IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE POWER & ENERGY SOCIETY SECTION

Received 18 October 2023, accepted 9 November 2023, date of publication 20 November 2023,
date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3334394

TOPICAL REVIEW

Review of Data-Driven Techniques for On-Line Static and Dynamic Security Assessment of Modern Power Systems

FABRIZIO DE CARO[1], (Member, IEEE), ADAM JOHN COLLIN[1], (Senior Member, IEEE), GIORGIO MARIA GIANNUZZI[2], COSIMO PISANI[2], AND ALFREDO VACCARO[1], (Senior Member, IEEE)

[1]Department of Engineering, University of Sannio (UniSannio), 82018 Benevento, Italy
[2]Terna Rete Italia SpA, 00138 Rome, Italy

Corresponding author: Fabrizio De Caro (fdecaro@unisannio.it)

ABSTRACT The secure operation of the transmission grid is of primary importance for any power system operator. However, the introduction of new technologies, market deregulation, and increasing levels of interconnectivity have made the analysis and assessment of power system security both more challenging and essential than ever. In this context, data-driven-based methodologies are being increasingly employed to classify and anticipate insecure future states, and make inferences on potential triggers of undesired operational conditions. This paper provides a comprehensive and systematic review of this fast-moving research area and covers data-driven-based methodologies deployed in both static and dynamic security assessment. Particular attention is paid to recent trends, such as the use of spatiotemporal feature selection algorithms and the increasing research activity in short-term voltage stability and frequency stability, which are not yet widely assessed as a collective in the existing literature.

INDEX TERMS Data-driven methodology, dynamic security assessment, frequency stability, machine learning, online, power system security, rotor stability, static security assessment, transient stability, voltage stability.

I. INTRODUCTION

A reliable power system satisfies the customer demand at a given quality level for the amount of time desired even in the presence of disturbances (i.e., adversarial conditions) that can affect its designed operation. These disturbances include a plethora of events, which vary in type and magnitude. Small disturbances are load variations (or in general, any difference between aggregate load and generation profiles), whereas large disturbances are faults to one or more components of the system. Usually, faults trigger the intervention of protection systems, which isolate the rest of the system from the faulted part, changing grid topology and affecting power flows. From a mathematical point of view, the disturbance affects the state of the system, forcing it to reach a new one through a system evolution trajectory, with different operating conditions. In this context, the goal of the system operator is to assess if the new operating conditions still satisfy the technical performance quality constraints. This task is known as Security Assessment (SA), and it can be split into two different subtasks according to the type of analysis: Static Security Analysis (SSA), if only the final steady-state (or post-disturbance) state is analyzed, and Dynamic Security Analysis (DSA), if the transitory system evolution is analyzed.

The associate editor coordinating the review of this manuscript and approving it for publication was Pratyasa Bhui.

A. POWER SYSTEM SECURITY ASSESSMENT

According to the joint IEEE/CIGRE working group [1] power system security "*refers to the capability of a power system using its existing resources to maintain reliable power supplies in the face of unexpected shocks and sudden*

disruptions in real-time, such as the unanticipated loss of key generation or network components, loss of fuel, or rapid changes in demand.'' SSA involves the assessment of the power system's steady-state operating condition after a contingency has occurred. The voltage levels at busbars and the current flow through lines should remain within their respective safe operating ranges to ensure a ''secure'' system state. Broadly speaking, if any of these constraints are violated, the state is considered ''not secure'' or ''insecure'', however, further states can be defined.

DSA is defined as ''*an evaluation of the ability of a certain power system to withstand a defined set of contingencies and to survive the transition to an acceptable steady-state condition*'' [2]. Particularly, an unstable system cannot be considered secure since instabilities can lead to dangerous operating conditions in power systems [1].

DSA examines one or more power system stability categories, such as rotor angle, frequency, and voltage, in the presence of small/large disturbances. In DSA analysis, the security level is assessed using performance indicators linked to the stability level of the system assuming a certain contingency [3], where the stability metric is computed after the hypothesized fault-clearing.

As such this is concerned with electromagnetic and electromechanical phenomena, i.e., the response of rotating machinary and control in the time scale up to a few seconds. However, the response of system control equipment, at time scales of up to tens of minutes, is also covered by DSA.

One of the main classifications of SA is based on the number of contingencies considered. Indeed, SSA of a power system is classified solely according to the number of contingencies (or events) considered. Typically, the ''$N-1$'' criterion is considered, and the system is defined as ''$N-1$ secure'' if all constraints are respected, i.e., the system will continue to function with the loss of one single component. In general, a system is deemed ''$N-k$ secure'' if all system constraints are met for k concurrent contingencies. This also serves as the basis of analysis for certain aspects of DSA, in which the system response to faults are considered. For others, small variations in load and generation around the equilibrium point (classified as small disturbances) are assessed using a range of operating conditions. These operating conditions are defined using probabilistic methods to vary the load and generation around the expected or known value.

The set of possible contingencies and/or operating conditions can be established by considering the most credible ones via operator experience. However, relying solely on operator experience may leave the system vulnerable to rare but significant events, which could cause cascading effects if not promptly resolved. On the other hand, considering ''$N-k$'' contingencies or too many load/generation operating conditions raises the issue of the explosion of possible combinations, which affects the computational burden and analysis time of all the possible states [2].

Thus, an important part of online applications is the careful selection of the number of multiple contingencies to analyze to promptly support power system operators. This aspect is particularly important for DSA since dynamic simulations, based on a deterministic set of nonlinear differential algebraic equations, are computationally expensive, especially in the presence of a large network, with a high number of machines, inverter-based resources, and non-linear components.

Hence, one of the greatest limitations of traditional SSA and DSA tools is that system operators use them multiple times per day. Each invocation requires numerical methods, particularly in DSA, which demand high-performance computing capabilities and efficient software routines. This process becomes computationally burdensome and costly, especially for large grids and extensive lists of probable contingencies. Data-driven methods can address the issues related to these limitations by introducing tools with reduced computational burden, leading to quicker and more cost-effective analyses. This makes them highly suitable for quasi-real-time power system operation strategies.

B. MOTIVATION AND CONTRIBUTION

Data-driven-based methodologies have become increasingly important, and power systems are not immune to the potential offered by this branch of science. When Machine Learning (ML) was first introduced into power systems, there was some scepticism about relying on tools that were not based on deterministic models but only on linking input and output data derived from historical data. However, with the improvements in data storage capacity, the proliferation of sensors that collect real-time system operation data, and the development of more advanced ML-based tools, the effectiveness of ML in power systems applications has been proven, even in very complex scenarios. Accordingly, ML-based tools are a well-established and important part of power system analysis and are currently used in a wide range of applications in power systems, from planning to real-time operation. A comprehensive list of these applications is available in [4].

ML has also been deployed to address SA problems. These applications are diverse and involve approaches for anticipating possible events, saving computational time by estimating system evolution due to contingencies and classifying and clustering events in post-disturbance analysis.

This review aims to analyze and discuss the role of data-driven-based models in SA problems, the unique features of the analyzed works, open problems, and the most promising research trends. While this topic has been the subject of review articles in the recent past, e.g., [5] and [6], the research area, in terms of the evolution of stability issues and the implementation of more sophisticated ML techniques, is developing at a rapid rate. With respect to the aforementioned papers, the contributions of this review paper are:

1) Processing of article metadata to provide a comprehensive, visual summary of the main features of the

analyzed works. This includes consideration not only of the number of publications but also the number of citations, an important indicator of overall interest and an aspect not covered in previous review papers;
2) Comprehensive overview of the ML workflow, introducing the ML-based models utilized for different SSA and DSA problems, which is extended here to include a detailed discussion of the stability criteria employed for each task;
3) Inclusion of the recent, important developments of the integration of spatiotemporal features in power system SA, which have occurred since the publication of previous reviews;
4) A clear distinction is made between the applied ML methods for all different types of DSA which adds new perspectives to the aggregated analysis presented in [5], while frequency stability was not covered in [6];
5) It is the first comprehensive review since the formal definition of two new types of DSA in [7] and [8], the implications of which are discussed as an important research topic moving forward (since the new forms of stabilities to analyze will affect the DSA tools, which must be set-up to consider the system insecurity caused by these new forms of system instabilities).

C. STRUCTURE

The review is structured as follows: Section II presents the literature search methodology and provides a summary of the article metadata; Section III provides an overview of the typical ML workflow and summarises the functionality of the feature selection/cardinality reduction techniques, ML models and performance evaluation indexes identified during the literature review; Sections IV and V analyze the state-of-the-art in ML applications for SSA and DSA, respectively, and compares the different input features, ML models and output data structures utilized; Section VI highlights critical issues, open challenges, and the most promising research trends; brief conclusions are provided in Section VII.

II. LITERATURE SEARCH METHODOLOGY

This section presents an analysis of the research works related to ML-based SA analysis in power systems, to reconstruct their storyline. Data was gathered from the Scopus database using several relevant research keywords (Data-driven / machine learning / artificial intelligence AND (security assessment OR static/dynamic security assessment, classification AND security assessment)) to search for relevant manuscripts using the title, abstract, and/or author's keywords. The extracted lists were merged, and manual filtering was applied to remove manuscripts that were not perfectly coherent with the topic area under analysis. Works submitted to the most relevant journals and conferences in the power system sector were considered for analysis. This procedure led to a collection of 147 articles in the period 1974 to 2023. All papers were read and further filtered to ensure that papers providing clear advances to the state-of-the-art and novel perspectives could be adequately discussed within this paper.

Fig. 1 displays the works collected and categorized by the type of SA (SSA vs DSA). It is interesting to note that the interest of researchers in applying ML to SA really started after 2010 and surged after 2018, with over 30 articles per year in 2021, although some manuscripts existed before that period. The presence of articles before 2000 indicates that researchers had already recognized the potential issues associated with SA in large power systems and the potential of ML-based applications in addressing the problem. However, it was only with the advent of economical computational resources and the introduction of novel and effective ML techniques that researchers began to actively explore and apply ML-based methodologies to power system SA. The figure demonstrates the growing interest in applying these models to DSA.

Fig. 2 shows the number of citations collected by type of SA (SSA vs DSA). The citations obtained by the articles referred to their year of publication seem to be equally distributed between DSA and SSA, indicating that the attention to both topics is still quite similar, despite the prevalence of DSA-topic related published articles as shown in Figure 1.

