LUT University

# Data-Driven Small-Signal and N-1 Security Assessment Considering Missing Data

Mostafanezhad Majid, Mohammadi Mohammad, Afrasiabi Shahabodin, Afrasiabi Mousa, Aghaei Jamshid, Chung C. Y.

This is a Author's accepted manuscript (AAM) version of a publication

published by IEEE

in IEEE Transactions on Power Systems

**DOI:** 10.1109/TPWRS.2023.3298090

**Copyright of the original publication:**

© 2023 IEEE

**Please cite the publication as follows:**

Mostafanezhad, M., Mohammadi, M., Afrasiabi, S., Afrasiabi, M., Aghaei, J., Chung, C. Y. Data-Driven Small-Signal and N-1 Security Assessment Considering Missing Data. IEEE Transactions on Power Systems. DOI: 10.1109/TPWRS.2023.3298090

© 2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

**This is a parallel published version of an original publication. This version can differ from the original published article.**

# Data-Driven Small-Signal and N-1 Security Assessment Considering Missing Data

Majid Mostafanezhad, Mohammad Mohammadi, *Member, IEEE*, Shahabodin Afrasiabi, *Graduate Student Member, IEEE*, Mousa Afrasiabi, *Member, IEEE*, Jamshid Aghaei, *Senior Member, IEEE* C.Y Chung, *Fellow, IEEE*

***Abstract*— This paper proposes a modulated data-driven method to assess the small-signal and $N-1$ security status in large power systems. To do so, a three-module data-driven framework is designed, including: i) auto-encoder embedded feature selection to reduce the measurement dataset dimension and enhance the computational efficiency; ii) modified generative adversarial networks to improve the robustness against incomplete data and partial observability, which preserves the interpretability of the data measured by phasor measurement units (PMUs) using a reformulated loss function and a new noise generation process; and iii) a convolutional neural network (CNN) as a strong classifier for the assessment of the small-signal and $N-1$ security status. The proposed method is implemented on a 162-bus NESTA benchmark system. The results show the performance of the designed network for different cases and in comparison with several state-of-the-art methods in terms of accuracy, reliability, and computational burden.**

***Index Terms*— Auto-encoder embedded feature selection, convolutional neural network, incomplete data/partial observability, small-signal security assessment, N-1 security assessment.**

## I. Introduction

### A. Outlook

POWER system security assessment is one of the utmost challenging issues in the operation of the power systems due to the demand growth, complexity, advanced communication infrastructure, etc. The emergence of new concepts such as smart grids, microgrids, and distributed generation/storage devices alongside the rapid demand growth has overcomplicated the security assessment problem in today's power systems. The progressive change and high growth rate of electrical energy consumption have led the operation of power systems to follow volatile behaviors, increasing the risks of blackouts. Security assessment is crucial to take the necessary control actions in case any disturbance occurs in the power system. Power system security assessment includes several aspects, i.e., static and dynamic security assessment, small-signal security assessment, transient security assessment, etc. [1]. This paper aims to develop a framework to assess the small-signal and $N-1$ security as a form of static security. To do so, traditional methods such as the application of performance indices and modal analysis cannot be potential solutions for real-time security assessments. With recent developments in the wide-area measurement system (WAMS), particularly the phasor measurement units (PMUs), advanced data-driven methods can be applied. Despite the efforts in recent years, several challenges remain: i) the requirement of a large amount of data inflicts a high computational cost on the process, evaluation, and implementation in practice; ii) the accuracy might be reduced by considering features with little relevance and low importance; and iii) latency in communication, interruption, and malfunction of measurement devices such as PMUs are unavoidable phenomena in power system monitoring. Therefore, a practical solution is required to address these issues. We are motivated to address the abovementioned problems/challenges by designing a multi-module structure including deep feature selection, a data-driven missing data correction module, and a strong deep learning-based classification module.

### B. Literature review

Power system security assessments have been widely investigated in previous studies. The existing methods can be classified as: i) model-based; ii) signal-based; and iii) artificial intelligence (AI) [2]. The model-based models use a mathematical expression of the power system to analyze the security status [3]. The main disadvantages are: i) high dependency on the physical model, as it would be too difficult to provide an accurate model for the power system in most cases; and ii) unsuitability for real-time assessment based on accurate modeling. Signal-based methods analyze the measured signals in the time/frequency domain and assess the security of the power system. However, these methods depend on a predefined threshold and are highly sensitive to the complete information of the power system's status. They might fail

M. Mostafanezhad, M. Mohammadi, are with the School of Electrical and Computer Engineering, Shiraz University,Shiraz, Iran (E-mail: majidmn95@gmail.com, m.mohammadi@shirazu.ac.ir).

S. Afrasiabi is with the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N5A9, Canada. (E-mail: shpower77@yahoo.com).

Mousa Afrasiabi and Jamshid Aghaei are with the Department of Electrical Engineering, School of Energy Systems, Lappeenranta University of Technology (LUT), 53850 Lappeenranta, Finland (E-mail: musa.afra@shirazu.ac.ir, aghaei@sutech.ac.ir)

C. Y. Chung is with the Department of Electrical Engineering, The Hong Kong Polytechnique University, Hong Kong, China (E-mail: c.y.chung@ply.edu.hk)

considering new technologies such as HVDC links and oscillations due to the fluctuation of renewables [4].

AI-based methods only require a set of historical data to assess the power system security status. These methods are categorized into two subcategories: shallow-based methods and deep networks. Shallow-based methods, such as random forest for static security assessment [5], and decision tree for dynamic security assessment [5], multi-class support vector machine (SVM) for both transient and static security assessment [6], core vector machine [7], etc., have been presented in previous studies. These methods are restricted in terms of extracting the comprehensive features in complex and nonlinear systems such as power networks.

A deep neural network is an evolutionary machine learning concept that has demonstrated high capabilities in power system applications, including power system protection [8], load prediction [9], renewable power generation prediction [10], etc. Deep networks have also been employed in power system security assessment.

For instance, in [11], a long short-term memory (LSTM) network is used to learn the oscillatory response of the power system and monitor the critical, poorly damped low-frequency oscillations. In [12], a deep neural network was designed to assess small-signal stability in the presence of renewable energy, which dramatically increases the problem uncertainty. In [13], this problem is addressed by a static security module based on artificial neural networks (ANN). Moreover, recent works comprise more complex models, such as the convolution neural network (CNN), which assess the dynamic and static security at once under the assumption of availability of all measurement data [14]. Furthermore, the presented work has a high computational burden because of the massive data dimensions and the deep model parameter counts. In [15], the initial input space is reduced, and static security is assessed with a CNN model; however, the other challenges, such as partial observability and dynamic security aspects, are neglected. Despite all of these efforts, previous studies suffer from: i) the high computation costs due to the huge amount of measurement data; and ii) failure of the models to accurately assess the security status in cases with missing data.

To reduce the computational cost, a practical solution can be found by feature selection methods. Reducing the dimensions and selecting the features with higher importance can reduce the data required for the power system security assessment. In this regard, relief and partial mutual information are respectively applied in [16], [17] for the power system security assessment. Although these statistical methods are relatively fast, they cannot perform properly in power system analyses owing to their distance-based performance. Embedded feature selection methods such as random forest can extract a feature subset from the data provided by the PMUs but might be unable to avoid overfitting due to the lack of sufficient control parameters.

