified manner. With no limitation on missing data types, it is more applicable than existing alternatives in practical contexts with various complicated data loss scenarios.
2) By systematically integrating some advanced DL techniques, the approach is realized with a well-designed ERLS which robustly handles diverse missing data conditions in a divide-and-conquer fashion via ensemble decision-making.
3) Extensive tests reveal that the proposed approach achieves superior online performance in diverse defective PMU measurement conditions. Hopefully, it would help enhance the applicability of the data-driven DSA paradigm in practice.

The remainder of the paper is structured as follows. The missing data problem related to online SVS assessment is described in Section II. Section III details the proposed ERLS for data loss-tolerant online SVS assessment. In Section IV, numerical case studies are carried out on the benchmark Nordic test system for performance verification. Finally, concluding remarks are summarized in Section V.

II. Problem Description

Given a regional receiving-end power grid where m major buses are configured with PMUs for online SVS monitoring,

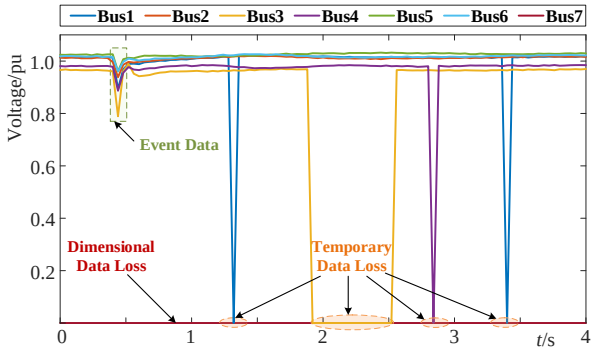


Fig. 1. Illustration of PMU measurements in a real-world power grid.

an observation time window (OTW) with time span T_{win} is employed to acquire consecutive PMU measurements from individual buses. Specifically, assuming the starting point of the OTW is set to the time instant of transient fault occurrence, it collects fault-on and early post-fault system responsive data via PMUs. To sufficiently capture system dynamics, multiplex PMU data, including bus voltage magnitude, injected active power, and injected reactive power, i.e., $\{V, P, Q\}$, are sequentially collected from the m buses. Note that, here P and Q are derived data that are estimated with raw voltage and current phasors initially measured by potential transformers (PTs) and current transformers (CTs) in individual PMUs. All of these data are integrated as a three-channel data matrix $\boldsymbol{X} = \{\boldsymbol{V}, \boldsymbol{P}, \boldsymbol{Q}\}$. Taking the voltage channel for instance, it is described as

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \\ \vdots \\ \boldsymbol{V}_m \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ V_{m1} & V_{m2} & \cdots & V_{mn} \end{bmatrix} \tag{1}$$

where $\boldsymbol{V}_i = [V_{i1}, V_{i2}, ..., V_{in}]$ is the voltage time series (TS) acquired from bus i ($1 \leq i \leq m$), $n = T_{\text{win}}/\Delta T$ is the number of data points within the OTW, and ΔT is the time interval of PMU data acquisition. With $\boldsymbol{X}$ fed into a well-designed DL model for representation learning, critical SVS features can be sufficiently captured to help quickly predict whether the system can maintain SVS after transient fault clearance.

However, due to the inevitability of defective PMU data acquisition and transfer conditions in practice, e.g., PMU/WAMS malfunctions and even failures, practical PMU data matrices have a high risk of undergoing diverse data losses. Taking field voltage measurements acquired from seven substations in a practical power grid for example, the obtained 7×100 voltage measurement matrix (for $T_{\text{win}} = 4$ s and $\Delta T = 40$ ms) with multiple missing data points is depicted in Fig. 1. Note that all the missing data points are filled with zero values by default.

As shown in Fig. 1, all the missing data can be generally grouped into two categories, i.e., dimensional data loss and temporary data loss. The first category implies that all the data points of a certain row (bus) in $\boldsymbol{V}$ are missing, which may result from the breakdown of the corresponding PT or other related devices in the PMU. For the second category, it may

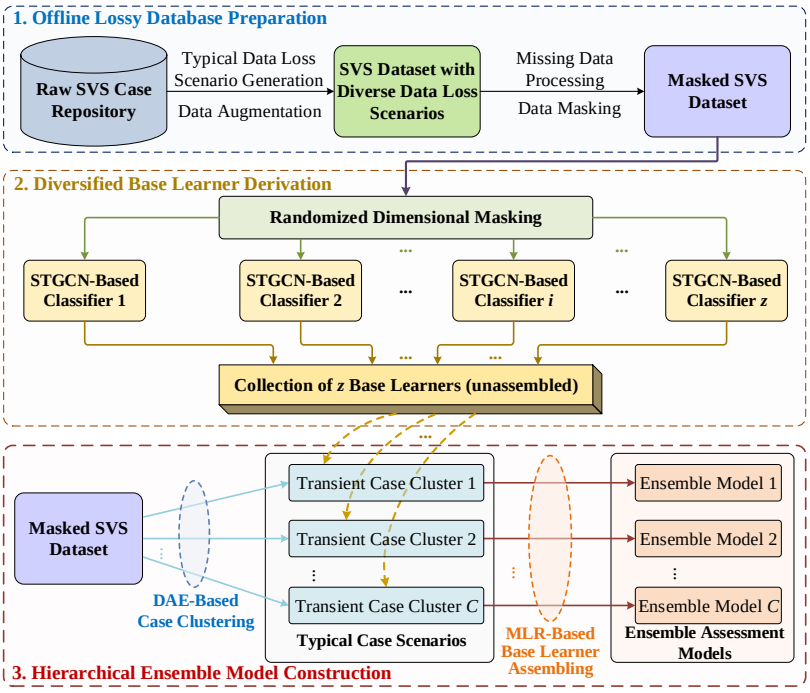


Fig. 2. Overall framework of the proposed ERLS.

be caused by temporary measuring malfunctions during data acquisition or occasional data dropout during data transfer. If not carefully tackled, such undesirable data losses inducing incomplete PMU data matrices could undermine the reliability of follow-up data-driven SVS assessment schemes.

To address the above problem, this paper will develop a data-driven ERLS that can be leveraged to implement reliable online SVS assessment under various PMU data loss conditions in a unified fashion. As will be illustrated below, instead of trying to fix missing data, this scheme derives DL-based SVS assessment models against diverse missing data scenarios by directly performing robust representation learning with lossy PMU measurements.

III. Proposed Learning Scheme

As depicted in Fig. 2, the realization of the proposed ERLS includes three phases: 1) offline lossy database preparation; 2) diversified base learner derivation; 3) hierarchical ensemble model construction.

A. Offline Lossy Database Preparation

For a specific receiving-end system, considering various representative operating conditions and SVS-related transient events, a raw SVS case repository is first produced via batch time-domain (TD) simulations. In each case, by collecting fault-on and early post-fault $\{V, P, Q\}$ TS trajectories from the m buses deployed with PMUs, a three-channel PMU data matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n\times 3}$ is acquired. PMU data losses are simulated by randomly choosing $r_1\%$ of data points from $\boldsymbol{X}$ and setting their values to NaN. In particular, taking the voltage measurements in $\boldsymbol{X}$ (i.e., $\boldsymbol{V}$) for instance, the data matrix $\boldsymbol{V}$ with size $m \times n$ is first reshaped as an $mn \times 1$ vector denoted by $\boldsymbol{v}$. Then, $[mn * r_1\%]$ elements ($[\ ] \rightarrow$ rounding towards $+\infty$) in $\boldsymbol{v}$ are randomly picked out to generate missing data points, where their sequence numbers in $\boldsymbol{v}$ follow a specific random distribution. To boost the diversity of data loss conditions, three types of commonly-used random distributions are considered here to help generate missing data points, including uniform, normal, and exponential distributions. These three distributions are also applied in a random manner, where each of them is chosen with a probability of 1/3. With the $[mn * r_1\%]$ elements chosen from $\boldsymbol{v}$, their identical data points in $\boldsymbol{V}$ are set to NaN, which results in missing values. Similarly, missing values are set for the other channels of PMU data matrices. For each SVS case, the above procedure of randomly generating missing data points in the form of NaN values is executed multiple times with repeated sampling, so that multiple missing data positions in the data matrices following different random distributions are generated. With missing data generation iteratively performed for all the SVS cases in the case repository, numerous lossy data matrices are produced to cover diverse missing data scenarios. As unstable cases are generally scarce compared to stable ones in practice [27], more data loss scenarios can be generated for unstable cases to mitigate the imbalance degree of class distribution in the repository.