The methodology used to collect and count publications and references relies on using a developed algorithm that: a) extracts the relevant features extracted by the SCOPUS database using the previously mentioned keywords b) merges in a unique list for each keywords combination, c) rearranges data to be processed by R-based visualization tools.

III. MACHINE LEARNING FOR SECURITY ASSESSMENT

This section provides an analysis of the ML algorithms and feature selection techniques employed in data-driven-based security assessment applications. Its primary focus is to highlight the key characteristics and distinctions among the utilized techniques, thereby enhancing our understanding of their applicability and effectiveness.

A. TYPICAL WORK FLOW

The typical workflow for developing a ML model is shown in Fig. 3. Several variations of this workflow exist depending on the characteristics of the data and system to be modeled, as well as the decisions made by the developer. However, ML models and their application can be broadly classified depending on whether they output continuous values (regression problems) or class labels (classification problems).

The functionality of the feature selection algorithms, ML algorithms, and evaluation metrics identified during the literature review are discussed here. Input data and output are included in the wider analysis in Section IV (SSA) and Section V (DSA).

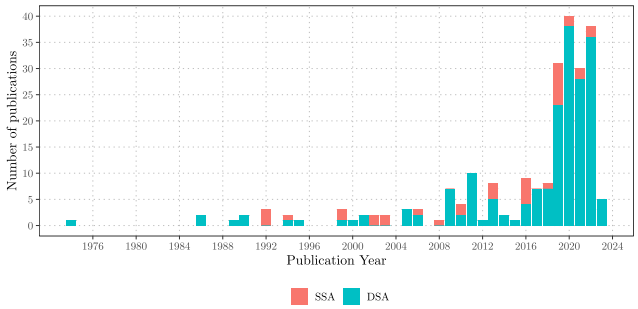

FIGURE 1. Number of publications on machine learning-based static security assessment (SSA) dynamic security assessment (DSA) per year.

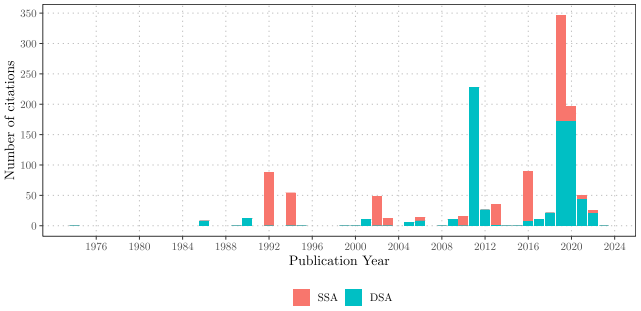

FIGURE 2. Number of citations of machine learning-based static security assessment (SSA) dynamic security assessment (DSA) per year.

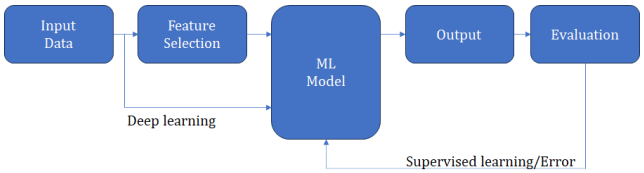

FIGURE 3. Typical machine learning model development workflow.

B. FEATURE SELECTION ALGORITHMS

In classification/regression problems involving a large number of input variables it is not probable that all have the same importance. In many cases, only a subset of them includes a large part of the knowledge to effectively build a map between predictor and target variables. Hence, it is much more probable that considering the full set of variables leads only to large training time and ineffective predictions, since a large number of variables may infer the algorithm training performance due to the curse of dimensionality [9].

In such scenarios, cardinality reduction algorithms are widely applied in data science, as these techniques identify the most promising input variable set or manipulate the original variables transforming them into new ones more suitable to build effective predictor-target mapping. Specifically, these two classes of algorithms are known as ''Feature Selection'' and ''Cardinality Reduction'' techniques [10]. Since the former class does not alter the original variables it is preferred where the interpretability of the results plays a crucial role. Hence, it is widely adopted in data-driven SSA/SDA. Feature selection algorithms can be further split into two classes: *filter* and *wrapper* methods. The former identify the most correlated input variables to the target one, whereas the latter identify the most ''useful'' in supplying the final prediction. On the other hand, another class of tools to reduce the size of the input matrix is represented by the cardinality reduction algorithm, such as the principal component analysis (PCA), which transforms the original features into new ones.

1) SEQUENTIAL FEATURES SELECTION

Sequential Features (or Forward) Selection (SFS) is a greedy technique that builds a set of input variables through an iterative approach [11]. Particularly, given the set of available M variables and an empty set of the most promising ones, a univariate model is built by considering every single m-th variable per time, and the one linked to the lowest error metric is picked up without replacement from the initial set to the most promising variables set. Thereafter, a bi-variate model is built by considering the previously selected variable combined with each one of the remaining $M - 1$. Also in this case the variable linked to the most effective bi-variate model is picked up from the initial set. The process ends when there is no significant accuracy improvement in the built multi-variate models at the k-th step respect with to the previous iteration. The final set of selected variables has $k - 1 < M$ variables. Clearly, this approach results in significant time consumption for datasets having large cardinality.

An alternative approach is to utilize the Sequential Backward Selection (SBS) technique, where the least promising features are gradually removed from the initial set. In contrast to SFS, SBS aims to eliminate features iteratively instead of adding them.

While both SFS and SBS can be used for feature selection, SFS is generally preferred over SBS due to its increased robustness against multicollinearity issues. Multicollinearity refers to the presence of high correlations among predictor variables, which can lead to unstable and unreliable model results. SFS is more effective in handling such situations and tends to produce more reliable feature subsets.

Therefore, SFS is often the preferred choice when conducting feature selection, as it provides a more robust and stable selection process compared to SBS.

2) MINIMUM REDUNDANCY MAXIMUM RELEVANCY

The objective of the minimum Redundancy Maximum Relevancy (mRMR) algorithm is to select a subset of input variables from an initial set. This subset consists of variables that exhibit maximum correlation with the target variable (maximum relevance) while minimizing redundancy among them (minimum redundancy).

Particularly, mRMR utilizes a forward procedure to identify a subset of relevant features, denoted as M_R, which is significantly smaller than the initial set of features (M) [12]. At each step of the procedure (for $g = 1$ to M_R), mRMR

selects the predictor variable that is both the least redundant and the most informative.

$$\underset{\mathbf{x}_j \in x - \Phi_{g-1}}{\arg\max} \left[I(\mathbf{x}_j; \mathbf{y}) - \frac{1}{g-1} \sum_{\mathbf{x}_i \in \Phi_{g-1}} I(\mathbf{x}_i; \mathbf{x}_j) \right] \quad (1)$$

where y is the vector of the target variable, and $\mathbf{x}_i$ and $\mathbf{x}_j$ are the i-th and j-th variable stored in the input matrix $\mathbf{X}$, respectively.

To achieve this, mRMR considers the set of previously selected variables, denoted as Φ_{g-1}, and calculates the mutual information between each candidate variable and the target variable. The mutual information, denoted as $I(x, y)$, measures the statistical dependency or information shared between two variables x and y. In the context of mRMR, the mutual information can be efficiently estimated using the formula $I(x, y) = \frac{1}{2} \ln(1 - \rho(x, y)^2)$, where ρ represents the Pearson correlation coefficient.

3) LINEAR DISCRIMINANT ANALYSIS

Fisher Discriminant is a statistical technique to perform cardinality reduction similar to PCA [13]. The novel variables are obtained through a linear combination of the features that maximally separates the data having different classes. Different from PCA, Linear Discriminant Analysis (LDA) takes into account the label of each point. Particularly, it reduces the dimensionality of input data whereas it preserves the discriminatory information, seeking to maximize the distance between different classes whereas it minimizes the variation within each class. Mathematically, given a matrix $\mathbf{X}$ with dimensions $[M \times S]$, where S is the number of samples and M of features where data belongs to C classes stored in vector $\mathbf{y} \in \mathbb{N}$, the LDA tries to find a linear map $\mathbf{y} = \mathbf{w}^T\mathbf{X}$ maximizing the following function:

$$F(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \quad (2)$$

where $\mathbf{w}$ is the column vector of weights with dimensions $[M \times 1]$, and $\mathbf{S}_b$ ($\mathbf{S}_w$) is the between-class (within-class) scatter square matrix. The order of these matrices is M. S_b is computed as:

$$\mathbf{S}_b = \sum_{i=1}^{C} N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (3)$$

where $\bar{\mathbf{x}}_i$ is the i-th the mean vector of the i-th class over, $\bar{\mathbf{x}}$ is the global mean column vector, and N_i is the number of samples of the i-th class.

On the other hand, $\mathbf{S}_w$ is computed as:

$$\mathbf{S}_w = \sum_{i=1}^{C} \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T \quad (4)$$

where $\mathbf{x}_{ij}$ and $\bar{\mathbf{x}}_i$ are the j-th sample and the mean vector of the i-th class. LDA solves the generalized eigenvalue problem by finding the eigenvectors and eigenvalues of the matrix $S_w^{-1} S_b$. The eigenvectors correspond to the directions in the input space that maximize the separation between classes. The eigenvectors obtained in the previous step are ranked according to their corresponding eigenvalues. Particularly, the eigenvectors with higher eigenvalues contain more discriminatory information. These eigenvectors, known as discriminant features, are selected to form a new feature subspace. The final step involves projecting the original data onto the subspace spanned by the selected discriminant features. The dimensionality of the data is reduced, as the new subspace has fewer dimensions than the original feature space.

4) RELIEF-BASED FEATURE SELECTION

Relief-based feature selection is a suitable method for extracting relevant features in classification problems. The main advantages of these techniques are their simplicity and their effectiveness, as well as their capability to avoid the curse of dimensionality issues. The relevancy of each feature towards a target variable is determined using a weight statistic, which ranges from -1 to 1, representing the worst and best relevance to the binary target variable, respectively [14]. Given a matrix $\mathbf{X}$ with S samples and M features, along with a binary vector $\mathbf{Y}$ containing S samples, the original Relief algorithm computes the feature rank through the following steps: a) Randomly extract a subset of s samples from the dataset. b) For each of the s samples, calculate the distance vector with respect to all other samples in the dataset. c) Identify the Nearest Hit, which are instances with the same class as the current sample, and the Nearest Miss, which are instances with a different class, based on the calculated distances. d) Update the weight for each feature by considering the distance between the feature values of each sample and the values of the Nearest Hit and Nearest Miss samples. Repeat steps a) to d) for all of the s samples to obtain a stable feature ranking. Sort the feature weights in descending order to visualize the most relevant feature ranking. By following these steps, the Relief-based algorithm identifies the features that have the most significant impact on the classification task, providing valuable insights for feature selection in ML problems.