Auto-encoders (AEs) have been considered as strong deep networks for extracting hidden patterns. However, AEs transfer raw data into the latent space, making the converted data uninterpretable. Reference [18] significantly modified the AE structure by relaxation and parameter regulation techniques to address this issue. This paper develops the modified AE feature selection-based method to select features while preserving the data's interpretability.

To enhance the robustness of the security assessment framework, it is crucial to consider the lack of complete measured data arising from missing PMU data, communication system interruption/latency, etc. To this end, two general solutions can be considered. First, the data-driven structure is designed to be robust against the missing data. In other words, an AI-based method is employed to evaluate the power system security with incomplete data. For instance, a robust ensemble decision tree is presented in [19]. However, robust AI-based structures are significantly affected by different PMU missing data conditions. The second solution group replaces the missing data with features having higher relevance or artificially generated data. In line with this group of missing data compensation methods, generative adversarial networks (GANs) have been utilized [20]. A GAN generates the distribution function of the dataset. Upon the occurrence of missing PMU data phenomena, newly generated data are applied to assess the security of the power systems. The major disadvantage of using a GAN as the missing data correction block is the possibility of data imputation and interpretability, which can reduce the accuracy and reliability of the method. This issue can be addressed by modifying the conventional loss function of the GANs and the noise generation process [21].

### *C. Contributions to this work*

This paper provides a comprehensive framework for power system security assessment. The designed data-driven structure can monitor the small-signal security as well as the $N-1$ security. The designed structure consists of three major modules, i.e., an AE-based feature selection module, an improved GAN network-based module for missing data/partial observability, and a convolutional neural network (CNN) module as the power system status classifier. The AE-based feature selection reduces the amount of required input information and enhances the accuracy and reliability. The interpretability of the raw data is preserved by some modifications on the AEs, including the application of relaxation and parameter regulation techniques. The improved GAN replaces the missing data with the available data. As mentioned above, the data generated by conventional GANs suffer from a lack of interpretability. This issue is addressed in the presented generative adversarial imputation network (GAIN) by modifying the loss function and employing a new noise generation process.

Consequently, the CNN network is developed to assess the small-signal and $N-1$ security status in power networks. The numerical results in the 162-bus network using several state-of-the-art feature selection and missing data correction methods demonstrate the effectiveness and superiority of the designed three-module data-driven framework. The key contributions of this paper can be summarized as follows:

- An improved AE-based feature selection has been designed to reduce the power system security assessment's information redundancy and computational cost. The

designed data-driven feature selection technique can preserve the data interpretability of raw signals measured by PMUs. Thus, the designed AE embedded feature selection structure enhances the computational efficiency, reliability, and accuracy.

- To enhance the robustness of the designed framework considering incomplete/partial observability of the PMUs, the GAN structure is improved by adding the reformulated loss function and a new noise generation process. The improved GAN can preserve the interpretability of data and replace the missing data.
- The designed modular network can simultaneously assess the $N-1$ and small-signal security status. Therefore, a decision-maker can set the power system setpoints with lower costs and time. Through this assessment, the required model has to extract a more complex pattern from a much lower space variable. Thus, the design of a model structure must be more elegant, and determining the hyperparameters of the model becomes one of the key features.

## II. Proposed Methodology for Security Assessment of Power Systems

The proposed method consists of three major modules: i) an AE structure for embedded feature selection, ii) a GAIN-based interrupted/missing data correction, and iii) CNN-based small signal and N-1 security assessment. The proposed approach is illustrated in Fig. 1. In the following, each module and the overall procedure are explained in detail.

### A. Auto-encoder-based Embedded Feature Selection

In modern power systems, the emergence of advanced measurement devices has led to a massive amount of information. Although a large amount of data is essential to evaluate power system security, unnecessary information can reduce the security assessment accuracy and increase the computational complexity. To address this problem, feature selection is a potential solution. Feature selection is a technique to reduce the dimensionality and search space for learning algorithms and decrease the computational burden. This paper utilizes an AE for deep learning-based embedded feature selection. AE-based feature selection is fast and can fully understand the difference between complicated and nonlinear features due to a large parameter hypothesis.

Let $\rho_X$ be a distribution over a $d$-dimensional space. In the AE-based approach, the encoder first maps the input features $X\epsilon\mathbb{R}^{n\times d}$into a $k$-dimensional space $X_s\epsilon\mathbb{R}^{n\times k}$ to feed into the decoder. The decoder then reconstructs the data $\acute{X}\epsilon\mathbb{R}^{n\times d}$in $d$–dimensional space using a reconstruction function $g_\gamma(\bullet)$. The AE-based feature selection is constructed based on:

$$\underset{X_s,\gamma}{\operatorname{argmin}} \left\|g_\gamma(X_s)\text{-}X\right\|_F \tag{1}$$

where $\|.\|_F$ represents the Frobenius norm.

To transfer the input feature maps from $d$–dimensional space to $k$–dimensional space, a stochastic process is considered. The stochastic process in feature dimension reduction enhances the computational efficiency. To do so, a weighting conversion is applied to the input dataset, as:

$$\Phi_i = \frac{\exp\left[\left(\frac{\log\beta_i + g_i}{\lambda}\right)\right]}{\sum_{k=1}^{d}\exp\left[\left(\frac{\log\beta_k + g_k}{\lambda}\right)\right]} \tag{2}$$

where $\beta_i$ , $g_i$ , and $\lambda$ are discrete random variables, the probability distribution function (Gumbel distribution function), and regulation parameter, respectively.

If $\lambda$ consists of small values (close to zero), then $\Phi$ would be a one-hot vector. In this condition, $\Phi$ most likely follows a discrete distribution. Accordingly, the output of the encoder $O_n = X\odot\Phi\ \ n\epsilon\{1\ldots k\}$ is fed to the decoder. The outputs of the AE embedding features selection are obtained from:

$$g_\gamma(Z) = \Theta(W^h * Z + b^h) \tag{3}$$

$$\underset{\beta,\gamma}{\operatorname{argmin}} \left\|g_\gamma(O) - X\right\|_F \tag{4}$$

where $\Theta$,$W$, and $b$ respectively represent the leaky rectified linear unit (ReLU) activation function [20], weight matrices, and bias matrices. Furthermore, the number of hidden layers is represented by $h$, and O results from concatenating all $O_n$.