Based on the above procedure of random missing data generation, an augmented SVS dataset involving a wide variety of data loss scenarios is obtained. To make such lossy data be compatibly handled by subsequent DL procedures, all the missing data points marked by NaN are filled with zero values. Meanwhile, a binary mask matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ is formed to

explicitly mark all the missing data points in $\boldsymbol{X}$ [28]:

$$\boldsymbol{M} = \begin{bmatrix} 1 & 0 & \cdots & 1 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdots & 1 \end{bmatrix} \tag{2}$$

where 0 denotes the occurrence of missing values for the corresponding entries in $\boldsymbol{X}$, while 1 represents normal values.

As demonstrated in [28], such a binary mask matrix characterizing the positions of missing values in the data matrix can be taken as an informative channel to guide standard representation learning approaches to pay special attention to missing data points. Without limitations on missing data types or specific assumptions on the characteristics of missing data, this additional masking channel can help achieve superior performance over conventional data imputation-based solutions (as will be shown in subsequent case studies). Considering this potential, here $\boldsymbol{M}$ is used to augment $\boldsymbol{X}$, which results in a masked four-channel data matrix $\boldsymbol{X}' = \{\boldsymbol{V}, \boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{M}\}$ ($\boldsymbol{X}' \in \mathbb{R}^{m\times n\times 4}$). With all the masked data matrices gathered together, a masked SVS dataset is formed for SVS assessment model derivation in the sequel.

B. Diversified Base Learner Derivation

In order to enhance the reliability and robustness of online SVS assessment, ensemble learning is carried out in this paper for data-driven decision-making model construction. To this end, a group of diversified base learners are first derived from the above masked SVS dataset for SVS feature learning and representation. Analogous to random subspace-based ensemble learning [29], here a randomized dimensional masking strategy is taken to generate z different learning datasets for base learner derivation. For a certain data matrix $\boldsymbol{X}'$, $r_2\%$ of dimensions (entire rows) in the fourth channel $\boldsymbol{M}$ are masked with 0, and the corresponding positions in the $\boldsymbol{V}/\boldsymbol{P}/\boldsymbol{Q}$ channels are set to NaN values to mimic dimensional data losses; this procedure is repeated z times to produce z different masked versions of $\boldsymbol{X}'$, thereby forming z distinct datasets. With the help of this strategy, not only the diversity between the z datasets is augmented, but also multiple dimensional data loss scenarios are extensively covered.

Further, considering the complicated spatial-temporal correlations between individual load buses during regional SVS dynamics [7], [16], [18], z base learners are separately built from the z datasets via in-depth spatial-temporal representation learning. Considering the graph-like structural couplings between individual buses, the emerging STGCN that fully learns both spatial and temporal features from irregular network dynamics [18], [26] is employed as the primary algorithm in this paper to build SVS assessment-oriented base learners. The architecture of the STGCN is shown in Fig. 3, where a fully convolutional structure is devised to perform SVS feature learning and representation. As can be seen, the STGCN is composed of an input layer (with four-channel structural input data), two spatial-temporal graph convolution (STGC) blocks, a temporal gated convolution (TGC) layer, two fully connected (FC) layers, and an output layer. Among the hidden layers, the two STGC blocks are the major feature learning layers devoted to comprehensive spatial-temporal representation learning. The TGC layer is utilized for further temporal feature learning, while the subsequent two FC layers are leveraged to abstract the learned features and map them to the eventual output. For each of the STGC blocks, it consists of two TGC layers and a spatial graph convolution (SGC) layer, with the three layers alternatively connected to form the whole block (see Fig. 3). The key learning mechanisms of the SGC layers, the TGC layers, and the other layers are detailed below.

1) SGC Layers: The SGC layers are placed in the middle of the STGC blocks to implement networked spatial representation learning. Given the m-bus receiving-end system for SVS monitoring, it is described as a graph with an adjacency matrix $\boldsymbol{W} = [w_{ij}]_{m\times m}$. Here the element w_{ij} in $\boldsymbol{W}$ is computed by

$$w_{ij} = \begin{cases} |Y_{ij}|, \text{ if } i \neq j \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

where $|Y_{ij}| = \sqrt{R_{ij}^2 + X_{ij}^2}$ denotes the magnitude of the mutual admittance of the transmission line that links bus i and bus j, with R_{ij} and X_{ij} representing the resistance and reactance of the transmission line, respectively. On the basis of $\boldsymbol{W}$, the graph Laplacian of the system is normalized as

$$\boldsymbol{L} = \boldsymbol{I}_m - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2} \tag{4}$$

where $\boldsymbol{I}_m \in \mathbb{R}^{m\times m}$ represents the identity matrix; $\boldsymbol{D} \in \mathbb{R}^{m\times m}$ stands for a diagonal matrix, for $D_{ii} = \sum_j w_{ij}$. Let the snapshot of the data matrix $\boldsymbol{X}'$ at time frame t ($1 \leq t \leq n$) be denoted as $\boldsymbol{X}'(t)$, for $\boldsymbol{X}'(t) \in \mathbb{R}^{m\times 4}$. Taking the graph Laplacian-based network description into account, $\boldsymbol{X}'(t)$ can be regarded as four-channel structural data acquired from the power network. If $\boldsymbol{X}'(t)$ is taken as the input, the SGC manipulation can be formulated as

$$\boldsymbol{\mathcal{X}}_j(t) = \Theta \star \boldsymbol{X}'(t) = \Theta(\boldsymbol{L})\boldsymbol{X}'(t) = \boldsymbol{U}\Theta(\boldsymbol{\Lambda})\boldsymbol{U}^T\boldsymbol{X}'(t) \tag{5}$$

where $\boldsymbol{\mathcal{X}}_j(t)$ is the jth-channel ($1 \leq j \leq 4$) graph convolution result; $\Theta \in \mathbb{R}^{K_s\times 4\times 4}$ stands for the graph convolution kernel with size K_s; $\star$ denotes the graph convolution operation; $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ represents the matrix of eigenvectors of the graph Laplacian $\boldsymbol{L}$; $\boldsymbol{\Lambda} \in \mathbb{R}^{m\times m}$ corresponds to the diagonal matrix of eigenvalues of $\boldsymbol{L}$ (for $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$); $\Theta(*)$ denotes the graph convolutional filtering operation (with $*$ concisely representing $\boldsymbol{L}$ and/or $\boldsymbol{\Lambda}$). Due to the need for complicated eigenvalue decomposition and eigenvector matrix multiplication, it is computationally expensive to directly implement SGC via (5). To alleviate the computational burden, the SGC can be accelerated by approximating the filtering operation in (5) with the help of Chebyshev polynomials [26], [30], [31]:

$$\boldsymbol{\mathcal{X}}_j(t) = \Theta \star \boldsymbol{X}'(t) \approx \sum\nolimits_{i=1}^{4} \sum\nolimits_{k=0}^{K_s-1} \theta_{k,i,j} T_k(\tilde{\boldsymbol{L}}) \boldsymbol{X}'(t) \tag{6}$$

where $\theta_{k,i,j} \in \mathbb{R}$ and $T_k(\tilde{\boldsymbol{L}}) \in \mathbb{R}^{m\times m}$ denote the kth-order ($0 \leq k < K_s$) Chebyshev coefficient and polynomial for the approximation of graph convolutional filtering; $\tilde{\boldsymbol{L}} = 2\boldsymbol{L}/\lambda_{\max} - \boldsymbol{I}_m$ represents the normalized graph Laplacian, with $\lambda_{\max}$ denoting the maximum eigenvalue of $\boldsymbol{L}$. Here the

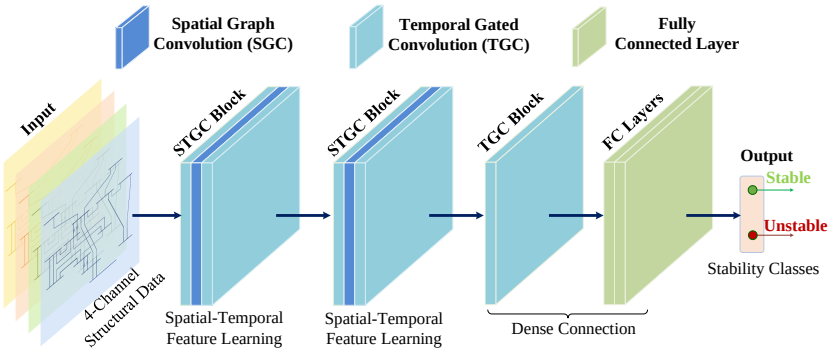


Fig. 3. Architecture of STGCN for base learner derivation.