5) PRINCIPAL COMPONENT ANALYSIS

Data reduction techniques differ from feature selection since they transform original ones having new features useful for the specific task. PCA is a widely used method for linear data reduction [15]. It accomplishes this by projecting the data from the original space into a lower-dimensional space, where the new axes, known as Principal Components (PCs), are computed by combining the original variables. The first PC is oriented along the direction with the maximum variance of the data [16].

The algorithmic procedure is as follows for a given matrix $\mathbf{X}$ with dimensions $[S, M]$, where S and M are the number of samples and features, respectively. The first step is to normalize the matrix $\mathbf{X}$ so that each column of $\bar{\mathbf{X}}$ has a zero mean and unit variance. Thereafter, the Singular Value

Decomposition is performed on $\bar{\mathbf{X}}$:

$$\bar{\mathbf{X}} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{5}$$

where $\mathbf{U}$ is an orthogonal matrix of order S, $\mathbf{D}$ is a rectangular diagonal matrix with dimensions $[S, M]$, with diagonal elements $d_1 \geq d_2 \geq \ldots \geq d_M$, and $\mathbf{V}$ is an orthogonal matrix of order M. Hence, the new variables in the lower-dimensional space are computed by selecting the first $k \leq M$ columns of the matrix $\mathbf{Z}$, where:

$$\mathbf{Z} = \bar{\mathbf{X}}\mathbf{V} = \mathbf{U}\mathbf{D} \tag{6}$$

There are various ways to choose the optimal number of PCs, one of which is to consider the percentage of variance captured by the selected components, with a value greater than 95% generally considered satisfactory.

C. AUTOENCODERS

Autoencoders are artificial neural networks (ANNs) used for dimensionality reduction in datasets. The objective of an autoencoder is to reconstruct the original data by minimizing the reconstruction error. The simplest form of autoencoder is a three-layer neural network (referred to as ''shallow''). In this process, the data in the first layer with dimensions $[S, M]$ is transformed into a new representation with dimensions $[S, M']$ in the second layer using a function called the ''encoder,'' denoted as f. Subsequently, in the third layer, another function g transforms the reduced data back into an estimation of the original dataset, thus restoring its original dimensionality; this process is known as the ''decoder.'' The number of neurons in the second layer controls the reduction in dimensionality [17]. If the functions f and g are linear them the autoencoder acts like PCA, whereas if they are non-linear then the autoencoders diverge significantly from PCA. They gain the ability to capture multi-modal aspects of the input distribution, a characteristic not shared with PCA. Autoencoders with multiple layers are called ''Deep-autoencoders'' [18]. Unfortunately, one of the greatest drawbacks of autoencoders is that they are prone to overfitting, since they are characterized by more parameters than input features.

D. CAUSALITY-BASED FEATURE SELECTION

Causality-based feature selection refers to a class of methods that apply inference analyses to identify the most promising predictor variables. Unlike correlation, which indicates only a statistical association, causality examines the direct relationship between predictors and target variables [19]. Approaches utilizing Bayesian Networks first construct a model based on the available dataset, creating an acyclic graph to determine the probabilistic dependencies between input and target features. Subsequently, features lacking a direct causal relationship with the target feature are identified and removed. This process is known as applying the Markov Blanket of the target feature [20]. Causal Graphical Models, Score-Based Models, and Differential Causality Analysis are other widely employed causality-based feature selection methods [19].

E. MACHINE LEARNING CLASSIFICATION/PREDICTION ALGORITHMS

Once the set of input variables has been reduced to the most promising ones, it is time to proceed with training the model for data-driven SSA/DSA. Training an effective model requires the selection of an algorithm capable of capturing the complex input-output relationship in the multi-input single-output problem, which involves exploring a high-dimensional space in $\mathbb{R}^{M_R}$. The choice of the algorithm should be guided by several factors, including the characteristics of the available data, the presence of non-linear relationships, the potential multicollinearity among input variables, and the diverse types of data being considered (such as real-valued data for active and reactive power spatial profiles, and binary data indicating active/inactive buses and lines). Since this research has evolved together with advances in ML algorithms, different algorithms have been applied to this task, defined as either ML or Deep Learning Algorithms. However, it is important to note that Deep Learning is a subset of ML, where deep stands for the usage of complex algorithms composed of a large number of interconnected layers interacting between them.

1) DECISION TREES

Decision Trees (DT) are algorithms largely used in classification tasks [21]. Given $\mathbf{X}$ and $\mathbf{y}$ as training data, it builds a hierarchical tree of rules where each node is labeled with an input feature, whereas the linked branch is labeled with a condition associated with the possible occurrences of the node the branch coming from. The lowest nodes in the tree are called leaves and they are associated with the target outputs occurring according to the realized data partitioning.

The DT is built by recursively partitioning the dataset based on the feature values that best separate the data according to the target variable. This process involves selecting the best feature and the optimal splitting criteria at each internal node. Common criteria used for splitting include information gain, Gini impurity, or entropy [22]. Given unknown data, their features drive the patch from the root node to the leaves over the tree according to the built set of rules.

DT are known for their interpretability, as they can be easily visualized and understood. They provide a clear hierarchical structure that allows users to trace the decision-making process. However, DTs can suffer from overfitting if not properly controlled, leading to poor generalization on unseen data. To mitigate this issue, techniques such as pruning and setting stopping criteria can be applied.

2) RANDOM FOREST

A random forest (RF) is created by combining multiple DTs. The number of DTs used in the RF is set before the learning

process and optimized using trial and error approach. Each DT is trained on a different subset of the original dataset, selected through a process called bagging, where samples are uniformly sampled with replacement. Random selection does not only involve samples but also the features, decorrelating the trees, and improving the performance.

The RF algorithm is based on the concept that a collection of weak predictors, which are uncorrelated and unbiased, can collectively improve prediction accuracy compared to using a single predictor alone [23]. In the RF, multiple weak models in the form of regression trees are generated from different subsets of the training data. These subsets are created by randomly selecting samples and features from the original training set, aiming to reduce the correlation between the weak predictors. The final prediction is obtained by considering the class more voted from all the weak models in the random forest. Generally, RF is preferred to DT since it better avoids overfitting, increasing the generality of the model.

3) SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) for classification is an algorithm that basically tries to find the hyperplane that optimally separates data points belonging to different classes [24]. In more detail, as the hyperplane divides the feature space into different regions, the algorithm finds a hyperplane maximizing the margin between the classes. This means maximizing the distance between the hyperplane and the closest points. Clearly, finding a hyperplane separating two classes is possible only when the data are linearly separable. For non-linearly separable data, SVM transforms the data from the original feature space to another one (with higher feature dimensions) which makes data linearly separable. This is performed through the so-called ''Kernel Trick'' [25], where the common kernel functions used in SVM include linear, polynomial, radial basis function (RBF) [26], and sigmoid. Mathematically, in a binary classification problem, the hyperplane is represented by a linear equation of the form:

$$\mathbf{w}^T\mathbf{X} + \mathbf{b} = 0 \tag{7}$$

where $\mathbf{w}$ is the normal vector to the hyperplane and $\mathbf{b}$ is the bias term. The parameters are estimated by addressing a constrained optimization problem. Since the original SVM can deal with binary classification problems, several methods are deployed to overcome this limitation such as ''Winner-takes-all'' and ''one-versus-one'' strategies [27].

4) K-NEAREST NEIGHBORS

The k-nearest neighbors (kNN) algorithm is a simple yet powerful approach that does not require training since the data itself serves as the model [28]. In this algorithm, the input variables are organized into a matrix $\mathbf{X}$ of dimensions $[S \times M]$, where S represents the number of samples and M denotes the number of input variables. Similarly, the corresponding output variables are stored in a vector $\mathbf{y}$ of dimensions $[S \times 1]$.

During the validation step, when a new input $\mathbf{x}_q$ of dimension $[1 \times M]$ is given, the algorithm calculates the distances between $\mathbf{x}_q$ and all S samples in $\mathbf{X}$. The samples with the shortest distances are selected, forming subsets $\mathbf{X}_{NN}$ and $\mathbf{y}_{NN}$. To generate the final output, the algorithm assigns weights to the selected samples based on their inversely proportional distances, which are then combined with $\mathbf{y}_{NN}$. The specific manner in which the output is determined depends on the nature of the problem being addressed, whether it involves classification or regression.

The kNN algorithm provides a flexible and intuitive approach for making predictions without the need for explicit training. By considering the distances between samples, it identifies the most similar instances to a new input and utilizes their corresponding outputs to generate a prediction.

5) ARTIFICIAL NEURAL NETWORK

ANNs have gained tremendous popularity in recent decades for their ability to capture nonlinear relationships in data by adjusting their parameters based on observed data. ANNs are inspired by the functioning of the human brain, with interconnected nodes arranged in layers. Each node receives input, performs a function, and produces an output. The function's weights within each node are tuned based on input-output data, enabling the network to adapt to specific tasks.

ANNs can be constructed with multiple layers, resulting in deep learning models. The term ''deep'' signifies the presence of multiple layers within the model. Deep learning models have achieved remarkable accuracy due to advancements in optimization techniques, increased computational power for large-scale computations, and the availability of vast datasets.

However, deep learning models also pose challenges, including limited interpretability. This arises from the significant number of parameters involved, which are necessary to account for complex nonlinear dependencies in the data. The abundance of parameters makes it difficult to comprehend the inner workings of the model and understand the factors influencing its predictions.

A vast number of techniques have been developed over the years and applied to data-driven-based SSA/DSA problems. Particularly, one-layer feedforward neural networks (FNN) were initially deployed, whereas the most recent applications rely on the usage of convolutional neural networks (CNN).

a: MULTI-LAYER FEEDFORWARD NETWORK

A Multi-layer FNN is a foundational architecture extensively employed in various applications, notably in classification tasks [29]. These networks, also known as Deep Neural Networks (DNN), derive their name from their multiple layers.

In a DNN, information flows through several layers in the same direction. Each layer performs transformations on the inputs using activation functions, such as Rectified Linear Unit (ReLU) or Exponential Linear Unit (ELU), where a more complete list is available in [30]. These activation

functions introduce crucial non-linearities, enabling the network to discern intricate patterns in data. When radial functions are used, the network becomes a Radial-Basis Neural Network, which is particularly effective for function approximation.