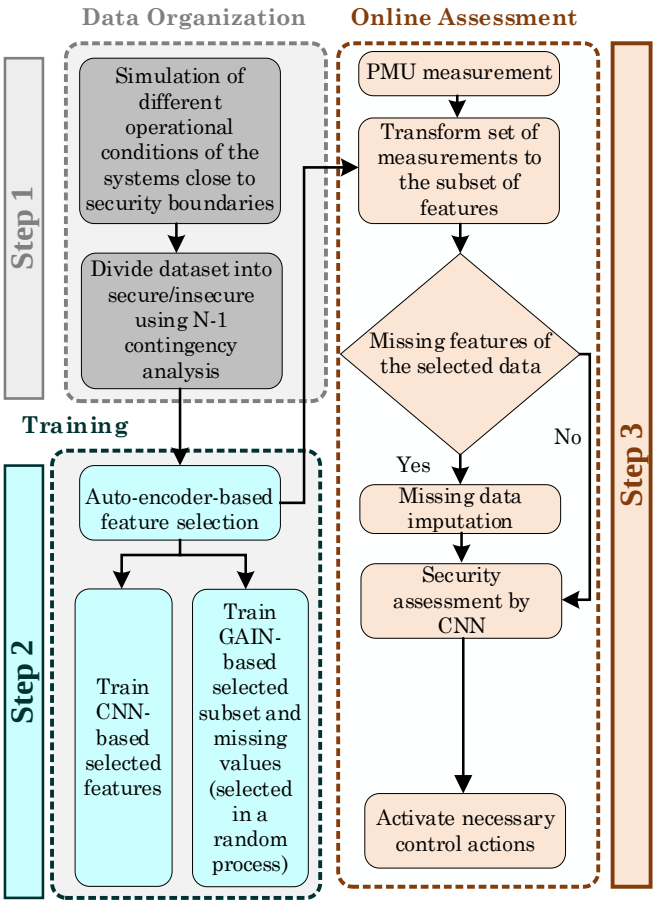



**Fig. 1.** Overall structure of the proposed methodology

### B. GAIN-based Missing Data Correction

Missing data in the wide-area measurement are an unavoidable phenomenon. To enhance the robustness of the designed data-

driven approach, this paper develops a GAIN structure. GAIN is a GAN-based deep network. GAN was originally presented in [23], and it is composed of two major neural networks, i.e., generator and discriminator [24]. In GAN, the generator generates fake data based on the noise and the original data, and then, the discriminator detects the fake data from the original data. The training process continues until the discriminator can no longer detect fake and original data. Hence, the generator generates data that follow the same distribution as the data measured by the PMUs. To improve the processing time, GAIN, as an improved version of GAN [21], is utilized in this paper. In GAIN, only missing/imputed data are used instead of the whole dataset. GAIN is composed of a generator, discriminator, and hint blocks, which are depicted in Fig. 2.

1) **Generator**

The generator produces a dataset to replace the missing data. The input of the generator is the selected (using AE-based feature selection) measured data by the PMUs, $X_k$, associated with distribution function $\rho_S(X_k)$, noise $Z$ with the distribution $\rho_N(Z)$, random binary vector $M$ over the $k$ –dimensional space, with the missing data $\tilde{X}_k$, and $\bar{Z}$ being the noise associated with the missing data defined as:

$$\bar{Z} = (1 - M) \odot Z \tag{5}$$

In the generating process, the dataset, $\tilde{X}_k$ is generated to replace with the missing data. Thus, the new space consists of two sub-spaces, $\tilde{X}_k$ and $X_k$. The new space is defined as:

$$\hat{X}_k = \begin{Bmatrix} X_k & M_k = 1 \\ \tilde{X}_k & M_k = 0 \end{Bmatrix} \tag{6}$$

2) **Discriminator**

The discriminator (shown by $D$) discriminates the imputed missing data from the generator and the actual measurement of the PMUs. The data generated by the generator are used to define $\overline{X_k}$, as:

$$\overline{X_k} = M \odot X_k + (1 - M) \odot G\left(\hat{X}_k, M, \bar{Z}\right) \tag{7}$$

$\overline{X_k}$ and $H$ are considered as the inputs of the discriminator, where $H$ is the hint matrix (described in the next subsection). Similar to the generator, discriminator $D$ is a multi-layer perceptron.

3) **Hint**

The generator can generate data with different distribution functions. To enhance the reliability of the generated data for replacing missing PMU data, the generated data should follow a unique distribution function. To this end, an additional random variable has been added to the conventional GAN, namely the hint, defined to provide enough information about $M$. By providing the required information about $M$, the missing PMU data can be replaced by data with a unique distribution function.

The hint is a random variable upon the defined space, and its value depends on $M$. In the GAIN, the generator outputs and hints are fed into the discriminator. To do so, the discriminator's output probability depends on the hint, and the hint regulates the output probability of the discriminator (generation of the discriminator follows a unique distribution function). The hint, $H$ is computed as:

$$H = Y \odot M \tag{8}$$

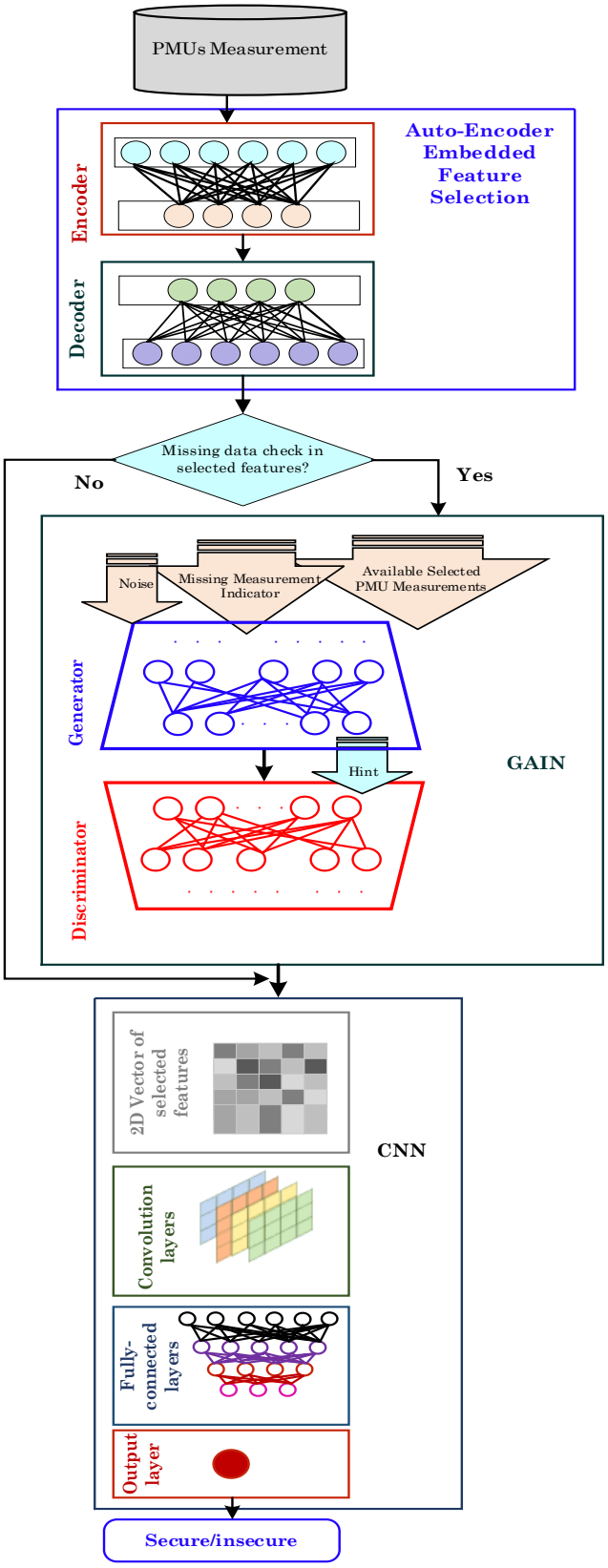



**Fig. 2.** Overall structure of the designed network

where $\Upsilon$ is a binary random variable, and the symbol $\odot$ shows element-wise multiplication.