Chebyshev polynomial $T_k(\tilde{\boldsymbol{L}})$ is calculated in a recursive way:

$$T_k(\tilde{\boldsymbol{L}}) = \begin{cases} 1, & k = 0 \\ \tilde{\boldsymbol{L}}, & k = 1 \\ 2\tilde{\boldsymbol{L}}T_{k-1}(\tilde{\boldsymbol{L}}) - T_{k-2}(\tilde{\boldsymbol{L}}), & 2 \le k < K_s \end{cases} \tag{7}$$

Similar to conventional regular convolution in Euclidean spaces, the SGC in (6) is K_s-localized for information reception. For example, if $K_s = 3$, it receives information from buses with no more than $K_s - 1 = 2$ hops w.r.t. the central bus for convolution. By stacking multiple SGC layers, the overall receptive field of graph convolution can be enlarged to perform deeper feature learning and representation [26]. For each time frame in $\boldsymbol{X}'$, the SGC manipulation is carried out in parallel with the same graph convolution kernel Θ, and the overall calculation across all the n frames yields a four-channel matrix again, which is represented by $\boldsymbol{\mathcal{X}} = \Theta \star \boldsymbol{X}'$, for $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m \times n \times 4}$.

2) SGC Layers: Taking the learning results of the above SGC layer as the intermediate inputs, another module called temporal gated convolution (TGC) is utilized for temporal representation learning. Although the RNN specializing in temporal dependency learning [32] can also be employed to fulfill this task, the complicated gate manipulations and step-by-step serial computations during RNN training would induce heavy computational burdens. Comparatively speaking, the parallel-enabled TGC method [26], [33] can perform much faster temporal representation learning with high reliability. Thus, it is taken as the major algorithm to learn temporal features from $\boldsymbol{\mathcal{X}}$. Let the SGC results at bus i ($1 \le i \le m$) be denoted as $\boldsymbol{\mathcal{X}}_i$ ($\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{n \times 4}$). The TGC is manipulated as

$$\boldsymbol{\Gamma} * \boldsymbol{\mathcal{X}}_i = (\boldsymbol{\Gamma}_1 * \boldsymbol{\mathcal{X}}_i + \boldsymbol{a}) \odot \sigma(\boldsymbol{\Gamma}_2 * \boldsymbol{\mathcal{X}}_i + \boldsymbol{b}) \tag{8}$$

where $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_1 \ \boldsymbol{\Gamma}_2]$ ($\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2 \in \mathbb{R}^{K_t \times 4 \times 4}$) is the regular (Euclidean) convolution kernel with size K_t; $*$ denotes the regular 1-D convolution; $\boldsymbol{a}$ and $\boldsymbol{b}$ represent biases; $\odot$ is element-wise product; $\sigma(\rho) = 1/(1 + e^{-\rho})$ represents the sigmoid activation function; $\boldsymbol{A} \odot \sigma(\boldsymbol{B})$ is the gate mechanism controlling the information propagation of $\boldsymbol{A}$ into subsequent neural layers via the sigmoid gate $\sigma(\boldsymbol{B})$. In fact, this mechanism resembles gate manipulations in conventional RNNs to mitigate the vanishing gradient issue for deep temporal dependency learning, yet with a much higher learning efficiency [33]. With the same kernel $\boldsymbol{\Gamma}$, TGC is efficiently performed for all the m buses in the system, which is compactly denoted by $\boldsymbol{\Gamma} * \boldsymbol{\mathcal{X}}$.

3) Other Layers: As illustrated in Fig. 3, the SGC and TGC layers described above are taken as the basic modules to construct two successive STGC blocks. Then, a single TGC layer and two FC layers are supplemented to relate the learned features to the eventual output. For the two FC layers, the first one maps multiple channels of learned features to one channel, while the second one links the summarized one-channel features to discrete SVS classification. In particular, as the main target of SVS assessment is to predict whether a given SVS case would remain stable or not, the second FC layer is configured with two neurons. For a specific SVS case, supposing the outputs of the two neurons are ν_0 and ν_1, the probability of predicting the case as stable is estimated as

$$\tilde{y} = e^{\nu_1} / (e^{\nu_0} + e^{\nu_1}) \tag{9}$$

Based on the probabilistic value $\tilde{y}$, the eventual stability status of the case is inferred by

$$\hat{y} = \begin{cases} 1 & \text{for } \tilde{y} > \varepsilon \quad \text{(stable)} \\ 0 & \text{otherwise} \quad \text{(unstable)} \end{cases} \tag{10}$$

where ε is the threshold for binary SVS classification. For the purpose of unbiased decision-making, ε is specified as $\varepsilon = 0.5$ in this paper. Supposing the actual stability status of the given case is y (for $y \in \{0, 1\}$, $0 \to unstable, 1 \to stable$), the per-instance learning objective of the overall STGCN is formulated by minimizing the following loss function:

$$\min \mathcal{L}_1 = \min \left[-y \log_2 \tilde{y} - (1 - y) \log_2 (1 - \tilde{y}) \right] \tag{11}$$

Note that the actual stability status of a certain SVS case is determined by the following widely-used criterion: in a specific power grid, if all the major bus voltages recover to an acceptable equilibrium (with voltage magnitudes being no less than 0.9 pu) in the transient time period of 0∼10 s [11], [12], [34], the corresponding SVS case would be taken as stable (i.e., y = 1), otherwise unstable (i.e., y = 0).

As z randomly masked datasets are available, z STGCN-enabled classifiers can be efficiently trained in parallel based on (6)-(11). These classifiers are taken as base learners for subsequent ensemble SVS assessment model construction.

C. Hierarchical Ensemble Model Construction

In this paper, an array of ensemble SVS assessment models are systematically built in a hierarchical manner. First, by

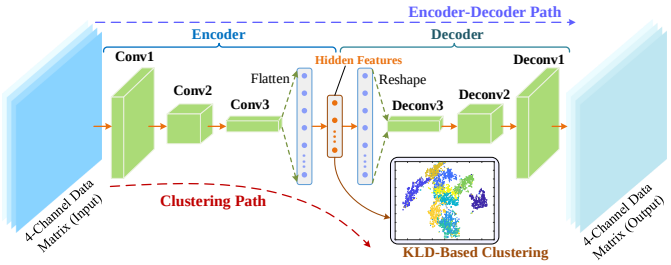


Fig. 4. DAE-based case clustering.

exploring the intrinsic dissimilarities between different SVS cases with a DAE, they are automatically grouped into C typical clusters. Then, a specific ensemble model is constructed for each cluster by assembling the z base learners with the help of MLR. Based on such a divide-and-conquer strategy, distinct SVS cases would be adaptively addressed by different ensemble models, which is expected to enhance the overall robustness and reliability of SVS assessment.