During training, the network's parameters (weights and biases) are fine-tuned through backpropagation, a process where optimization algorithms like gradient descent minimize the disparity between the network's predicted output and the true output. This iterative process refines the network's parameters, enhancing its performance on the given task [31].

The architecture's effectiveness is affected by the number of hidden layers and nodes per layer. Networks with too few layers may fail to capture complex relationships, leading to inaccurate models. Conversely, networks with excessive layers can grasp intricate input-output mappings, but an abundance of layers can induce the vanishing gradient problem. This phenomenon hinders effective training and optimization.

In classification tasks, the output layer comprises neurons equal to the classes to be predicted. These output neurons typically employ activation functions like the sigmoid function (logistic function), squashing the output between 0 and 1 and giving a decision user threshold. Each neuron's output can be interpreted as the probability of the input belonging to a particular class.

b: CONVOLUTIONAL NEURAL NETWORK

The CNN is well-suitable for addressing problems where there is a spatial relationship between data like images. It is a DNN composed of convolutional, pooling, and fully connected layers [32].

The convolutional layers are the core blocks of a CNN. They include multiple kernels, acting as filters, and perform convolution operations sliding over the images. These filters aim to detect local patterns or features in the input data. Pooling layers are placed after the convolutional layers, and their role is to perform spatial dimension reduction of the feature maps. Finally, fully connected layers elaborat the output of pooling layers, transforming the features into class probabilities or regression output.

One of the drawbacks of a CNN is the tremendous number of parameters, which require a large computation effort for the training process. Additionally, training deep CNNs often requires large labeled datasets and substantial computational resources to handle the increased model complexity.

c: FUZZY NEURAL NETWORK

A fuzzy feedforward neural network (FFNN) combines FNN with fuzzy logic to handle uncertainty in data. In an FFNN, the inputs are first fuzzified, which means that they are mapped to fuzzy sets [33]. Fuzzy sets allow the representation of uncertainty by assigning membership values between 0 and 1 to each element. The fuzzified inputs are then fed into the neural network, and the output values are then defuzzified to obtain crisp values, which are the final predictions of the network.

The main difference between FFNNs and FNNs lies in their handling of uncertainty. In a FNN, the inputs and outputs are precise values, and the network learns to map inputs to outputs based on the given training data. Hence, it does not take into account any uncertainty in the data. On the other hand, a fuzzy neural network incorporates fuzzy logic to represent and handle uncertainty in the data. Fuzzy sets are used to represent inputs, and the activation functions in the neurons are fuzzy membership functions. This allows the network to model and deal with uncertain data, making it more suitable for applications where uncertainty is present.

d: EXTREME MACHINE LEARNING

Extreme Learning Machine (ELM) is a special type of single-hidden layer FNN. Particularly, in an Extreme Learning Machine, the hidden layer's weights, and biases are randomly generated and fixed, unlike traditional neural networks where these parameters are learned through an iterative optimization process [34]. The output layer's weights are then analytically determined using a least-squares approach to minimize the error between the network's predictions and the actual target values. ELM has several advantages over traditional learning algorithms for neural networks such as a much faster training process since the hidden layer parameters are randomly generated and do not need to be updated through iterations, and fewer hyperparameters to tune.

6) ENSEMBLE PREDICTION STRATEGIES

Ensemble prediction involves combining multiple machine/deep learning models to improve accuracy compared to using individual models. An example of an ensemble prediction is the RF algorithm, whereas this section extends the concept to the hybrid ensemble prediction, where different types of algorithms are considered. The rationale behind ensemble prediction is that a single data-driven model may not always perform well consistently over time. By combining multiple models, the ensemble can reduce the risk of making large prediction errors, as there will always be one method that performs better than others at any given time [35].

In ensemble prediction, each model independently provides its prediction. The final prediction is obtained by merging the predictions from all models. In classification-based ensembles, the final prediction is determined by majority voting among the models. On the other hand, in regression-based ensembles, the final prediction is computed as a weighted average of the predictions from different methods. The weights are locally updated based on the recent performance of the individual methods in the ensemble over time, making the model dynamically adaptable.

Ensemble prediction is a powerful technique that leverages the diversity of multiple models to achieve better overall

performance and robustness in handling different patterns and variations in the data. It has been widely used in various applications to enhance predictive accuracy and stability [36].

F. PERFORMANCE EVALUATION

During the review process, it was observed that the vast majority of papers utilized the same standard evaluation metrics. As such, they are discussed here for completeness, but no comparison is included in the analysis presented in Section IV and V.

1) CLASSIFICATION PROBLEMS

The performance evaluation of models that tackle classification tasks in data-driven SSA/SDA can be done by utilizing metrics based on the analysis of the confusion matrix. In the case of a binary classification problem, assuming that a positive event is labeled $y = 1$ and a negative event as $y = 0$, the confusion matrix takes the following standard form shown in Table 1.

TABLE 1. Confusion matrix for binary classification problems.

		Observed Class	
		$y = 1$	$y = 0$
Predicted Class	$\hat{y} = 1$	TP	FP
	$\hat{y} = 0$	FN	TN

In the confusion matrix, *TP* (True Positive) represents the number of correct predictions of the positive event ($y = 1$), *FP* (False Positive) represents the number of times the negative event ($y = 0$) is incorrectly predicted as positive ($\hat{y} = 1$), *TN* (True Negative) represents the number of correct predictions of the negative event ($y = 0$), and *FN* (False Negative) represents the number of times the positive event ($y = 1$) is not predicted ($\hat{y} = 0$). In the context of system SA, the positive event typically corresponds to an insecure state for a given contingency.

Various metrics can be derived from the elements of the confusion matrix to assess the accuracy of the model as shown in [37]. The most commonly used metrics identified during the literature review include Accuracy, TPR (True Positive Rate or Sensitivity), TNR (True Negative Rate or Specificity), FPR (False Positive Rate), and FNR (False Negative Rate):

$$ACC = (TP + TN)/(TP + TN + FP + FN) \quad \in [0, 1] \quad (8)$$

$$TPR = TP/(TP + FN) \quad \in [0, 1] \quad (9)$$

$$TNR = TN/(TN + FP) \quad \in [0, 1] \quad (10)$$

$$FPR = FP/(FP + TN) \quad \in [0, 1] \quad (11)$$

$$FNR = FN/(TP + FN) \quad \in [0, 1] \quad (12)$$

It is crucial to note that the scores in the confusion matrix and related accuracy metrics are dependent on a specific decision threshold. Several classification techniques offer probability scores indicating the likelihood of input $\mathbf{x}$ belonging to a specific class. The conversion from predicted score to label is dictated by applying a decision threshold $\beta \in [0, 1]$. In binary scenarios, this can be represented as:

$$\hat{y} = \begin{cases} 1 & \text{if} \quad \mathcal{P}(\hat{y} = 1|\mathbf{x}) \geq \beta \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

and visualised using the Receiver Operating Characteristic (ROC) curve. An illustration of the ROC curve is depicted in Fig. 4. In the graph, an ideal classifier corresponds to the coordinates (0,1).

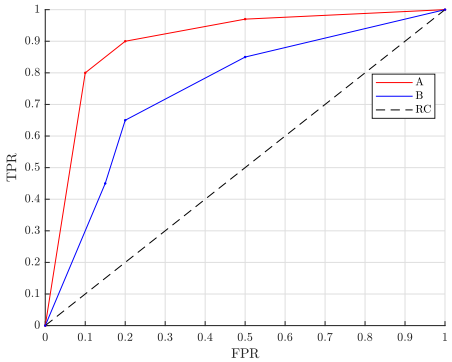

FIGURE 4. Example of ROC curve.

The Area Under the Curve (AUC) measures a classification model's effectiveness across various decision threshold values, where a higher AUC value signifies superior model performance. In Fig. 4 model "A" outperforms model "B", while RC represents a threshold at 0.5, indicating models must surpass this to outperform random classification.

However, AUC does not reveal the balance between TPR and FPR in the prediction model. Thus, metrics like the Geometric Mean or "G-Means" (14) are valuable for determining the trade-off between TPR and FPR on the ROC curve.

$$GM = TPR \times (1 - FPR) = TPR \times TNR \ \in [0, 1] \quad (14)$$

GM provides insight into the balance between TPR and FPR in predictions. The highest *GM* value indicates the point where this balance is optimized. A lower maximum *GM* value signifies poorer performance in predicting positive events, even if negative events are accurately predicted. A frequently employed metric to assess the performance of a classifier model is the F1-score (F1s), which is the harmonic mean between Precision $= TP/(TP + FP)$ and Recall $= TP/(TP + FN)$:

$$F1s = \frac{2 \times TP}{2 \times TP + FP + FN} \in [0, 1] \quad (15)$$

The metrics defined above could be extended to multi-classification problems by setting a class and considering all the remaining ones as a single class. In this case, more analyses are necessary to find the best operation point.

2) REGRESSION PROBLEMS

The most widely utilized metrics found during the literature review for comparing the continuous output prediction of

regression problems in SA applications are the mean absolute error (MAE), the mean average percentage error (MAPE), the mean squared error (MSE), and the root mean squared error (RMSE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{16}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{\hat{y}_i - y_i}{y_i}| \tag{17}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2 \tag{18}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}{N}} \tag{19}$$

where N is the number of tests, $\hat{y}_i$ and y_i are the predicted and known value of the i-the test, respectively.

IV. STATIC SECURITY ASSESSMENT

A. STATIC SECURITY INDEXES

SSA requires computing state variables (magnitude/phase voltages and power flows for all branches) for a given operating state (active and reactive spatial power profiles) and considering outages. However, relying solely on solving a power flow equation does not provide a comprehensive assessment of the system's security grade. To address this limitation, suitable metrics have been introduced to transform violations of power systems constraints into more interpretable and comparable forms. These metrics play a crucial role in providing insightful information regarding the security status of the system. Particularly, the computed indexes are transformed into security system labels according to heuristic rules, given specific thresholds, where the label to estimate will be the output of a data-driven SSA problem. In this section, the main static security indexes used in the analyzed manuscripts are reported.