4) *GAIN for PMU Missing Data Correction*

The GAIN is designed based on the loss function. Similar to the conventional GAN, the loss function is a $min-max$ optimization problem. In this loss function, the probability of predicted $M$ is maximized in the discriminator, while the value of the loss function using predicted $M$ is minimized in the generator. The loss function in the GAIN is:

$$\max_D V(D,G) = \mathbb{E}_{\overline{X_k},M,Z,H}\,[M^T * \log D(\overline{X_k},H)] + \mathbb{E}_{\overline{X_k},M,Z,H}[(1-M)^T * \log(1-D(\overline{X_k},H))] \quad (9)$$

Similarly, the loss function in the generator network is:

$$\min_G V(D,G) = \mathbb{E}_{\overline{X_k},M,Z,H}[(1-M)^T * \log(1-D(\overline{X_k},H))] \quad (10)$$

Consequently, the loss function of the GAIN is:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\overline{X_k},M,Z,H}[M^T * \log D(\overline{X_k},H)] + \mathbb{E}_{\overline{X_k},M,Z,H}[(1-M)^T * \log(1-D(\overline{X_k},H))] \quad (11)$$

The optimal parameters of the GAIN are obtained based on the iterative process by a $min-max$ game corresponding to the value function until it reaches the optimum global point. According to [21], to guarantee the global optimum in the GAIN, two different conditions should be satisfied:

1. The optimal parameters of the discriminator considering the fixed generator parameters are obtained through:

$$D^*(\overline{X_k},H) = \frac{\rho(\overline{X_k},H,M=1)}{\rho(\overline{X_k},H,M=1)+\rho(\overline{X_k},H,M=0)} \quad (12)$$

2. The global minimum for $\max_D V(D,G)$ can be obtained if and only if:

$$\rho(\overline{X_k}|H,M) = \rho(\overline{X_k}|H) \quad (13)$$

The mathematical proof can be found in [20].

In the GAIN-based missing PMU reconstruction, an additional term is added to the loss function. This additional term keeps the generated data within a minimum difference from the actual (not missing) data measured by the PMUs. The additional term is:

$$\min_G MSE(\hat{X}_k,\overline{X_k},M) = \sum M \odot [(\hat{X}_k - \overline{X_k})^2] \quad (14)$$

## C. Convolution Neural Network

The raw data selected by the deep feature selection module and missing data reconstructed by the GAIN module are considered as the inputs to the 2D-CNN network. A CNN is a powerful classifier that can fully understand information with diverse spatial features, such as wide-area power systems [25]. The CNN module consists of convolutional, flattening, and fully-connected layers [26]. The input feature maps are then fed into the convolutional layers to recognize patterns based on 2D filters. The output of the convolution layers is:

$$X^l = f_{RELU}(\sum_{i=1}^{F} W_i^l \otimes X^{l-1} + b_i^l) \quad (15)$$

where $l, F, W$, and $b$ represent the layer number, filter number, weights matrix, and bias matrix, respectively. The activation function is ReLU [27], which is represented by $f_{RELU}$.

The outputs of convolution layers are flattened by the *flattening* layers and converted into 1D arrays. The 1D arrays then pass to fully connected layers with ReLU as the activation function. The last fully connected layers use *sigmoid* as the activation function and determine the classes (secure/insecure status of the power system). The loss function of the CNN, namely binary-cross entropy, is:

$$f^{loss} = -\frac{1}{N}\sum_{i=1}^{N} y_i * \log \hat{y}_i + (1-y_i) * \log(1-\hat{y}_i) \quad (16)$$

where $\hat{y}_i$ and $y_i$ represent the predicted and actual status.

## D. Data Generation

In previous studies, the use of historical data in small-signal and $N-1$ power system security assessment, and the generation of a comprehensive dataset, are not properly addressed. However, an efficient data generation method has recently been presented in [28]. To assess all possible operational conditions, it is required to handle an $NP$-hard problem. The method presented in [28] resolves this problem. In this method, Latin hypercube sampling (LHS) is used to generate uniform samples based on the operating point of the power systems. Second, the infeasibility certificate technique is applied to guarantee the feasibility of the operating points generated by the sampling technique. The infeasibility certificate is a technique that relaxes the non-convex AC optimal power flow (AC-OPF). Utilization of the infeasibility certificate technique can also reduce the operating point space.

The next step in the methodology involves simulating a specific contingency set on a test case and evaluating the system's operating point based on small-signal security criteria. In this study, the considered contingencies focus on the outage of each existing power line, which represents a potential disruption in the system. Following the initial assessment, sensitivity analysis is employed to identify operating points that are in close proximity to the small-signal security boundary. These points are of particular interest because they are more critical for the system's security assessment. The purpose of this analysis is to ensure that the selected operating points capture potential vulnerabilities in the system. To validate the compliance with $N-1$ security requirements, the selected operating points undergo $N-1$ security assessment. This assessment involves checking for violations of line thermal limits and voltage constraints. By doing so, the analysis aims to confirm that the system remains within acceptable bounds and avoids any compromising conditions. Specifically, the line thermal limits are examined to ensure that the current on each transmission line remains within safe levels to prevent excessive heating or damage. Additionally, the voltage constraints are assessed to verify that the voltage levels at each busbar maintain stability and prevent voltage collapse.

### *E. Training Process*

Figure 2 illustrates the overall structure of the designed modulated power system security assessment, as follows:

*Step 1*: Input organization

A subset of the raw features measured by PMUs is selected using AE-based embedded feature selection considering the loss function in (4).

*Step 1-1*: The raw input data are fed into the feature selection module as a $(N, D)$ matrix and the output is a $(1, K)$ indexed vector of selected features, where the parameters $N$ and $D$ represent the number of operating points during an event and the number of features in the training dataset, respectively. $K$ represents the number of selected features, which is a constant parameter and selected in a way that the inputs can be converted from 1D to 2D vectors (example:$K = a \times a$ )….

*Step 2*: GAIN training

Based on the selected $K$ features, the input $(N, D)$ matrix transforms into a $(N, K)$ matrix. To mimic the missing data, a proportion of the data $\zeta$ are deliberately removed based on the missing rate of PMU data and a random process. The size of the outputs in the GAIN network is $(N, K)$, considering imputed missing data.

*Step 3*: CNN-based classification

The imputed matrix with a size of $(N, K)$ is fed into the CNN module, which then outputs the probability of secure/insecure status. The output probability is compared with the constant, and the security class of the power system is determined accordingly.

***Remark:*** Each module is a neural network-based structure, and each module's parameters are optimized by minimizing the loss function with the Adam optimizer [29]. Adam is a stochastic gradient descent algorithm, the details of which are provided in [30].

### *F. Practical Implementation of the Proposed Deep-Based Power System Security Assessment*

To implement the proposed method in the power system, the following steps should be conducted:

*Step 1*: PMUs measure and then send the data to the power system central control system.

*Step 2*: Missing data occurrence is checked.

*Step 3*: The feature selection module is activated, and the features are extracted corresponding to the selected subset in the trained AE model.