1) DAE-Based Scenario Clustering: To simultaneously perform feature representation and cluster assignment, a powerful convolutional DAE algorithm [35] is introduced in this paper for adaptive SVS scenario clustering. The CNN-based DAE designed for SVS scenario clustering is illustrated in Fig. 4.

As can be seen, it is composed of a CNN-based encoder, a CNN-based decoder, and a Kullback-Leibler divergence (KLD)-based clustering module. Similar to conventional DAEs, the CNN-based encoder and decoder work together to learn in-depth feature representations from the data matrices of substantial SVS cases in an unsupervised way. While the encoder tries to learn a concise feature representation of the input data matrices in a low-dimensional embedded space, the decoder strives to reconstruct the data matrices as precisely as possible. Mathematically, given the data matrix $\boldsymbol{X}'$ of a certain case, the per-instance learning objective along the whole encoder-decoder path is described by minimizing the following reconstruction error:

$$\min \mathcal{L}_r = \min \|\mathcal{H}\left(\mathcal{G}(\boldsymbol{X}', \Phi_G), \Phi_H\right) - \boldsymbol{X}'\|_2^2 \tag{12}$$

where $\mathcal{G}$ and $\mathcal{H}$ stand for the overall mappings of the encoder and decoder; Φ_G and Φ_H denote their respective convolutional neural parameters to be trained; $\|*\|_2$ represents the square norm. Since the encoder and decoder share the same structures and operations with conventional CNN layers [32], their detailed computations are not presented here.

As shown in Fig. 4, for the intermediate hidden layer between the encoder and the decoder, denoted as $\boldsymbol{\varphi} = \mathcal{G}(\boldsymbol{X}', \Phi_G)$, it captures the primary characteristics of the input data matrices. Hence, it is taken as the embedded features for SVS case clustering. By connecting it to the KLD-based clustering module, a clustering path is built to perform cluster assignment. With the emphasis on assigning all the SVS cases to C clusters, the corresponding C cluster centers are randomly initialized as $\{\boldsymbol{\kappa}_j\}_{j=1}^C$ in the embedded feature space specified by $\boldsymbol{\varphi}$. Given N SVS cases for clustering, based on the Student's t-distribution, the probability of assigning a certain SVS case featured by $\boldsymbol{\varphi}$ to the jth cluster is estimated as

$$p_j = \frac{(1 + \|\boldsymbol{\varphi} - \boldsymbol{\kappa}_j\|_2^2)^{-1}}{\sum_{j'=1}^{C}(1 + \|\boldsymbol{\varphi} - \boldsymbol{\kappa}_j'\|_2^2)^{-1}} \tag{13}$$

Taking $\{\boldsymbol{\kappa}_j\}_{j=1}^C$ as trainable neural parameters, the learning objective of cluster assignment is quantified by minimizing the KLD between p_j and a target distribution q_j:

$$\min \mathcal{L}_c = \min \sum_{j=1}^{C} q_j \log_2(q_j/p_j), \tag{14}$$

where q_j is calculated as

$$q_j = (p_i^2/f_j)/(\sum_{j'=1}^{C} p_{j'}^2/f_{j'}). \tag{15}$$

Here f_j refers to the cluster frequency obtained by summarizing all the N cases' probability values p_j w.r.t. cluster j.

Taking feature representation and cluster assignment into comprehensive account, the overall learning objective for clustering all the N SVS cases is summarized as:

$$\min \mathcal{L}_2 = \min \left(\sum\nolimits_i \mathcal{L}_r + \gamma \sum\nolimits_i \mathcal{L}_c\right), \text{ for } 1 \le i \le N \tag{16}$$

where γ is the regularization parameter that regularizes the clustering path for scenario clustering. Following the guideline in [35], γ is empirically set to $\gamma = 0.1$ in this paper. After learning convergence, each SVS case is assigned to a cluster that wins the largest probability value p_j. Based on (16), the whole DAE is able to adaptively group all the SVS scenarios into C highly representative clusters.

2) Ensemble Model Construction: For each of the C clusters obtained above, the SVS cases in it are expected to share similar data characteristics with each other. Thus, they can be collectively handled by a specific decision-making model. For cluster j ($1 \le j \le C$), suppose N_j SVS cases are clustered into it (for $\sum_{j=1}^{C} N_j = N$). Given the z STGCN-enabled base learners derived in Section III-B, an ensemble SVS assessment model is constructed for cluster j by performing MLR:

$$\boldsymbol{Y} = \tilde{\boldsymbol{Y}}\boldsymbol{\alpha} + \boldsymbol{\xi} \tag{17}$$

$$\text{for } \boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N_j} \end{bmatrix}, \ \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_z \end{bmatrix}, \ \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{N_j} \end{bmatrix},$$

$$\tilde{\boldsymbol{Y}} = \begin{bmatrix} 1 & \tilde{y}_{11} & \tilde{y}_{12} & \cdots & \tilde{y}_{1,z} \\ 1 & \tilde{y}_{21} & \tilde{y}_{22} & \cdots & \tilde{y}_{2,z} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \tilde{y}_{N_j,1} & \tilde{y}_{N_j,2} & \cdots & \tilde{y}_{N_j,z} \end{bmatrix}, \tag{18}$$

where y_i is the actual class lable of case i ($1 \le i \le N_j$); $\tilde{y}_{ij}$ is the probabilistic output of predicting case i as the stable class by base learner j [see (9) for detailed computation]; $\boldsymbol{\alpha}$ is the vectorized MLR weight coefficients that assemble the z base learners for weighted prediction; $\boldsymbol{\xi}$ represents the MLR errors. In essence, the MLR model is determined by the $(z+1)$ weight coefficients in $\boldsymbol{\alpha}$. Based on the classical least square method, the MLR problem in (17) can be efficiently solved without iteration to estimate the optimal $\boldsymbol{\alpha}$.

Following the above procedure, C MLR models are con-

structed for the C SVS clusters. Each MLR model outputs the weighted softmax decision values for the cases belonging to the corresponding cluster. The binary class label of each case can be eventually determined by the threshold-based rule in (10). Hence, these MLR models adaptively coordinating the z base learners for decision-making can be taken as ensemble models, and they are readily available for online SVS assessment.

D. Online Application for SVS Assessment

During online SVS monitoring, when the system encounters a transient fault or event, the given OTW with length T_{win} is activated immediately to acquire sequential PMU measurements of $\{V, P, Q\}$ TS trajectories from the m buses in the system. Based on missing data processing and data masking, a measured four-channel data matrix $\boldsymbol{X}^{\text{mea}}$ is obtained from the raw PMU measurements. By feeding $\boldsymbol{X}^{\text{mea}}$ into the z well-trained base learners, z probabilistic decision values for stability classification are derived. Meanwhile, $\boldsymbol{X}^{\text{mea}}$ is taken as the input of the DAE to adaptively infer which SVS cluster the current case belongs to. Assuming it is assigned to cluster j ($1 \leq j \leq C$), the jth MLR model constructed above is employed to estimate the weighted softmax decision value, which is further sent to the threshold-based rule (10) for binary SVS assessment. If the system is inferred to be unstable, alerting signals would be triggered immediately to warn that corrective countermeasures be taken as quickly as possible. Otherwise, the system is deemed to be stable, and continuous SVS monitoring can be performed by sliding the OTW.

To examine the overall SVS assessment performance of the ERLS, three statistical indices, including *misdetection rate*, *false-alarm rate* and *accuracy* (denoted by Mis, Fal, and Acc, respectively), are computed from online stability assessment results under multiple PMU data loss conditions. In particular, with N' cases subject to diverse missing data conditions for online SVS assessment, the three indices are calculated as

$$Mis = N_{fs}/N' \times 100\% \tag{19}$$

$$Fal = N_{fu}/N' \times 100\% \tag{20}$$

$$Acc = \left(N' - N_{fs} - N_{fu}\right)/N' \times 100\% \tag{21}$$

where N_{fs} is the number of unstable cases mistaken as stable; N_{fu} is the number of stable cases falsely labeled as unstable.