1) PERFORMANCE INDEX

The following metrics, which are all defined as Performance Indexed (PI), analyze only a specific type of violation per time [38]:

$$PI_b = \sum_{j=1}^{N_b}\left(\frac{\Delta V_j}{\Delta V_j^{max}}\right)^{2n} \tag{20}$$

where ΔV_j and ΔV_j^{max} are the observed and maximum allowable voltage deviation in the j-th bus, and N_b is the number of buses. Another metric considers voltage deviation as:

$$\begin{cases} PI_v = \sum_{j=1}^{N_b} w_j\left(\dfrac{V_j - V_j^{sp}}{\Delta V_j^{lim}}\right)^{2n} \\ V_j^{sp} = \dfrac{V_j^{max} + V_j^{min}}{2} \\ \Delta V_j^{lim} = \dfrac{V_j^{max} - V_j^{min}}{2} \end{cases} \tag{21}$$

where w_j is the bus priority index assigned to the j-th bus, and V_j^{min} and V_j^{max} are the lower and upper tolerance voltage security bounds.

Other versions of the Performance Index analyze violations in terms of the branch current/power (line/transformer) of the systems:

$$PI_{ol} = \sum_{i=1}^{N_l}\left(\frac{I_i}{I_i^{max}}\right)^{2n} \tag{22}$$

$$PI_{wmw} = \sum_{i=1}^{N_l}\frac{w_i}{2n}\left(\frac{S_i}{S_i^{max}}\right)^{2n} \tag{23}$$

where I_i (S_i) and I_i^{max} (S_i^{max}) are the observed and maximum current (apparent power) flows, respectively, and N_l is the number of branches.

Given the considered PI, the system state label is assigned according to the following heuristic rules:

$$\text{system state} = \begin{cases} \text{secure} & PI = 0 \\ \text{alarm state} & \text{if } 0 < PI \leq 1 \\ \text{insecure} & \text{if } 1 < PI \end{cases} \tag{24}$$

2) STATIC SECURITY INDEX

The Static Security Index (SSI), as proposed by [39], incorporates the assessment of constraints violations pertaining to line overloading and bus over/under voltages. This index serves as a valuable measure to evaluate the static security of a power system by considering these critical aspects:

$$SSI = \frac{W_1\sum_{i=1}\text{LOI}_i^{N_l} + W_2\sum_{j=1}^{N_b}\text{VDI}_j}{N_l + N_b} \tag{25}$$

where the indexes i and j refer to the i-th line/transformer and the j-th bus, N_l and N_b, is the total number of lines/transformers and bus, W_1 and W_2 are weights selected according to the security priority of power system operator, and LOI and VDI are the Line Overloading Index and Voltage Deviation Index, respectively. LOI is defined as:

$$\text{LOI}_i = \begin{cases} \dfrac{S_i - S_i^{max}}{S_i} \times 100 & \text{if } S_i > S_i^{max} \\ 0 & \text{if } S_i \end{cases} \tag{26}$$

where S_i is the apparent power flow through the ith line/transformer and S_i^{max} is the corresponding rate value.

On the other hand, VDI assumes the following form:

$$\text{VDI}_j = \begin{cases} \dfrac{|V_j^{min}| - |V_j|}{|V_j^{min}|} \times 100 & \text{if } |V_j| < |V_j^{min}| \\ \dfrac{|V_j| - |V_j^{max}|}{|V_j^{max}|} \times 100 & \text{if } |V_j| > |V_j^{max}| \\ 0 & \text{if otherwise} \end{cases} \tag{27}$$

where V_j is the voltage in the j-th bus V_j^{min} and V_j^{max} are the lower and upper voltage bound thresholds, respectively. Once the SSI is computed, the security label of the system can be assigned by the following ruleset:

$$\text{system state} = \begin{cases} \text{secure} & \text{if } 0 < SSI \leq 1 \\ \text{critically secure} & \text{if } 1 < SSI \leq 5 \\ \text{insecure} & \text{if } 5 < SSI \leq 15 \\ \text{highly insecure} & \text{if } SSI \geq 15 \end{cases} \tag{28}$$

Equation (25) is modified in [40] by adjusting its denominator according to:

$$SSI = \left(\frac{W_1 \sum_{i=1}^{N_l} \text{LOI}_i^{N_l} + W_2 \sum_{j=1}^{N_b} \text{VDI}_j}{100} \right)^{1/(2n)} \tag{29}$$

where this version of SSI removes the weight selection and masking problem [38]. In this case, the Authors defined only two state labels according to $SI = 0$ (secure state) and $SI > 0$ (insecure state).

3) COMPOSITE SECURITY INDEX

The Composite Security Index (CSI), developed by [41], considers violations on both lines/transformers and voltages:

$$\text{CSI} = \left[\sum_j^{N_b} \left(\frac{d_{v,j}^u}{g_{v,j}^u} \right)^{2n} + \sum_j^{N_b} \left(\frac{d_{v,j}^l}{g_{v,j}^l} \right)^{2n} + \sum_i^{N_l} \left(\frac{d_{p,i}}{g_{p,i}} \right)^{2n} \right]^{\frac{1}{2n}} \tag{30}$$

where factors d take into account voltages (subscripts u and v) and line power flow deviations (subscript p) from alarm limits, and g are normalization factors:

$$d_{v,j}^u = \begin{cases} \dfrac{|V_j - F_j^u|}{V_j^d} & \text{if } V_j > F_j^u \\ 0 & V_j \leq F_j^u \end{cases} \tag{31}$$

$$d_{l,j}^u = \begin{cases} \dfrac{|F_j^l - V_j|}{V_j^d} & \text{if } V_j < F_j^u \\ 0 & V_j \geq F_j^u \end{cases} \tag{32}$$

$$g_{v,j}^u = \frac{V_j^u - F_j^u}{V_j^d} \tag{33}$$

$$g_{v,j}^l = \frac{F_j^l - V_j^l}{V_j^d} \tag{34}$$

where V_j^u (V_j^l) are the voltage upper and lower secure limits, whereas V_j^u (V_j^l) are the upper and lower alarm limits.

Finally, the line flow limits are considered as follows:

$$d_{p,i} = \begin{cases} \dfrac{|P_i| - P_{F,i}}{\text{base MVA}} & \text{if } |P_i| > P_{F,i} \\ 0 & \text{if } |P_i| < P_{F,i} \end{cases} \tag{35}$$

$$g_{p,i} = \frac{P_{P,i} - P_{F,i}}{\text{Base MVA}} \tag{36}$$

where $P_{F,i}$ is the line power flow limit.

Once the CSI index has been calculated, the system state label is computed according to the following rules:

$$\text{system state} = \begin{cases} \text{secure} & \text{if CSI} = 0 \\ \text{alarm} & \text{if } 0 < \text{CSI} \leq 1 \\ \text{Insecure} & \text{if } 1 < \text{CSI} \end{cases} \tag{37}$$

4) APPLICATION

SSA aims to verify whether power system operational constraints are violated in the steady-state reached after a contingency through post-disturbance analysis. The monitored constraints typically comprise bus voltages (magnitudes and phase angles) and line current flows. The system security is assessed by linking these variables to the system state label according to heuristic rules, in some cases related to the values of computed severity indexes. A power flow computation is necessary to monitor the constraints in the presence of an outage. Security is linked to specific spatial load and generation profiles, called operation scenarios, and outages, respectively. The security of the system changes with the considered outage to the same operating condition. Iterative approaches are necessary to assess the security of the system for a list of contingencies. However, the time required for this process, although the advancements in computation resources, may be comparable to the available time to make decisions in real-time operation. For this reason, ML has been applied to speed up the process in several manuscripts. Table 2 and Table 3 present the main features of SSA-related works, including details such as the type of indices used, type of learning (classification, regression, etc.), deployed algorithm, analyzed grids, and input data. Manuscripts proposing clustering or classification data-driven algorithms to address SSA are included in Table 2, whereas those using regression-based data-driven algorithms are included in Table 3.

The literature analysis has revealed significant similarities between the reviewed works. Most methods address a classification problem where the input features include the operating conditions and the considered outage, and the output is a binary or multiclass label that reflects the state of the system. Interestingly, synthetic data are predominantly used, and the direct use of real data is limited. The input data differ slightly, with some methods considering basic features like spatial demand and generation profiles, while others use the pre-contingency system state. In the latter case, a power

TABLE 2. Works addressing classification problems in SSA-related tasks.

Ref	ML Algorithm	Test Case	Input Data	No. of States
[42]	Kohonen's SOM	3 / 7 / 19 bus	Injected $\mathbf{P}$,$\mathbf{Q}$; $N-1$ contingency	3
[43]	RST	IEEE 118 bus	**Power Systems State**; $N-1$ contingency	2
[44]	SVM	Nordic32 system	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{V}$, $\mathbf{\Theta}$; $\mathbf{S}$	2
[45]	Enhanced RBF NN and WTA Network	IEEE 14 / 30 / 118 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{V}$, $\mathbf{\Theta}$; $\mathbf{S}$ (Pre-contingency)	2
[46]	RF	30 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{V}$, $\mathbf{\Theta}$;	4
[47]	C4.5	IEEE 30/57/118/300 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{V}$, $\mathbf{\Theta}$; $\mathbf{S}$	2
[48]	CNN	Mexican Power Grid (81 bus)	$\mathbf{P}$,$\mathbf{Q}$ (bus data)	2
[49]	DT	Hellenic Equivalent Grid (240 bus)	$\mathbf{P}$,$\mathbf{V}$ (not all power plants)	2
[50]	SOFM	IEEE 14 30 bus	Power flow output	2
[40]	ANN	IEEE 14 bus / 30 bus / 246	Direct and Indirect PMU measurements	2
[51]	ANN	IEEE 8 bus	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{V}$, $\mathbf{\Theta}$;	2
[52]	SVM	30 / 118 bus	Injected $\mathbf{P}$,$\mathbf{Q}$ and $N-1$ congestion	3

Note: Bold symbols mean the input for a specific variable class is multivariate.
$\mathbf{P}$: Injected active powers, $\mathbf{Q}$: Injected reactive powers, $\mathbf{V}$: bus voltage magnitudes, $\mathbf{\Theta}$: bus voltage phases, $\mathbf{S}$: line power flows

TABLE 3. Works addressing regression problems in SSA-related tasks.