*Step 4*: If the selected subset feature maps include missing data, then the GAIN module is activated and imputes the data it generates; otherwise, goes to the next step.

*Step 5*: In the final step, the selected features (imputed by GAIN in the case of missing data occurrence) are fed into a 2D-CNN to determine the security status in terms of small-signal and $N - 1$ security constraints.

## III. CASE STUDY

This section discusses the numerical results of the proposed deep-based power system security assessment criterion. To this end, a 162-bus power system is considered. The optimal location of the PMUs in the benchmark power system is selected based on the method presented in [31], considering zero injection bus (ZIB). Therefore, this paper assumes that the full observability condition is met (shown in Fig 3). However, PMU missing data and partial observability are unavoidable. To address this issue, several cases are considered: i) no PMU data missing; ii) 10 % of PMU data missing; and iii) 20% of PMUs data missing. To apply the feature selection technique, assume that 90% of the data should be reduced.

For comparison, several state-of-the-art methods are considered, as follows:

- Random forest feature importance with 100 trees and heights of 10 leaves for each tree;
- Mutual information with 16 neighbors;
- Relief [16]; and
- Robust ensemble learning [6].

The input dataset is generated in MATLAB software and further processed by Python in a computer system with an Intel Core i7 CPU and 32 GB of RAM.

### *A. Metrics*

The metrics are computed based on the confusion matrix with the general form as depicted in Table I.

TABLE I
CONFUSION MATRIX

| | | Predicted | |
|---|---|---|---|
| | | Secure | Unsecure |
| Actual | Secure | TP | FN |
| | Unsecure | FP | TN |

In this regard, evaluation metrics are defined as follows ([21]):

$$\mathrm{R^c} = \frac{TP}{TP + FN} \tag{17}$$

$$\mathrm{S^p} = \frac{TN}{TN + FP} \tag{18}$$

$$\mathrm{A^c} = \frac{TN + TP}{TN + FN + TP + FP} \tag{3}$$

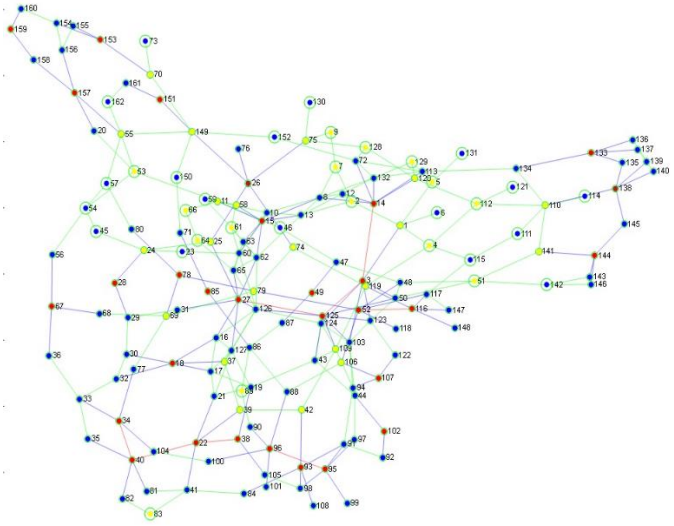


**Fig. 3.** 162-bus optimal PMU placement considering ZIB (red points are optimal PMU locations)

where $\mathrm{R^c}$ shows the recall metrics, also known as the true positive rate, as the proportion of correctly predicted secure

cases to overall secure cases. $S^p$ indicates the specificity (true negative rate), defined as the proportion of correctly predicted insecure cases to overall insecure cases. Finally, $A^c$ yields the correctly predicted cases over all cases. In cases of missing data occurrence, all combinations of PMU missing data have been taken into account. To compute the metric for the missing conditions, the average value of metrics is computed for each case.

### *B. Data Generation*

The 162-bus NESTA grid is considered as the benchmark system. This system includes 17 generators and 284 transmission lines. To generate a comprehensive dataset, 50,000 events are considered. The events cover about 23239 secure and 26761 insecure cases. Based on the data generation process presented in [28], the small-signal security boundary is determined based on the damping ratio rule (3%). The input data include active/reactive powers of the power system at lines/buses. To train the dataset, about 75% of the data are devoted to the training process, while the rest are considered for the testing process. The overall generated features include 568 features (in the benchmark system), and 50,000 samples (overall for the training and testing process).

The parameters of each module of the proposed method are given in Tables II to IV (where *B* represents the batch size in Tables).

TABLE II
AUTO-ENCODER PARAMETERS

| Layer | Shape | Weight | Bias | Parameters |
|---|---|---|---|---|
| Input | [B,568] | | | 0 |
| Encoder layer | [B,49] | [568,49] | | 27832 |
| Decoder layer (1) | [B,49] | [49,49] | 49 | 2450 |
| Decoder layer (2) | [B,49] | [49,49] | 49 | 2450 |
| Decoder layer (3) | [B,568] | [49,568] | 568 | 28400 |
| Total parameters | | | | 44,002 |

TABLE III
GAIN PARAMETERS

| Layer | Shape | Weight | Bias | Parameters |
|---|---|---|---|---|
| Generator Input | [B,98] | | | 0 |
| Generator layer (1) | [B,49] | [98,49] | 49 | 4,851 |
| Generator layer (2) | [B,98] | [49,98] | 98 | 4,900 |
| Generator layer (3) | [B,147] | [98,147] | 147 | 14,553 |
| Generator layer (4) | [B,147] | [147,147] | 147 | 21,756 |
| Generator layer (5) | [B,98] | [147,98] | 98 | 14,504 |
| Generator layer (6) | [B,49] | [98,49] | 49 | 4,851 |
| Generator layer (7) | [B,49] | [49,49] | 49 | 2,450 |
| Discriminator Input | [B,98] | | | 0 |
| Discriminator layer (1) | [B,49] | [98,49] | 49 | 4,851 |
| Discriminator layer (2) | [B,98] | [49,98] | 98 | 4,900 |
| Discriminator layer (3) | [B,147] | [98,147] | 147 | 14,553 |
| Discriminator layer (4) | [B,147] | [147,147] | 147 | 21,756 |
| Discriminator layer (5) | [B,98] | [147,98] | 98 | 14,504 |
| Discriminator layer (6) | [B,49] | [98,49] | 49 | 4,851 |
| Discriminator layer (7) | [B,49] | [49,49] | 49 | 2,450 |
| Total parameters | | | | 135,730 |

TABLE IV
CNN PARAMETERS

| Layer | Shape | Weight | Bias | Parameters |
|---|---|---|---|---|
| Input | [B,7,7,1] | | | |
| Conv2D layer(1) | [B,7,7,80] | [2,2,1,80] | 80 | 400 |
| Conv2D layer(2) | [B,7,7,140] | [2,2,80,140] | 140 | 44,940 |
| Conv2D layer(3) | [B,7,7,88] | [2,2,140,88] | 88 | 49,368 |
| Conv2D layer(4) | [B,7,7,108] | [2,2,88,108] | 108 | 38,124 |
| Flatten | [B,5292] | | | |
| Dense layer (1) | [B,80] | [5292,80] | 80 | 423,440 |
| Dense layer (2) | [B,88] | [80,88] | 88 | 7,128 |
| Output layer | [B,1] | [88,1] | 1 | 89 |
| Total parameters | | | | 563,489 |

- AE-decoder-based embedded feature selection

The parameters are given in Table II, including the number of hidden layers, neurons, weights, and biases determined based on trial and error. A key factor in the feature selection method is the regulation parameter $\lambda$, selected as an exponential function of epochs.