To sum up, the proposed ERLS is implemented via the following major steps. First of all, by randomly simulating various data loss conditions and masking missing data points, an offline lossy database for SVS classification learning is well prepared. Then, with the help of the STGCN technique, a series of base learners with strong capability in critical SVS feature learning and representation are derived. Further, by designing a CNN-based DAE, different SVS scenarios are grouped into a handful of representative clusters; for each cluster, the base learners derived above are systematically assembled via MLR, so as to carry out ensemble SVS assessment in a divide-and-conquer manner. With such a comprehensive learning procedure, the whole ERLS would be able to achieve robust SVS assessment performance in the presence of diverse complicated data loss conditions during online application.

IV. Case Study

In modern power grids, since the grids rarely experience severe transient events, especially those that can lead to system instability [27], [36], it is infeasible to collect a large number of actual transient cases from practical power grids for data-driven SVS assessment model construction. Taking this issue into account, case studies in this section were carried out with simulation data obtained from the numerical simulation model of a specific power grid. In fact, this is a common practice in many existing studies on data-driven SVS assessment [16]–[21]. To make the simulation model resemble practical contexts as closely as possible, here the Nordic test system modified from the real-world Swedish and Nordic power grid [37] was taken for realistic numerical simulations. In fact, this system is also a benchmark system strongly recommended by a specialized IEEE task force for voltage stability study [38] and widely adopted by existing voltage stability research in the literature [12], [16], [39], [40]. As shown in Fig. 5, the whole system is divided into four areas, where the Central area is the receiving-end region heavily consuming more than 55% of electric power. Hence, this area was taken as the target region for SVS monitoring. In this area, all the 130-kV and 400-kV buses were deployed with PMUs for online SVS monitoring, as highlighted in Fig. 5. The system was numerically modeled and simulated in PSD-BPA, a popular power system TD simulation package widely adopted in China. For the sake of adequately simulating system SVS dynamics, all the loads in the system were modeled in the form of composite loads consisting of equivalent IMs and static ZIP loads. Data preprocessing and computations were jointly conducted in MATLAB 2020 and Python 3.7 (with the TensorFlow back end).

A. Simulation Setting and Data Generation

Let the operating point A of the system [38] be the base operating condition. Starting from this point, substantial representative operation variations and transient contingencies were considered for batch TD simulations, so as to make the obtained SVS cases cover a wide variety of transient scenarios and characteristics. Specifically, the operation variations and contingencies were generated by setting diverse load power variations (80%$\sim$120% of the base level), distinct dynamic load (i.e., IM) proportions at individual load buses (60%, 75%, and 90%), 18 typical three-phase short-circuit faults (with different fault locations), and shifts of fault clearing times (0.1 s and 0.2 s). All of these combinations resulted in a raw SVS case repository with 10800 cases, among which unstable ones account for 13.7%. For each case, an OTW with length T_{win} = 0.3 s and PMU data sampling interval ΔT = 0.01 s was employed to acquire fault-on and early post-fault $\{V, P, Q\}$ TS data from the 13 major buses in the Central area.

To mimic severe missing data conditions, the proportion of missing data points was set to $r_1\%$ = 20% during lossy data generation. For the purpose of mitigating the imbalance

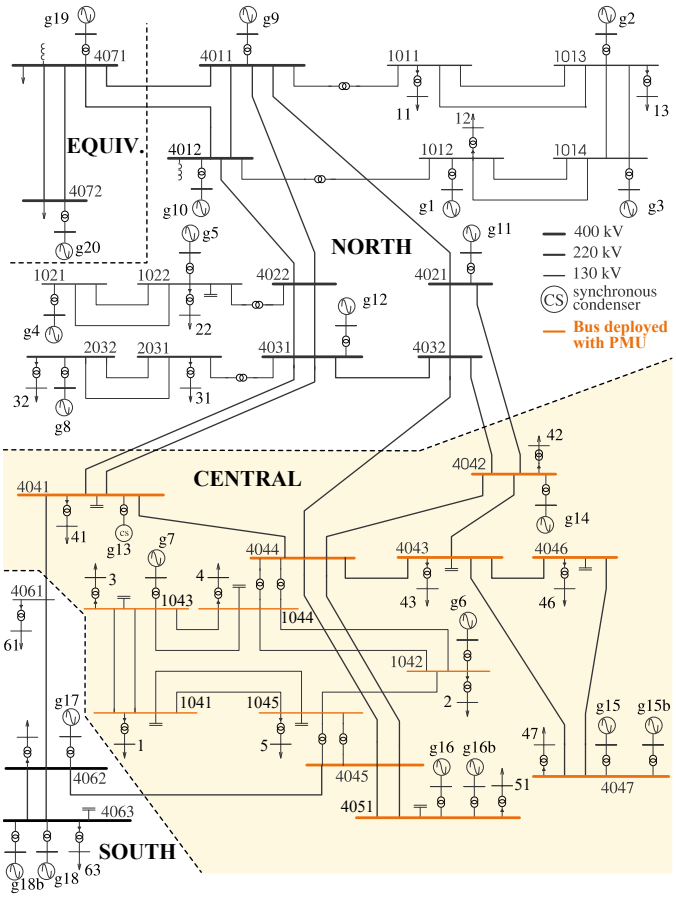


Fig. 5. Single line diagram of the Nordic test system.

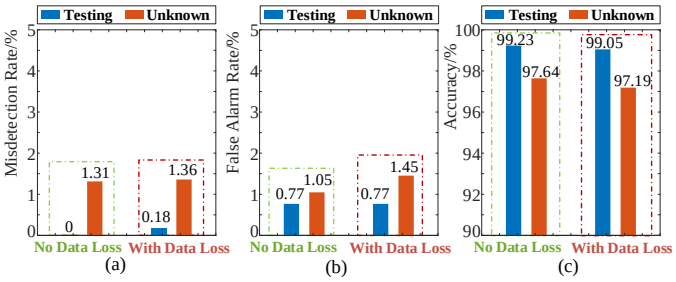


Fig. 6. SVS assessment performance with/without PMU data losses.

degree of stable/unstable cases, the procedure of random missing data point setting was repeated 12 times and twice for unstable and stable cases, respectively. Based on missing data processing and data masking, a masked SVS dataset with 18640 stable cases and 17760 unstable ones was obtained. The whole dataset was divided into two groups via deterministic separation: 1) 30000 cases involving the first 15 transient faults; 2) the remaining 6400 cases related to the latter three transient faults. The first group was further separated into a training set (with 24000 cases) and a testing set (with 6000 cases) via random sampling with no replacement, while the second group was used to test the ERLS's online performance. As this group of cases with distinct SVS events are totally unfamiliar to offline learning procedures, it can be regarded as an unknown set for online generalization tests. With the 24000 training cases, the ERLS was implemented for base learner training and ensemble SVS assessment model construction. The main parameters involved in the learning procedures were empirically specified as $r_2\%$ = 3/13 = 23%, z = 20, K_s = 3, K_t = 3 and C = 10. Note that the mutual admittances constituting the adjacency matrix $\boldsymbol{W}$ [see Eq. (3)] were obtained from the transmission line parameters summarized in Tables 2 and 3 of Ref. [37].

B. Comprehensive Verification of the ERLS

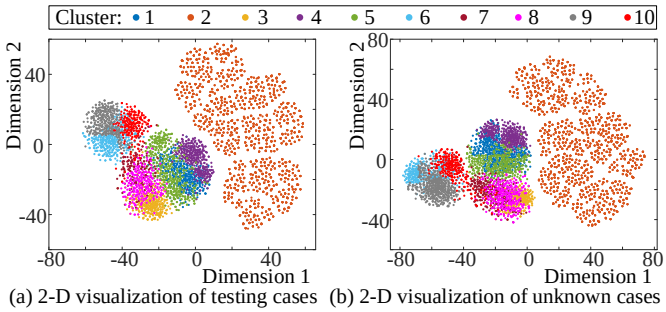
1) Overall Learning Performance: The proposed ERLS's overall learning performance was first tested by feeding all the 6000 testing cases and the 6400 unknown ones into it for classification. To demonstrate its competence in handling multiple missing data conditions during online SVS monitoring, each of the testing/unknown cases was further replaced by its original pure version without any missing data for comparative tests. All the test results are statistically summarized in Fig. 6.