Ref	ML Algorithm	Test Case	Input Data	Output Index
[53]	CNN	IEEE 9/30/57/118/181/300/1354 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{Y}$, $N-1$ Contingency	SI
[54]	ANN-MLP RBFN	IEEE 30 bus	Injected $\mathbf{P}$, $\mathbf{Q}$; $\mathbf{V}$, $\mathbf{\Theta}$, $\mathbf{S}$ (Pre-Contingency); $N-1$ Contingency	PI and SI
[55]	MLP	5 / IEEE 14 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$; $N-1$ Contingency	PI and SI
[56]	MLFNN	IEEE 118 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$ (Demand); $N-1$ contingency	PI
[57]	CNN	IEEE 9/30/57/118/181/300/1354 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$; $\mathbf{Y}$, $N-1$ Contingency	Power Flow output ($\mathbf{V}$,$\mathbf{\Theta}$)
[58]	CNN	IEEE 57/1354 bus system	Injected $\mathbf{P}$,$\mathbf{Q}$, $\mathbf{Y}$	ACOPF calculation
[59]	Deep Learning Model	(Multi-Area model)	Injected $\mathbf{P}$,$\mathbf{Q}$; transformer tap status; meteorological data;	Security Margins

See Table 2 notes.

flow analysis is carried out without outages. Notably, some methods do not include the outage as an input feature, which is essential for assessing system security under the same pre-contingency state. In these cases, a ML model is trained for each unique outage, which increases computational costs and partly offsets the benefits of data-driven approaches over traditional Newton-Raphson methods for security assessment in online operations. Finally, feature selection algorithms are widely used to reduce the number of features, where the considered algorithms are SFS, class separability index, and heuristic rules.

B. MACHINE LEARNING FOR SPEEDING UP THE SECURITY ASSESSMENT GIVEN A LIST OF OUTAGES

One of the first works on the deployment of machine learning in SSA was proposed by [42]. The main aspect of this manuscript was the deployment of a Kohonen Self-Organizing Map (SOM), which is a neural network specialized for unsupervised learning. Particularly, SOM is able to arrange multi-variate data (n-d) in a lower space (2-d) according to the data structure. The features, in this case, are the generation and demand active and reactive profiles ($\mathbf{P}$ and $\mathbf{Q}$) and the considered contingency. Once the data are arranged it is possible to identify the reduced dimension spaces since the area corresponds to an insecure/secure/critical state. When new inputs are available (for example in real-time operation), the position of new input in the map can easily return the security level of the system.

Another example of early work on the deployment of ML-based methods for power system SA was proposed by [51]. In this work, a basic neural network was used to predict system security in terms of voltage violations, where the input variables are the power flow variables in the pre-contingency states and the considered outages.

In [49], the authors proposed a data-driven SSA approach using DTs. They trained a DT model for each credible contingency by randomly generating spatial load profiles from a base case. The study was conducted on the Hellenic transmission grid, and the input features were reduced using an expert-based approach. The input features included tap transformer settings and the power injected by thermal power plants covering peak loads, such as hydroelectric and diesel plants. The output of the model is the system security label given a specific contingency. Although this approach shows promising results, it has limitations in handling large and complex grids with numerous contingencies since many models should be trained.

An RBF was proposed to detect the security of a power system given a set of inputs in [44] and [45]. In both cases, a system label (secure or not-secure) is assigned by considering a set of possible contingencies given a certain operational scenario, which is represented by active

and reactive generation and demand profile, spatial voltage magnitude, phase, and line current flow.

One interesting aspect of these two works is the proposal of a statistical method based on the class separability index to select the most informative features and discard the irrelevant ones. However, the authors claim that the obtained variables are linked to pre-contingency states, and the type of contingency is not reported in the feature list. Given the same operational scenario, the security of the system can vary depending on the topology after a contingency. Hence, to discriminate between pre-contingency operating scenarios with different security labels caused by different topologies a different neural network is considered for each contingency. The IEEE-14, IEEE 30, and IEEE-118 (only in [45]) bus grids were considered as test cases, where the operating conditions are randomly generated.

The authors of [56] deploy a multilayer FNN to estimate the severity index to assess the security of the system in the presence of a set of credible contingencies. The aim is to reduce the computational time linked to solving a huge number of power flow problems given certain operating conditions. Since the output of the problem is a numerical index, the addressed task can be represented as a regression problem; the estimated index is the PI. According to heuristic rules, the estimated value of PI is converted into a system security label. The input variables of the forecasting problems are loading conditions and the considered $N-1$ contingency. The IEEE-118 bus is considered in this study, where the loading conditions are randomly generated by changing the base value of a certain percentage threshold.

Specifically, this work considers a set of spatial active and reactive injection profiles at each bus, as well as binary variables indicating the status of the grid topology (active and inactive branches) and the considered contingency (line outages). By incorporating these various inputs, the proposed methodology aims to accurately assess the security of the transmission grid through the estimation of static security indexes as SI and PI. Unfortunately, this approach was tested only for small networks such as 5 and 14 bus networks, where the operative scenario data are randomly generated from the rate base case values.

The work presented in [60] employed a similar set of input variables presented in [56] and applied the least absolute shrinkage and selection operator (LASSO) to estimate the PI index based on the pre-contingency operating state and the specific line outage being considered. Operating scenarios were generated by randomly varying the load conditions between 50% and 100% of the rated value. The study was conducted on two power systems with 118 and 300 buses, respectively.

In [47], the authors proposed a data-driven security analysis using the C4.5 DT and SFS to reduce the number of input features. The C4.5 DT is a modified version of the ID3 DT, which uses the Gain Ratio as the splitting criterion instead of the Information Gain. Particularly, the authors of this manuscript use the pre-contingency operating conditions as input features and the system security label as the output. While this work is interesting for its consideration of high penetration of photovoltaic power capacity, it does show how the developed methodology discriminates between the same pre-contingency states and the security labels for different outages, without specifying the outages as input. On the other hand, the Authors have considered many networks for tests (30, 50, 118, and 300 bus grids), whereas the spatial load and PV power profiles are randomly generated from a base case.

The work proposed by [40] presents some interesting and novel aspects compared to the majority of data-driven-based SSA manuscripts. In particular, this manuscript proposes a case-based reasoning approach based on the deployment of a kNN to predict the system security label given a certain contingency. However, the most relevant aspect is the use of a reduced number of variables as input features. Only data collected by PMU buses and pre-contingency power flows of the lines connected to them are considered as input features. Furthermore, feature selection based on class separability measures is considered. Another innovative aspect is the presence of several modules aimed at predicting the type of violation, which is a missing aspect in all other analyzed works. Several test cases, up to the Indian 246 bus network, are considered in this work.

The work proposed by [52] is similar to other approaches in the field. It involves the deployment of a SVM to address a multiclass system security problem, where the input features include the traditional pre-contingency power flow problem and the considered outage. The authors also employ SFS to reduce the number of input features. 30 and 118 bus networks were considered in this study.

Deep learning algorithms have been utilized for this task to take advantage of their ability to learn from data with spatial structure. In particular, CNNs have been shown to be effective in processing both the grid structure and spatially correlated data. The authors of [53] applied CNNs to data-driven SSA, where injected active and reactive power profiles, along with grid structure, were used as input features. Specifically, only the non-diagonal elements of the susceptance matrix were considered, since the imaginary part of the mutual admittance becomes null in case of an outage on a line or transformer, whereas the conductance could already be zero for some transmission lines or transformers. The proposed approach was evaluated on several test cases, up to a 300-bus network, achieving satisfactory accuracy.

In [46] and [54], a methodology was proposed to predict the PI index for line congestions and voltage magnitudes using a combination of the multi-layer perceptron and radial basis function network. The goal was to predict these values based on pre-contingency variables and the considered $N-1$ contingency. The study was conducted on the IEEE-30 bus system, and the spatial injected active and reactive power profiles were randomly generated from the base case.

In [55], the authors proposed an ANN-based methodology using saturated linear couple neurons (sl-CONE) for transmission grid security assessment. One of the notable aspects

of this work is the set of considered data used to train the model and how the authors model the neural network to address potential issues that may arise from the usage of certain input data.

C. MACHINE LEARNING TO ACCELERATE THE CONTINGENCY SCREENING PROCESS

This section contains a collection of recent works where ML was utilized to expedite the contingency screening process. Unlike the previous collection, where ML was applied to return a label or severity index without solving power flow problems, here machine learning is used for subtasks such as estimating power flow solutions.

In [57], the authors proposed a deep CNN to speed up the process of $N-1$ contingency screening. The approach aims to estimate the state variables that are usually computed through a power flow problem. The main motivation for this work is that, although a power flow problem requires only a few seconds to be solved with current computation capacity, the number of scenarios becomes too large in the presence of wind power generation, hence uncertain spatial power injection profiles.

The authors of [59] proposed a method to forecast the security of power systems using both electrical and meteorological data. The security state is linked to the detection of low-security margins of each power transfer interface, which is defined as the complement of the ratio between the observed power flow and the total transfer capability. A power transfer interface is defined as a portion of an electric grid by Transmission System Operators for dimensionality reduction. The proposed method utilizes a deep spatial neural network to forecast security margins using historical data. A real case study of the Guangdong power system, consisting of 21 geographical zones, is considered in this study.

Another interesting approach was proposed by [58], where CNN and depth-first-search are employed for fast cascading outage screening. CNN is deployed to speed up the optimal power flow solution, whereas depth-first search is used to emulate multiple outage scenarios and find that having the highest expected cumulative security index. Particularly, for each step of the cascading sequence, the CNN is performed to estimate the SI at each outage stage.

V. DYNAMIC SECURITY ASSESSMENT

As explained in Section I-A, DSA examines one or more power system stability categories in the presence of small/large disturbances. This means that a data-driven DSA algorithm predicts an index/performance indicator or a security class by considering the system response to a fault (hence after it occurs) [3].

Traditionally, DSA categories were defined as rotor angle stability, voltage stability, and frequency stability. Further subdivision is normally applied based on either the magnitude of the disturbance or the time response of the disturbance. These are visualized in Fig. 5. Moreover, the proliferation of converter interfaced generation has recently led to the proposal of two new categories: converter-driven stability and resonance stability [7]. These are discussed further in Section VI-A.

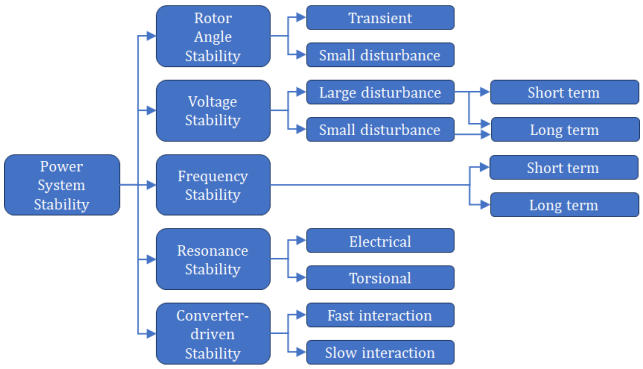

FIGURE 5. Classification of power system stability, adapted from [8].