- GAIN

Table III depicts the parameters of the designed GAIN network. The key parameter is missingness value $\zeta$ and is considered as a fixed value $\zeta = 0.4$. The missingness value is selected based on the portion of missing data in the selected data obtained by the feature selection module. In addition, a sensitivity analysis is performed on the GAIN module with respect to the $\zeta$ parameters to represent the performance of the GAIN module on imputing missing data. The mean absolute percentage error (MAPE) is selected as an evaluation metric, and is calculated as $MAPE = \frac{1}{N*K}\sum_{i=1}^{N*K}\left|\frac{Y_i^{Actual}-Y_i^{Predicted}}{Y_i^{Actual}}\right|$, where $N$ is the total number of testing events and $K$ is the number of selected features. The results are presented in Fig 4. Note that, with 50% of the selected features missing, the GAIN module can replace data with a MAPE of only 2.96 %.

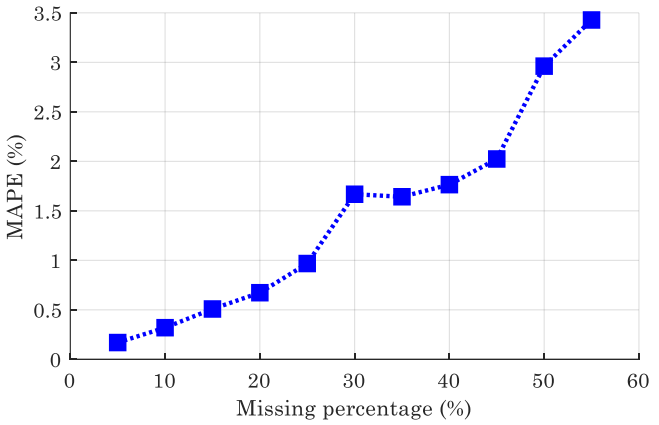



**Fig. 4.** GAIN module performance for different amounts of missing values of selected features

- CNN module

The CNN parameters and hyperparameters are selected using *Adam* and *random search* [32], respectively, and depicted in Table IV. Reducing the dataset based on the dimension reduction module reduces the convergence speed of the random search. Therefore, the values that can be assigned for designing the deep network require far less computational effort.

Note that the learning rate parameter for updating the weights in the backpropagation process is selected to equal 0.001.

### *C. Discussion of Results*

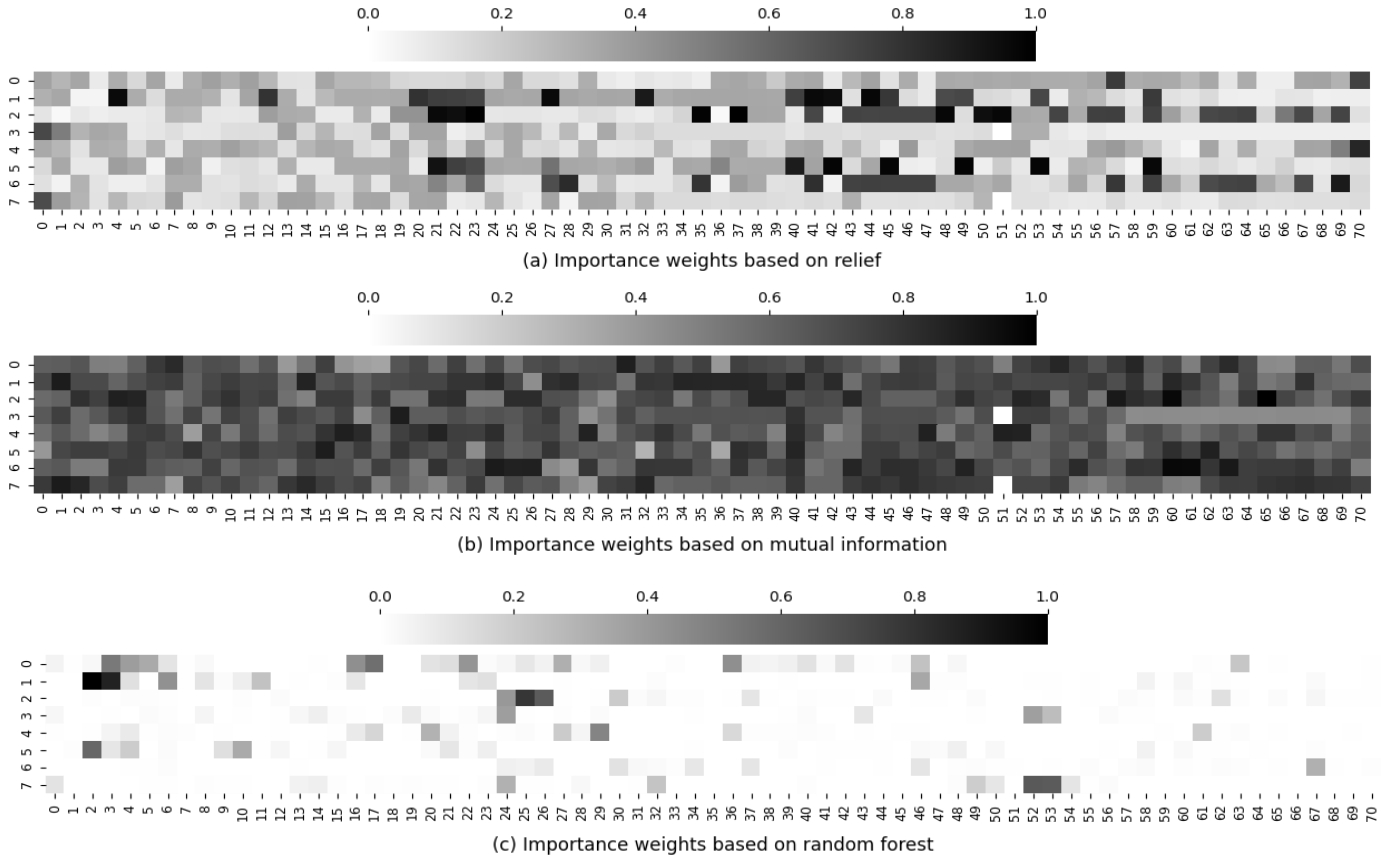
The results are evaluated based on feature selection impact and performance of the designed methodology in the presence of missing data.