Clearly, the ERLS achieves excellent online SVS assessment performance under both conditions with and without PMU data losses. In particular, with the misdetection and false-alarm rates kept below 1.5%, the overall assessment accuracy firmly remains above 97%. Comparatively speaking, after the introduction of 20% of temporary missing data and 23% of dimensional data losses, the online SVS assessment performance is slightly affected, with the assessment accuracy reduced by less than 0.5%. These results verify that the proposed ERLS has a strong capability in data loss-tolerant online SVS assessment. Further, it is observed that the ERLS can maintain an SVS assessment accuracy of more than 97% when faced with unknown cases totally unseen by offline learning. This implies that it generalizes well to unfamiliar conditions during online SVS monitoring.

2) Effect of SVS Scenario Clustering: To illustrate the effect of DAE-based scenario clustering on assisting SVS assessment, the t-distributed stochastic neighbor embedding (t-SNE) technique [41] was adopted to visualize the clustering results derived from the CNN-based DAE. As depicted in Fig. 7, all the testing/unknown cases were visualized as individual data points in an embedded 2-D space via the t-SNE technique.

It is found that the clustering results marked with different colors are largely consistent with the natural case distribution, with most intra-cluster cases staying together and inter-cluster cases remaining relatively far away from each other. This reveals that the CNN-based DAE successfully captures the primary data characteristics of different SVS cases, thereby resulting in highly representative SVS clusters. As will be verified below, such a clustering strategy would further help enhance online SVS assessment performance in a divide-and-conquer manner.

3) Comparison with Other Methods: To comprehensively verify the advantage of the proposed ERLS, more extensive tests were performed by comparing it with some representative and/or conventional data-driven solutions: a) a single STGCN



(a) 2-D visualization of testing cases (b) 2-D visualization of unknown cases

Fig. 7. t-SNE based visualization of DAE-based case clustering.

TABLE I
SVS ASSESSMENT PERFORMANCE ON 6000 TESTING CASES

Method	$Mis/\%$	$Fal/\%$	$Acc/\%$
Proposed	0.18	0.77	99.05
Single STGCN	1.15	1.27	97.58
STGCN ensemble	0.42	0.85	98.73
CNN + ERLS	1.88	1.43	96.68
LSTM + ERLS	2.53	0.42	97.05
GAN-based scheme [24]	0.77	0.87	98.36
kNN (imputation) + STGCN	1.82	1.08	97.10
ARMA (imputation) + STGCN	2.30	1.15	96.55

without ensemble learning; b) simplified STGCN-based ensemble learning that handles all the SVS cases using one ensemble SVS assessment model (without scenario clustering for hierarchical decision-making); c) a variant of the ERLS by replacing the STGCN with the conventional CNN for base learner derivation; d) a similar variant of the ERLS that takes the long short-term memory (LSTM)-based RNN to train base learners; e) the data-driven solution in [24] that imputes all the missing data with a well-trained GAN and employs RVFLs and ELMs to derive an ensemble assessment model; f) the conventional ARMA algorithm for missing data imputation plus the STGCN for SVS assessment realization; the classical kNN (k = 6) algorithm for missing data imputation plus the STGCN for SVS assessment implementation. All the comparative methods were tested with the same testing and unknown cases, as shown in Tables I and II.

Obviously, the proposed scheme achieves the highest online SVS assessment performance among all the methods. Specifically, compared with the two simplified versions, i.e., the single STGCN and the simple STGCN-based ensemble with no hierarchical decision-making, the ERLS is able to improve the assessment accuracies by 0.3%$\sim$1.5% and 0.6%$\sim$1.1% on the testing and unknown SVS cases. Such a superiority shows that the divide-and-conquer ensemble learning mechanism devised in this paper effectively enhances the ensemble decision-making performance during online application, especially when faced with unknown scenarios.

With further attention to the two variants adopting the CNN and LSTM to train base learners, they have much higher misdetection rates than the proposed scheme (1.5%$\sim$2.4% higher), thus leading to relatively lower SVS assessment accuracies. This implies that the CNN and LSTM are less competent than the STGCN in SVS-related spatial-temporal feature learning and representation. With a more natural learning manner that fully learns spatial-temporal SVS dynamics from the structural network perspective, the STGCN is able to capture the inherent SVS features for more accurate stability classification.

TABLE II
SVS ASSESSMENT PERFORMANCE ON 6400 UNKNOWN CASES

Method	$Mis/\%$	$Fal/\%$	$Acc/\%$
Proposed	1.36	1.45	97.19
Single STGCN	2.06	1.86	96.08
STGCN ensemble	1.58	1.83	96.59
CNN + ERLS	2.87	1.32	95.82
LSTM + ERLS	2.97	0.55	96.48
GAN-based scheme [24]	2.73	2.33	94.94
kNN (imputation) + STGCN	2.98	1.91	95.11
ARMA (imputation) + STGCN	3.19	2.16	94.65

For the GAN-based scheme, although it achieves very high performance on the initial 6000 testing cases, it does not generalize well to the 6400 unknown cases. This is attributed to the fact that the GAN closely following the data distributions of learning cases cannot reliably handle unknown cases with different data distributions (see Fig. 7) for data imputation. Besides, due to the adoption of shallow-structure RVFLs and ELMs for ensemble learning, it is difficult for this scheme to learn the in-depth SVS features from various complicated SVS scenarios. Consequently, it performs even worse than the CNN- and LSTM-based solutions when faced with unknown cases. Analogously, the last two alternatives striving to fix missing data values via the conventional ARMA and kNN algorithms perform worse than the first five solutions that directly learn from missing data assisted by data masking. As can be observed in Table II, these two alternatives have relatively low SVS assessment accuracies (94.6%$\sim$95.1%) in the presence of unknown cases. In fact, it is because the ARMA-based and kNN-based imputation procedures introduce non-negligible errors when a significant number of data points are missing. As a result, those erroneously imputed values would mislead the subsequent representation learning schemes for SVS assessment model derivation.

Considering the above result analyses, the proposed ERLS with superior capability in spatial-temporal SVS feature learning and representation under various missing data conditions would be preferred for online SVS monitoring in practice.

C. Computational Efficiency

For the sake of examining the computational efficiency of the ERLS, its computation time consumptions during both offline learning and online application were recorded. In particular, the 6400 unknown cases were used for online computational efficiency tests. All the computations were conducted on a PC with a 2.60-GHz$*$12 Intel Core i7-10750H CPU and an NVIDIA GeForce GTX-1650 Ti GPU. Multi-core parallel computing was conducted to accelerate offline learning. All the computation times are summarized in Table III.

Overall, the offline computational burden of the ERLS is a bit heavy, with more than four hours consumed by the whole offline learning procedure. Particularly, the procedures

TABLE III
SUMMARY OF OFFLINE AND ONLINE COMPUTATION TIME

Base learner training/s	Case clustering model training/s	MLR-based model assembling/s	Online SVS assessment/s
12091.4352	2941.3869	4.2647	0.0023

Note 1: The first three items represent the time consumptions of individual computation procedures w.r.t. offline ERLS construction; the 4th item stands for the average computation time of online execution of SVS assessment on the 6400 unknown cases, including masking-based missing data processing, DAE-based cluster assignment, and ensemble assessment model application.