The issue of computation time discussed in Section IV is particularly pertinent when it comes to DSA. The computational burden associated with solving the large systems of ordinary differential equations used to represent generator dynamics, uncertainty in the dynamic response of load, and the need to include models of control systems is not negligible, and the available response time to implement the necessary corrective actions is limited. Accordingly, the advantages proffered by data-driven techniques have led to significant interest in their use in this area.

This section is structured to reflect the general maturity of research into the use of ML for DSA. As such, this starts from rotor angle stability, which was the focus of the first research in this area, before proceeding through long-term voltage stability and short-term voltage stability. With respect to the other areas, research into frequency stability is still developing. The decision to separate voltage into long-term and short-term is due to the relatively even distribution of research into these topics, while rotor angle stability tends to focus on the system response to faults (i.e., transient stability). Each section discusses the stability criteria assessment, input data and feature extraction methods, and ML algorithms applied for each class of stability.

Regarding stability indexes, this review compiles and discusses the stability metrics as described above. More information about stability indexes can be found in [61], [62], and [63]. In [61], the authors explore stability indexes that can be computed from data collected by PMUs, while [62] places greater emphasis on voltage stability indexes. Reference [63] extends this analysis to the most sensitive metrics to assess voltage stability in the presence of massive PV penetration.

A. ROTOR ANGLE STABILITY (TRANSIENT STABILITY)

Rotor angle stability is the ability of synchronous machines to keep phase synchronization during disturbances. It was

also observed that this was commonly referred to directly as *transient stability* in the reviewed manuscripts.

1) STABILITY CRITERIA

The most widely applied criterion for stability is based on the phase angle difference between generating buses in the system within a given observation window. The applied difference will depend on the system, with either 180° [64], [65], [66], [67], [68], [69] or 360° [70], [71], [72] used as the threshold for binary classification.

Alternative scalar criteria are given considering the ratio with respect to 360° [73]:

$$\eta = \frac{360^\circ - |\Delta\delta|_{max}}{360^\circ + |\Delta\delta|_{max}} \tag{38}$$

or with respect to 180° [74]:

$$\eta = \frac{180^\circ - |\Delta\delta|_{max}}{180^\circ + |\Delta\delta|_{max}} \tag{39}$$

where $|\Delta\delta|_{max}$ is the maximum rotor angle difference among all generators and $\eta > 0$ is considered as stable. These are exploited by the proposed methodologies to provide further insight into any identified instability.

Other criteria, such as the transient security index in [75] and wide area severity indices [66], [76], have been utilized but are not as widely applied.

2) INPUT DATA AND FEATURE EXTRACTION

The input data utilized with the ML algorithm is relatively diverse, but still, the majority are still readily available from network PMU measurements. These are summarised in Table 4. The majority of methods make a clear distinction between generator buses and load/system buses. This is to be expected due to the nature of the phenomena. Note that the rotor angle, although specified explicitly by some methods, is denoted as V_{phase} in Table 4 as internal values can be easily calculated from the terminal PMU measurements [65].

Pre-processing of the input data using feature selection (extraction) methods has been applied in several papers. In [64], a simple approach, defined as the *single ranking method*, computed the importance of each input feature by calculating the number of correct classifications using the variable. The variable with the highest importance score and also those highly correlated with it were removed. Then the variable with the second highest (and those correlated) was removed and so on until a fixed number of features had been selected. In [64], the number of features was set as six as this was similar to the number of generators in the test network. In [83], the mRMR feature selection was utilized and identified generator electrical power output and busbar voltages as the most important features.

Other established algorithms (i.e., those defined in Section III-B) have also been applied. In [75], SFS was applied and obtained a dimensionality reduction of almost 4% for the IEEE 118-bus network and 18% for the IEEE NE 39-bus network. ReliefF was used in a semi-supervised learning framework in [80] to identify the 50 top-ranked features (of which 30 were randomly selected for model training). Reference [81] combined PCA and ReliefF for feature selection in an approach developed to address the problem of analyzing a large number of potential faults.

In [67], a novel feature selection method based on neighborhood rough set theory was used to characterize the separability and impact of 32 different input features, achieving classification accuracy of over 95% with a reduction rate of up to 80 %. Reference [68], introduced a *near real time* stage in the traditional offline-online two-stage workflow. This additional stage incorporated knowledge of the system to learn the causal structure between features in order to identify highly relevant features and update the model for the given operating conditions. This was compared against an exhaustive list of techniques comprised of filter techniques, i.e., mRMR, Correlation-based Feature Selection (CFS) and Joint Mutual Information (JMI), an embedded method, i.e., SVM Recursive Feature Elimination (RFE), and the SFS wrapper method. It was found that the best existing method was SFS, however, the proposed method reduced computation time by up to 75 %. Close to real-time applications can help to reduce the uncertainty in the network topology and operating conditions.

However, several recent works in this area are based on deep learning algorithms which are more adept at learning multiple dependencies in the input data and explicit feature reduction stages are not required.

3) ML ALGORITHMS

Some of the first work in this area, presented in the 1970s, was discussed referring to *pattern recognition*. One such example was presented in [64], in which simple linear and quadratic functions (and combinations of these) were successfully applied to classify power system disturbances into two classes (stable or unstable). More sophisticated ML methods were subsequently employed, with early examples of ANNs and DTs reported in [84] and [70], respectively. Reference [70] was also the first work to report the successful application to a ''non-trivial'' case, i.e., the IEEE New England 39-bus test system.

Since then, there has been a wealth of different methods applied in this area to improve particular aspects of the assessment process. For example, in [65] a fuzzy hyper-rectangular composite neural network (FHRCNN) was introduced with improved performance compared with existing FFNN approaches, an ensemble DT was implemented for the large-scale 783-bus Hydro Quebec grid planning model, while improved interpretability of the decision was explored using a multi-class SVM (which characterized four states) in [39]. Naturally, these trends are still the subject of ongoing research in this area.

Reference [77] considered the problem of generalization for different power system configurations using a CNN to

TABLE 4. Rotor angle stability (transient stability) assessment input data and machine learning (ML) algorithms.

Ref.	ML Algorithm	Gen Bus				Other Bus				Branch		System Inertia	Addressed Problem	Output (*No. of States*)
		P	Q	V_{mag}	V_{phase}	P	Q	V_{mag}	V_{phase}	P	Q			
[64]	Pattern Recognition (Sequential Method)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	Classification	(*2*)
[70]	DT	-	-	✓	✓	-	-	✓	✓	-	-	-	Classification	(*2*)
[65]	Fuzzy NN	-	-	-	✓	-	-	-	-	-	-	-	Classification	(*2*)
[66]	DT	-	-	✓	✓	-	-	✓	✓	-	-	✓	Classification	(*2*)
[75]	SVM	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	Classification	(*4*)
[77]	Twin Convolutional SVM	✓	-	✓	✓	✓	-	✓	✓	-	-	-	Classification	(*2*)
[78]	CNN	-	-	✓	✓	-	-	-	-	-	-	-	Classification	(*4*)
[79]	Cascaded CNN	-	-	-	✓	-	-	-	-	-	-	✓	Classification	(*4*)
[80]	Ensemble (kNN and ELM)	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	Classification	(*2*)
[81]	Hybrid (ELM)	✓	✓	✓	-	✓	✓	✓	-	-	-	-	Classification	(*2*)
[74]	Recurrent Graph CNN	-	-	✓	✓	-	-	✓	✓	-	-	-	Classification	(*2*)
[71]	DBN	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	Classification	(*2*)
[68]	Probabilistic Graphical Model	-	-	✓	-	-	-	✓	-	-	-	-	Classification	(*2*) per stage
[73]	Transfer Learning and CNN	✓	-	-	✓	✓	✓	✓	✓	-	-	-	Classification	(*2*)
[72]	Multi-Label-Learning	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	Classification	(*2*) per line
[82]	Graph Attention NN	-	-	✓	✓	-	-	✓	✓	-	-	-	Classification	(*3*)
[83]	WTA NN	✓	-	✓	-	-	-	✓	-	-	-	-	Classification	(*2*)

Where the output of regression problems is the index utilized and the output of classification problems is given by the number of states (*N*)

extract features to supply a twin SVM classifier. Indeed, due to the properties discussed in Section III-E5, CNNs are a common choice for DSA. In [78], a heatmap image was used to represent the measured system variables as an input to a CNN classifier. This work was also able to identify critical generators in the system to help the system operator's decision-making.

Large batch contingencies for day-ahead dispatch were the subject of research in [79]. The approach used cascaded CNNs to capture data from different lengths of time series data. Notably, this was applied to the 2,383-bus Polish power system. The analysis and handling of time series in DSA applications is increasingly investigated in recent publications. One such method, which combined a graph convolutional network (GCN), to handle spatial characteristics, and a long short-term memory (LSTM) unit, to handle temporal features, to form a graph convolutional network (RGCN), was proposed in [74]. Additionally, the assessment was treated as a multi-task learning problem which produced both the stability classification and identified critical generators.

While these approaches have proven successful, the development of deep learning techniques is opening further possible applications. Due to their nature, these solve problems that straddle input data, feature selection, and classification. Reference [80] presents one of the first semi-supervised learning methods, with the objective of reducing the need to generate a large number of data samples to update and retrain models. This approach used labeled and unlabelled training data and a data editing technique called depuration which updates/removes erroneously labeled samples using the kNN algorithm. This was combined with an Extreme Learning Machine (ELM) as the classifier engine. However, although powerful, deep learning algorithms are inherently complex, and the need to deliver interpretable decisions is acknowledged in [71]. In the presented approach, a local linear interpretation is applied to identify the importance of input features if the system is unstable.

The problem of missing input data was addressed using a generative adversarial network in [85]. Following this, an ensemble learning model was implemented using ELM and random vector functional link networks (RVFL). The same ML models (ELM and RVFL) were combined in [81] to address the problem of the large number of potential faults using a transfer learning-based method. Another transfer learning framework was proposed in [73] using CNN as the classifier. A further extension is considered in [72] which developed a multi-label learning method to simultaneously address different multiple faults by one model. New/unlearned faults were learned by label correlation.