#### 1) **Feature selection**

To address the superiority of the embedded feature selection technique, the proposed method is compared with mutual

TABLE V
METRIC RESULTS OF DIFFERENT FEATURE SELECTION METHODS

| Metrics | Selected features | Random forest + CNN | Relief + CNN | Mutual information + CNN | Proposed method |
|---|---|---|---|---|---|
| $A^c$ | 25 | 98.22 | 95.80 | 96.91 | **99.21** |
| | 36 | 97.89 | 95.55 | 96.89 | **99.07** |
| | 49 | 98.18 | 95.92 | 97.52 | **99.10** |
| $S^p$ | 25 | 98.16 | 94.99 | 96.57 | **98.58** |
| | 36 | 97.91 | 93.98 | 96.41 | **99.04** |
| | 49 | 98.54 | 95.65 | 97.25 | **98.17** |
| $R^c$ | 25 | 98.02 | 96.02 | 96.80 | **99.75** |
| | 36 | 97.54 | 96.60 | 96.92 | **99.10** |
| | 49 | 97.52 | 95.55 | 97.43 | **99.93** |



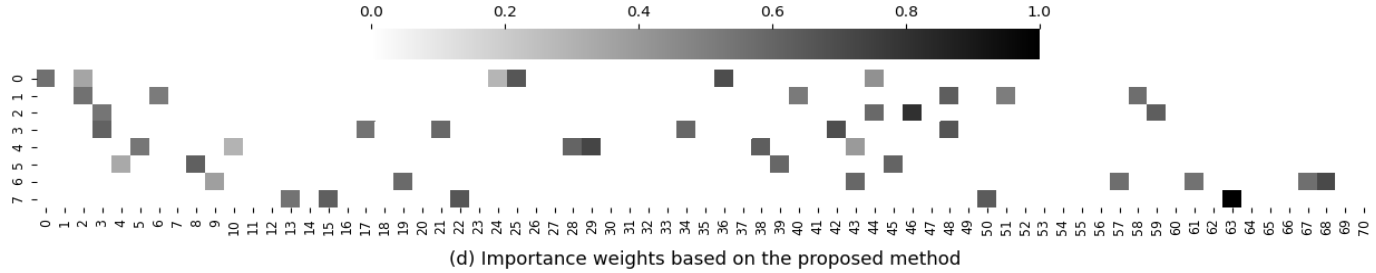



**Fig. 5.** Feature map of importance weights in different feature selection methods based on: a) relief; b) mutual information; c) random forest; and d) the proposed method.

information, relief, and random forest, which are widely used as powerful feature selection techniques. The results discuss the combination of CNN and these different feature selection techniques. In each case, the parameters and hyperparameters of the CNN are optimized by *Adam* and *random search* separately. The importance weights of the selected features are depicted in Fig. 5. Darker colors are assigned to the features with higher importance weights. The suitable feature selection can discriminate the features (assigning importance weights to sparse values). Therefore, the presented method is superior to other methods for the feature selection process. The performance of the different feature selection methods with CNN is illustrated in Table V. The higher accuracy and reliability of the proposed method is clear, which verifies the effectiveness of the designed AE-based feature selection in the power system security assessment.

Because dimension reduction might affect the performance of the CNN, a sensitivity analysis is performed with respect to the number of selected features. The sensitivity analysis results are shown in Fig. 6. The proposed deep-based feature selection method is significantly more reliable in terms of its specificity and recall. For instance, the designed feature selection combination with CNN improves the reliability of the random forest, mutual information, and relief approaches by about 1.15%, 2.7%, and 5.4%, respectively.

2) **Missing data consideration**

This section discusses the robustness of the designed network against missing data. To provide full observability in the 162-bus system (considering ZIB) [31], 30 different PMUs must be installed. Hence, $2^{30}$ possible different cases for PMU failure can occur. Evaluating these cases is not reasonable due to the high computational cost and their low probability of occurrence. Three possible conditions are considered to address this issue: 0%, 10%, and 20% missing PMUs data. In other words, the latter two cases represent three and six PMUs being unable to transfer their data; and the security of the power system is assessed under these partial observability conditions. The results are depicted in Table VI and compared with robust ensemble learning. The superiority of the proposed modulated deep network is obvious. For instance, consider the case with 20% of PMU data missing. Although it is a rare condition in practice, the accuracy of the proposed method is higher than 97.40%, while the accuracy of the robust ensemble learning is lower than 95.50%.

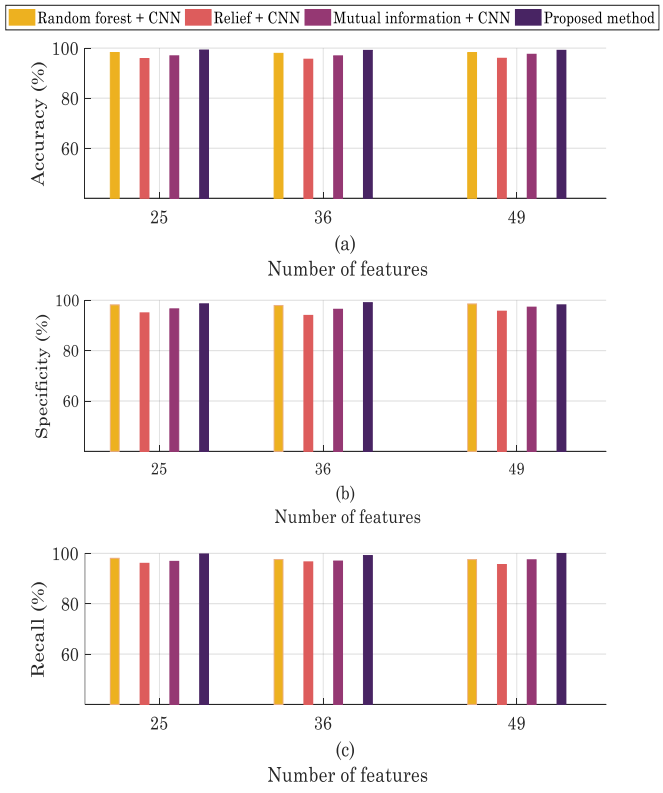



**Fig. 6.** Graphical view of Table VI for feature selection methods comparison: a) accuracy; b) specificity; and c) recall

TABLE VI
METRIC RESULTS FOR MISSING PMU DATA

| | Metrics | Robust ensemble learning [5] | Proposed method |
|---|---|---|---|
| 0% of PMU data missing | $A^c$ | 97.66 | **99.10** |
| | $S^p$ | 97.66 | **98.17** |
| | $R^c$ | 97.29 | **99.93** |
| 10% of PMU data missing | $A^c$ | 97.53 | **98.55** |
| | $S^p$ | 97.55 | **98.28** |
| | $R^c$ | 97.11 | **98.60** |
| 20% of PMU data missing | $A^c$ | 95.48 | **97.44** |
| | $S^p$ | 95.49 | **97.50** |
| | $R^c$ | 95.07 | **96.96** |

3) **Computational burden**

As mentioned before, one of the major advantages of the designed network, particularly using a feature embedded selection network, is the low computational burden. To illustrate this matter, Table VII compares the testing time of the CNN with/without the feature selection method. The results show that the testing time is reduced by approximately 91.66% by using the proposed feature selection method. Moreover, the dimension of the raw data in the power systems is reduced by 91% after implementing the dimension reduction method. Thus, feature selection can play an important role in the efficient improvement of computational performance for static/dynamic power system security assessment.