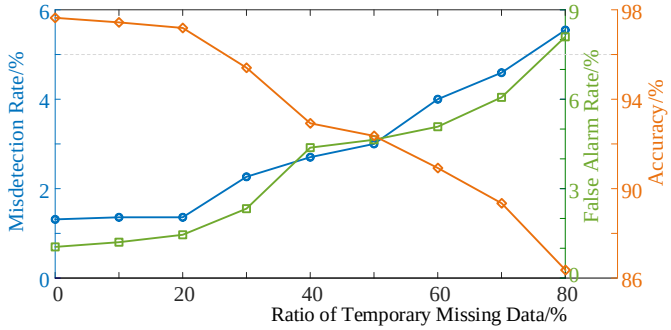


Fig. 8. SVS assessment performance with different ratios of temporary data losses.

of base learner derivation and clustering model training cost most of the offline computation time, while MLR-based SVS assessment model assembling can be quickly completed in less than 5 s. Note that the offline computational burden does not essentially affect the online application efficiency. In fact, as reported in Table III, it takes merely 2.3 ms for the ERLS to implement online assessment on a single SVS case. Given that the OTW begins to acquire sequential PMU measurements immediately after fault occurrence, the proposed ERLS can release online SVS assessment results in less than 0.21 s after fault clearance. Owing to such a high efficiency, pre-emptive control actions can be quickly taken to save the system from potential instability as early as possible.

D. Performance in Diverse Data Loss Conditions

To explore the potential of the proposed ERLS in addressing diverse severe PMU data loss conditions, additional tests were carried out here by varying the ratio of missing data points. Specifically, by separately setting different ratios of temporary missing data, i.e., $r_1\%$ = 0, 10%, 20%, ..., 80%, to the 6400 unknown cases, the ERLS's online performance was evaluated. As summarized in Fig. 8, both the misdetection and false-alarm rates gradually grow with the increase of r_1, which results in the decrease of the overall SVS assessment accuracy. However, it should also be noticed that the ERLS can maintain a high assessment accuracy of more than 96% when $r_1\% <$ 30%, which in fact covers most normal temporary missing data conditions in practical grids [42]. Even if 80% of PMU measurements encounter temporary data losses, the ERLS can still work effectively in most cases, with the assessment accuracy remaining above 86%.

TABLE IV
SVS ASSESSMENT PERFORMANCE UNDER PRACTICAL DATA LOSS CONDITIONS

Scenario	Missing Data Ratio/%	$Mis/\%$	$Fal/\%$	$Acc/\%$
1	12.63	1.09	1.69	97.22
2	15.41	1.06	1.58	97.36
3	17.28	1.12	1.63	97.25

In addition to temporary missing data, severe dimensional data loss conditions were considered for more extensive tests. Concretely, different ratios of dimensional data losses were simulated by assuming $\{0, 1, 2, ..., 10\}$ PMUs underwent total failures, respectively, which corresponds to setting $r_2\%$ to $\{0, 1/13, 2/13, ..., 10/13\}\times 100\%$. With such data loss conditions applied to the 6400 unknown cases, the ERLS's SVS assessment performance is presented in Fig. 9. Analogously, the overall performance generally degrades with the increase in the number of lost PMUs. Yet it is found that such dimensional data losses have a weaker influence on the ERLS's performance than temporary data losses. When no more than five PMUs are subject to total data losses ($r_2\% < 40\%$), the assessment accuracy is steadily kept above 96%. In the extreme condition of losing 10 PMUs ($r_2\% = 77\%$), the ERLS can still maintain an SVS assessment accuracy of nearly 91%.

For the sake of verifying the performance of the proposed ERLS in the presence of more realistic data loss conditions, here additional tests were carried out by mimicking practical data loss scenarios. Specifically, actual lossy PMU measurements were first acquired from three representative transient scenarios occurring in the real-world China Southern Power Grid (CSG) [43]; in each scenario, the data loss information of field PMU measurements was collected from a small 13-bus region of CSG to mimic data loss conditions in the Central area of the Nordic test system. Statistically, missing data points in the three scenarios account for 12.63%, 15.41%, and 17.28%, respectively. Given the information on missing data positions in these three scenarios, missing values were set to the corresponding data points of the data matrices of the 6400 unknown cases to mimic practical data loss conditions. With such newly set unknown cases, the SVS assessment performance of the proposed ERLS was examined again, as presented in Table IV.

Clearly, the proposed ERLS exhibits excellent SVS performance in the above-mentioned practical data loss conditions. The misdetection and false alarm rates firmly remain below 1.15% and 1.70%, respectively, resulting in an overall SVS assessment accuracy of more than 97.2%. Comparatively speaking, the SVS performance presented here is a bit better than that reported in previous tests where data loss conditions were simulated by randomly setting missing data points. In fact, with special attention paid to the three practical data loss scenarios in Table IV, it is observed that most of the missing data points belong to dimensional data losses, with only a small number of missing values occurring in the form of temporary data losses. As verified above, dimensional data losses generally have weaker impacts on the ERLS's

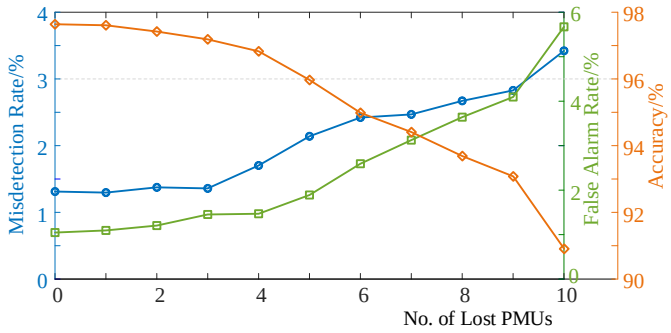


Fig. 9. SVS assessment performance with different amounts of dimensional data losses.

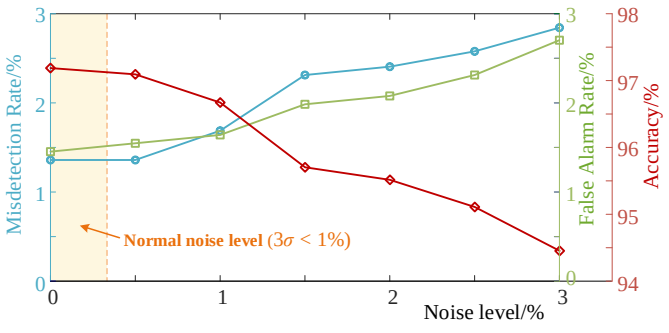


Fig. 10. Online SVS assessment performance in noisy measurement contexts.

performance. On the contrary, the random data loss conditions set up in Section IV-A are more diversified, being more difficult to be handled. This indicates that the previous tests with random data loss conditions are more stringent than that with practical data loss scenarios here. This is the major reason why the proposed ERLS performs slightly better in Table IV.

Considering its desirable tolerance capability w.r.t. both randomly simulated and practical data losses, the proposed ERLS is expected to perform well in various practical contexts with diverse complicated missing data conditions.

E. Robustness to PMU Measurement Errors

As PMU measurement errors are often inevitable in practical measurement environments, here the ERLS's performance was further tested by imposing measurement errors onto PMU data matrices. In particular, random noises following the normal distribution $N(0, \sigma^2)$ were generated to simulate online PMU measurement errors, and they were further imposed onto $\{V, P, Q\}$ measurements in individual PMU data matrices of the 6400 unknown cases. To sufficiently test the ERLS's performance under different levels of measurement errors, the standard deviation σ was set to 0, 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, and 3.0%, respectively. Fig. 10 summarizes the SVS assessment test results with such PMU measurement errors.

Evidently, the ERLS exhibits desirable robustness to PMU measurement errors. In the context of an extremely severe error level of $\sigma_3 = 3\%$, which corresponds to a relatively severe error level of as large as $\pm 9\%$ according to the 3σ rule, the ERLS can still keep the overall SVS assessment accuracy above 94%. In fact, as commercial PMUs deployed in practical grids are required to follow the IEEE Standard with $3\sigma < 1\%$ [44], the practical measurement errors would normally fall into this range. As can be seen in Fig. 10, the ERLS can hold the overall SVS assessment accuracy above 97% in the presence of normal measurement errors with $\sigma < 0.33\%$. Given such a satisfactory anti-error capability, the ERLS can be widely applied in practical defective PMU measurement contexts.