TABLE VII
EFFECTS OF FEATURE SELECTION ON CNN TESTING TIME

| | **CNN with input shape 162×162×1** | **CNN with input shape of 7×7×1** |
|---|---|---|
| Testing time | 3.12 ms | 0.26 ms |

## IV. CONCLUSION

This paper develops a data-driven framework to assess power system security in a real-time manner. To this end, a modulated network consisting of three modules, i.e., AE-based feature selection, GAIN-network-based missing data correction, and CNN-based main classification, is proposed. This multi-module approach benefits from: i) an unsupervised feature selection to reduce the computational burden; and ii) a data-driven method to enhance the method's robustness considering the failure of PMUs as an undesirable and unavoidable phenomenon. Previous investigations on data-driven security assessment suffer from the lack of interpretability of measured signals by PMUs and dependency on network observability. To overcome such challenges, the proposed GAIN network uses a new loss function and noise generation process.

Furthermore, a modified AE is equipped with new parameter regulation and relaxation techniques. The results show the designed feature selection module can select features with higher importance than several state-of-the-art methods. The accuracy of the combined AE-based feature selection is at least 1.12% higher than other strong feature selection approaches. The comparative results considering partial observability conditions also verify the proposed method's superiority and effectiveness, with an accuracy higher than 97.40% for the case with 20% of PMUs data missing, which corresponds to more than 2.1% improvement over robust ensemble learning.

## REFERENCES


[1] P. Kundur *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," IEEE Transactions on Power Systems, vol. 19, no. 3, pp. 1387-1401, Aug. 2004.

[2] You, S., Zhao, Y., Mandich, M., Cui, Y., Li, H., Xiao, H., ... & Zhang, Y. A review on artificial intelligence for grid stability assessment. In 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) (pp. 1-6). IEEE.,Nov.2020.

[3] L. Wang, X. Lin, F. Howell, and K. Morison, "Dynamic Security Assessment," Smart Grid Handbook, pp. 1-24, Jul.2016.

[4] L. G. Meegahapola, S. Bu, D. P. Wadduwage, C. Y. Chung, and X. Yu, "Review on oscillatory stability in power grids with renewable energy sources: Monitoring, analysis, and control using synchrophasor technology," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 519-531, Jan. 2020.

[5] P. Sekhar and S. Mohanty, "Classification and assessment of power system static security using decision tree and random forest classifiers," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 29, no. 3, pp. 465-474, May .2016.

[6] M. Mohammadi, G. Gharehpetian, and T. Niknam, "On-line small-signal stability assessment of power systems using ball vector machines," *Electric Power Components and Systems*, vol. 38, no. 13, pp. 1427-1445, Oct .2010.

[7] S. Kalyani and K. S. Swarup, "Classification and Assessment of Power system security using multi-class SVM," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 5, pp. 753-758, Dec. 2010.

[8] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep Neural Network Into Power Transformer Differential Protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 865-876, Jul .2019.

[9] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, and M. Khazaei, "Deep-Based Conditional Probability Density Function Forecasting of Residential Loads," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3646-3657, Feb. 2020.

[10] M. Afrasiabi, M. Mohammadi, M. Rastegar, and S. Afrasiabi, "Advanced Deep Learning Approach for Probabilistic Wind Speed Forecasting," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 720-727, Jul. 2020.

[11] S. K. Azman, Y. J. Isbeih, M. S. El Moursi, and K. Elbassioni, "A unified online deep learning prediction model for small signal and transient stability," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4585-4598, Jun. 2020.

[12] J. Cao and Z. Fan, "Deep Learning-Based Online Small Signal Stability Assessment of Power Systems with Renewable Generation," *Proceedings of 2018 IEEE SmartWorld Ubiquitous Intelligence & Computing Advanced & Trusted Computing Scalable Computing & Communications Cloud & Big Data Computing Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 216-221, Dec. 2018.

[13] R. S, R. S. Kumar, and A. T. Mathew, "Online Static Security Assessment Module Using Artificial Neural Networks," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4328-4335, Jun. 2013.

[14] J.-M. H. Arteaga, F. Hancharou, F. Thams, and S. Chatzivasileiadis, "Deep learning for power system security assessment," IEEE Milan PowerTech,IEEE, pp. 1-6. Mar. 2019

[15] M. Ramirez-Gonzalez, F. R. S. Sevilla, P. Korba, and R. Castellanos-Bustamante, "Convolutional neural nets with hyperparameter optimization and feature importance for power system static security assessment," Electric Power Systems Research, vol. 211, p. 108203, Oct. 2022.

[16] P. Beyranvand, C. Kucuktezcan, Z. Cataltepe, and V. M. I. Genc, "A novel feature selection method for the dynamic security assessment of power systems based on multi-layer perceptrons," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 6, no. 1, pp. 53-58, Mar. 2018.

[17] S. Liu *et al.*, "An integrated scheme for online dynamic security assessment based on partial mutual information and iterated random forest," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3606-3619, Apr. 2020.

[18] A. Abid, M. F. Balin, and J. Zou, "Concrete autoencoders for differentiable feature selection and reconstruction," *arXiv preprint arXiv:1901.09346*, Jan. 2019.

[19] Y. Zhang, Y. Xu, and Z. Y. Dong, "Robust ensemble data analytics for incomplete PMU measurements-based power system stability assessment," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1124-1126, Apr. 2017.

[20] C. Ren and Y. Xu, "A fully data-driven method based on generative adversarial networks for power system dynamic security assessment

with missing data," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5044-5052, Jun. 2019.

[21] J. Yoon, J. Jordon, and M. Schaar, "Gain: Missing data imputation using generative adversarial nets," 2018, arXiv:1806.02920. [Online]. Available: http://arxiv.org/abs/1806.02920

[22] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, Jun. 2013, vol. 30, no. 1, p. 3: Citeseer.

[23] I. J. Goodfellow *et al.*, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[24] S. Afrasiabi, M. Afrasiabi, B. Parang, M. Mohammadi, S. Kahourzade, and A. Mahmoudi, "Two-Stage Deep Learning-based Wind Turbine Condition Monitoring Using SCADA Data," *Proc. IEEE Int. Conf. Power Electron. Drives Energy Syst. (PEDES)*, pp. 1-6, Dec. 2020.

[25] S. Afrasiabi, M. Afrasiabi, M. Mohammadi, and B. Parang, "Fault localisation and diagnosis in transmission networks based on robust deep Gabor convolutional neural network and PMU measurements," *IET Generation, Transmission & Distribution*, vol. 14, no. 26, pp. 6484-6492, Dec. 2021.

[26] S. Afrasiabi, M. Mohammadi, M. Afrasiabi, and B. Parang, "Modulated Gabor filter based deep convolutional network for electrical motor bearing fault classification and diagnosis," *IET Science, Measurement & Technology*, vol. 15, no. 2, pp. 154-162, Mar. 2021.

[27] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, Jan. 2010.

[28] F. Thams, A. Venzke, R. Eriksson, and S. Chatzivasileiadis, "Efficient database generation for data-driven security assessment of power systems," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 30-41,Jan. 2019.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, Dec. 2014.

[30] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Designing a composite deep learning based differential protection scheme of power transformers," *Applied Soft Computing*, vol. 87, p. 105975, Feb. 2020.

[31] I. Abdulrahman and G. Radman, "ILP-based optimal PMU placement with the inclusion of the effect of a group of zero-injection buses," *Journal of Control, Automation and Electrical Systems*, vol. 29, no. 4, pp. 512-524, Aug. 2018.

[32] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of machine learning research*, vol. 13, no. 2, Feb. 2012.