V. Conclusion

Focusing on addressing the challenging issue of implementing reliable online SVS assessment in the presence of diverse PMU data loss conditions, this paper develops a unified data-driven solution by proposing a robust ERLS. With the incorporation of efficient missing data processing techniques and well-designed DL methods for hierarchical ensemble learning, the proposed ERLS gains high potential in handling multiple missing data conditions during online application of SVS assessment. Extensive numerical tests on the Nordic test system show that it achieves highly reliable online SVS assessment in an efficient way despite various severe PMU data loss conditions, with the overall assessment accuracy remaining above 97% in most cases. Compared with existing methods, the ERLS with strong capability in SVS-related spatial-temporal feature learning and representation performs better in online SVS monitoring, especially when faced with new cases unfamiliar to initial offline learning. Also, it exhibits better robustness to various complicated data loss conditions, being more applicable in practice.

This study mainly considers practical measurement conditions of PMU data losses for the implementation of SVS monitoring. How to cope with other defective conditions, e.g., WAMS communication delays and failures, would be investigated in future to derive a more applicable online SVS assessment scheme. Future work would also be devoted to developing an enhanced cost-sensitive SVS assessment approach, where the distinct costs of misdetection and false alarm would be incorporated into the DL procedure.

References

[1] J. Das, *Power Analysis Handbook: Short-Circuits in AC and DC Systems: ANSI/IEEE and IEC Standards*. CRC Press, 2017.

[2] T. Liu, Y. Song, L. Zhu, and D. J. Hill, "Stability and control of power grids," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 5, pp. 689–716, 2022.

[3] L. Xi, J. Wu, Y. Xu, and H. Sun, "Automatic generation control based on multiple neural networks with actor-critic strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2483–2493, 2021.

[4] C. Zheng, S. Wang, Y. Liu, C. Liu, W. Xie, C. Fang, and S. Liu, "A novel equivalent model of active distribution networks based on lstm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2611–2624, 2019.

[5] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, 2010.

[6] L. Zhu, C. Lu, and Y. Sun, "Time series shapelet classification based online short-term voltage stability assessment," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1430–1439, 2016.

[7] L. Zhu *et al.*, "Spatial-temporal feature learning in smart grids: A case study on short-term voltage stability assessment," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1470–1482, 2020.

[8] J. D. Pinzón *et al.*, "Real-time multi-state classification of short-term voltage stability based on multivariate time series machine learning," *Int. J. Electr. Power Energy Syst.*, vol. 108, pp. 402–414, 2019.

[9] H. Yang, W. Zhang, J. Chen, and L. Wang, "Pmu-based voltage stability prediction using least square support vector machine with online learning," *Electr. Power Syst. Res.*, vol. 160, pp. 234–242, 2018.

[10] Y. Zhang, Y. Xu, Z. Dong, and R. Zhang, "A hierarchical self-adaptive data-analytics method for power system short-term voltage stability assessment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 74–84, 2019.

[11] Y. Zhang, Y. Xu, R. Zhang, and Z. Y. Dong, "A missing-data tolerant method for data-driven short-term voltage stability assessment of power systems," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5663–5674, 2019.

[12] C. Ren, Y. Xu, Y. Zhang *et al.*, "A hybrid randomized learning system for temporal-adaptive voltage stability assessment of power systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3672–3684, 2020.

[13] T. Van Cutsem and C. Vournas, *Voltage Stability of Electric Power Systems*. Norwell, MA, USA: Kluwer, 1998.

[14] J. Zhang, Y. Luo, B. Wang, C. Lu, J. Si, and J. Song, "Deep reinforcement learning for load shedding against short-term voltage instability in large power systems," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–12, Early Access, 2021.

[15] Y. Xu, R. Zhang, J. Zhao *et al.*, "Assessing short-term voltage stability of electric power systems by a hierarchical intelligent system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1686–1696, 2016.

[16] L. Zhu *et al.*, "Intelligent short-term voltage stability assessment via spatial attention rectified rnn learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7005–7016, 2021.

[17] G. Wang *et al.*, "A short-term voltage stability online prediction method based on graph convolutional networks and long short-term memory networks," *Int. J. Electr. Power Energy Syst.*, vol. 127, p. 106647, 2021.

[18] Y. Luo *et al.*, "Data-driven short-term voltage stability assessment based on spatial-temporal graph convolutional network," *Int. J. Electr. Power Energy Syst.*, vol. 130, p. 106753, 2021.

[19] H. Cai and D. J. Hill, "A data-driven distributed and easy-to-transfer method for short-term voltage stability assessment," *Int. J. Electr. Power Energy Syst.*, vol. 139, p. 107960, 2022.

[20] Y. Li, M. Zhang, and C. Chen, "A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems," *Applied Energy*, vol. 308, p. 118347, 2022.

[21] C. Ren, X. Du, Y. Xu, Q. Song, Y. Liu, and R. Tan, "Vulnerability analysis, robustness verification, and mitigation strategy for machine learning-based power system stability assessment model under adversarial examples," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1622–1632, 2022.

[22] Y. Zhang, Y. Xu, and Z. Y. Dong, "Robust classification model for pmu-based on-line power system dsa with missing data," *IET Generation, Transmission & Distribution*, vol. 11, no. 18, pp. 4484–4491, 2017.

[23] B. Tan *et al.*, "Spatial-temporal adaptive transient stability assessment for power system under missing data," *Int. J. Electr. Power Energy Syst.*, vol. 123, p. 106237, 2020.

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

[25] X. Yi, Y. Zheng, J. Zhang, and T. Li, "St-mvl: filling missing values in geo-sensory time series data," pp. 2704–2710, 2016.

[26] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2018, pp. 3634–3640.

[27] L. Zhu, C. Lu, Z. Y. Dong, and C. Hong, "Imbalance learning machine-based power system short-term voltage stability assessment," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2533–2543, 2017.

[28] P. Finardi, I. Campiotti, G. Plensack, R. D. de Souza, R. Nogueira, G. Pinheiro, and R. Lotufo, "Electricity theft detection with self-attention," *arXiv preprint arXiv:2002.06219*, 2020.

[29] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall CRC, 2012.

[30] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[31] M. Defferrard *et al.*, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Adv. Neural. Inf. Process. Syst.*, 2016, pp. 3844–3852.

[32] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press Cambridge, 2016.

[33] Y. N. Dauphin *et al.*, "Language modeling with gated convolutional networks," in *Proc. Int. Conf. Machine Learning*, 2017, pp. 933–941.

[34] P. Kundur, J. Paserba, V. Ajjarapu *et al.*, "Definition and classification of power system stability," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1387–1401, 2004.

[35] X. Guo *et al.*, "Deep clustering with convolutional autoencoders," in *Int. Conf. Neural Information Processing*, 2017, pp. 373–382.

[36] R. Ma and S. Eftekharnejad, "Data generation for rare transient events: A generative adversarial network approach," in *Proc. IEEE Industry Applications Society (IAS) Annual Meeting*, 2021, pp. 1–6.

[37] T. Van Cutsem and L. Papangelis, "Description, modeling and simulation results of a test system for voltage stability analysis," Tech. Rep., 2013.

[38] T. Van Cutsem *et al.*, "Test systems for voltage stability studies," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4078–4087, 2020.

[39] M. Lashgari and S. M. Shahrtash, "Fast online decision tree-based scheme for predicting transient and short-term voltage stability status and determining driving force of instability," *Int. J. Electr. Power Energy Syst.*, vol. 137, p. 107738, 2022.

[40] H. Yang, N. Li, Z. Sun, D. Huang, D. Yang, G. Cai, C. Liu, T. Zhang, and W. Zhang, "Real-time adaptive uvls by optimized fuzzy controllers for short-term voltage stability control," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1449–1460, 2022.

[41] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[42] M. Wu *et al.*, "Online detection of low-quality synchrophasor measurements: A data-driven approach," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2817–2827, 2017.

[43] L. Zhu and D. J. Hill, "Cost-effective bad synchrophasor data detection based on unsupervised time-series data analytic," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 2027–2039, 2021.

[44] *IEEE Standard for Synchrophasor Measurements for Power Systems*, IEEE Std. C37.118.1-2011, 2011.