Recent research in this area has also sought to combine the assessment of rotor and short-term voltage stability, using either ensemble RF DTs [69] or graph attention networks (GAN) [82]. This has interesting real-world applications as, depending on the characteristics of the system and fault, it can be difficult to distinguish between dominant stability modes in certain cases.

The ML algorithms used for rotor angle stability assessment are included in Table 4. A comparison with ML algorithms used for other types of DSA is included Table 8.

B. LONG-TERM VOLTAGE STABILITY

Long-term voltage stability (LTVS) is dictated by equipment with slow response times, such as thermostatic loads, or network control devices, such as tap-changing transformers. Based on the response times of such equipment, and the subsequent perturbation of the system operating point, the analysis interval can be up to tens of minutes.

1) STABILITY CRITERIA

Early work in [86] utilized the energy method, which is closed form vector integration of the real and reactive

power mismatch equations between a stable and unstable operating point, to provide a quantitative measure of system vulnerability to voltage instability. A minimum energy margin is used to label the states.

More generally, LTVS is measured by the distance between the operating point (OP) and the voltage collapse point (VCP). This is typically illustrated using the P-V curve, which is calculated using the continuation power method. When the OP moves to a point below the VCP, voltage collapse is likely to occur and is quantified by the voltage stability margin (VSM):

$$VSM = \frac{P_{max} - P_{op}}{P_{max}} \tag{40}$$

where P_{max} is the maximum deliverable power, and P_{op} is the load demand at the current OP.

This scalar value can be used directly to estimate voltage stability margins, i.e., as regression problems in, e.g., [87], [88], [89], [90], [91], [92], [93], and [94], or be converted to class labels. Class labels have been defined with respect to the percentiles [95], three-class labels using quartiles [96], or binary classification using a predefined voltage stability criterion [97] (e.g., using WECC criteria: 7% for normal conditions, 5% for contingency [98]). Additional voltage stability indices are compared in [99]. However, their use is not widespread.

2) INPUT DATA AND FEATURE EXTRACTION

The vast majority of data utilized in LTVS assessment is readily available from PMUs (assuming they are present in the system) and associated with busbar quantities. A small number of algorithms also consider branch quantities in the assessment procedure. The identified input data utilized in the reviewed manuscripts is summarised in Table 5.

Several different feature selection procedures for LTVS have been considered. Early work in this area in [86] considered a ''sensitivity analysis'' using the inversion of the Jacobian matrices of the power flow equations to identify important input features. Clearly, this requires considerable knowledge of the system and is subject to obvious limitations in terms of computation and uncertainty. Since this initial exploration, the most often applied techniques found in the literature are PCA and a heuristic approach.

In [97] eight different groups of input features were evaluated and it was found that current magnitude and fault location were the two most important features. Four different feature sets were compared in [97] and it was found that voltage magnitudes and angles are the most appropriate features; this has physical meaning as phase angle is a good predictor of power flow and voltage is a good predictor of proximity to voltage collapse. A similar conclusion was reached in [88].

PCA was used in [87] and reduced the number of input features from 241 to 97 for the IEEE 118-bus network. PCA was also used in [93] and [100], however, the performance was not explicitly stated. In [89], PCA was used in combination with a multi-resolution wavelet method for feature selection, which could be applied to voltage profile (i.e., time series) data.

3) ML ALGORITHMS

An early application of ML methods to LTVS assessment was presented in [86]. This work applied a DNN to map the system operating condition to the VSM using the energy method to determine stable equilibrium points and unstable equilibrium points. An ANN (feedforward, back propagation) was also utilized in [87]. Interestingly this paper tried, unsuccessfully, to train the DNN to adapt to the network configuration.

However, this problem was addressed in [95], which proposed a transfer learning method to account for variations in time and network conditions. The classification was performed using a DNN model which was shown to function well with a limited number of labeled samples.

As can be expected when using data-driven methods for online applications, there has been significant attention devoted to the model training and the model training time. Reference [97] tackled the issue of training by periodically updating a DT model on an hourly basis. The implementation of a more sophisticated framework using an ensemble of ELMs to reduce model training time was proposed in [90]. An alternative idea was presented in [91], which developed a genetic algorithm optimization method to find the optimal parameters of the SVM, thus minimizing training time.

Model training was also the focus of [96], which applied pool-based active learning to select the most representative operating points to be added to a continuously updated training set. A number of ML algorithms (ANN, SVM, and RF) were compared for the three-state classifier, with RF found to be the most accurate on average. Interestingly, a similar conclusion was proffered in [99], although it should be clearly stated that the performance of any model depends on a multitude of factors, covering the training, implementation, and testing phases. However, RFs were also shown to perform well for LTVS assessment in [98], which proposed an enhanced RF approach with online bagging to create an online learning framework to minimize offline training time.

Another important aspect relating to the input data was presented in [94], which combined a Kernel ELM with a Mean-variance Mapping Optimization algorithm, to address the problem of noise in the input data. This was demonstrated in a case study assuming noise with a zero-mean normal distribution superimposed on the ideal voltage data (i.e., additive white Gaussian noise) to replicate possible issues in real-world data capture systems (e.g., PMUs). The optimal location of PMUs for LTVS assessment was addressed in [88] by combining an SFS feature selection method and an ANN classifier to determine the optimal input feature set consisting of phasor voltage data. The optimal feature set is then interpreted as the optimal location of PMUs.

TABLE 5. Long-term voltage stability assessment input data and machine learning (ML) algorithms.

Ref	ML Algorithm	Gen Bus			Other Bus						Branch			Fault location	Addressed Problem	Output (*No. states*)
		P	Q	Q_{res}	P	Q	V_{mag}	V_{phase}	I	I_{phase}	P	S	Z			
[86]	FNN	✓	✓	-	✓	✓	-	-	-	-	-	-	-	-	Regression	VSM
[87]	FNN	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	Regression	VSM
[97]	DT	-	-	-	-	-	✓	✓	✓	✓	-	✓	✓	✓	Classification	(2)
[88]	ANN (MLP)	-	-	-	✓	✓	✓	✓	-	-	-		-	-	Regression	VSM
[89]	FNN (Radial Basis)	-	-	-	-	-	✓	-	-	-	-	-	-	-	Regression	VSM
[90]	ELM	✓	✓	-	✓	✓	✓	✓	-	-	-	-	-	-	Regression	VSM
[91]	SVM	-	-	-	-	-	✓	✓	-	-	-	-	-	-	Regression	VSM
[92]	CNN	-	-	✓	-	-	✓	-	-	-	✓	-	-	-	Regression	VSM
[96]	Active Machine Learning (SVM, ANN, DT, RF)	-	-	-	-	-	✓	✓	✓	✓	-	-	-	-	Classification	(*3*)
[93]	Multi-Linear Regression	-	-	✓	-	-	✓	-	-	-	✓	-	-	-	Regression	VSM
[98]	Enhanced RF	-	-	-	-	-	✓	✓	-	-	-	-	-	-	Classification	(*2*)
[94]	Extreme Learning	-	-	-	-	-	✓	-	-	-	-	-	-	-	Regression	VSM
[100]	FNN	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-	Classification	(*2*)
[95]	Tri-net Deep Neural Network	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-	Classification	(*2*)

Where the output of regression problems is the index utilized and the output of classification problems is given by the number of states (N)

In [92], CNN was to classify the LTVS problem into i states, whilst also outputting the probability of each state. The probability of each state was also considered in the three-state classification in [96]. Such probabilistic approaches can support confidence in the decision-making process. An alternative approach, using a multi-linear regression model, which outputs both the VSM and its associated confidence interval was presented in [93].

Finally, a method with interesting real-world application was proposed in [100]. Using an ANN (feedforward, back-propagation), the power system voltage stability conditions were mapped to control actions (for load tap changers, capacitor banks, and load shedding) to remedy any stability issues. In this case, the output it not the stability margin of the system but a priority list of control actions.

The ML algorithms used for LTVS assessment are included in Table 5. A comparison with ML algorithms used for other types of DSA is included Table 8.

C. SHORT-TERM VOLTAGE STABILITY

Short-term voltage stability (STVS) is driven by the response of power system components that exhibit fast dynamics. This covers traditional components such as induction motors, but also modern power electronic interfaced loads and generators, as well as High-Voltage Direct Current (HVDC) links. The first few seconds of the power system response are used to define the STVS, and dynamic models with a high level of detail are required. Accordingly, it is not feasible to assess online using traditionaly approaches. A broad review of the subject, also covering conventional assessment methods, is available in [101].

1) STABILITY CRITERIA

Unlike rotor (transient) stability and LTVS, many different approaches are presented in the literature for the assessment of STVS.

They can generally be defined by magnitude and duration, using the logic that a system is classified as unstable if the magnitude of any bus voltage exceeds a threshold value for a certain amount of time [102], [103], [104], [105], [106], [107].

In [108], the maximal Lyapunov exponent (MLE) was used. The MLE originates from the ergodic theory and can be evaluated using the following equations:

$$|\Delta V(t)| = e^{\lambda t}|\Delta V_0| \tag{41}$$

where ΔV_0 is the initial voltage deviation, $\Delta V(t)$ is time-dependent deviation of the voltage trajectories and λ is the *Lyapunov exponent* given as:

$$\lambda = \frac{1}{k\Delta t}\ln\left(\left|\frac{V'((n_0+k)\delta T))}{V'(n_0\Delta t)}\right|\right) \tag{42}$$

where k is the size of the data window, Δt is the sampling interval, n_0 is the first data point index, and V' is the time derivative of the voltage at the specific data point

If the *Lyapunov exponent* is positive at any bus, the system is classified as unstable.

The Transient Voltage Severity Index was used in [109]:

$$TVSI = \frac{\sum_{i=1}^{N}\sum_{t=T_c}^{T} TVDI_i, t}{N \times (T - T_c)} \tag{43}$$

where N is the total number of buses in the system, T is the considered transient time frame, T_c is the fault clearing time, and $TVDI$ is the transient voltage deviation index, calculated by:

$$TVDI_{i,t} = \begin{cases} |\frac{V_{i,t}-V_{i,0}}{V_{i,0}}|, & \text{if } |\frac{|V_{i,t}-V_{i,0}}{V_{i,0}}| \geq \mu \\ 0, & \text{otherwise} \end{cases} \tag{44}$$

where $V_{i,t}$ denotes the voltage magnitude of bus i at time t, and μ is the threshold to define unacceptable voltage deviation level, which can be set according to the industrial criteria, e.g., 20 %. As such, TVSI is a regression index.

Reference [110] proposed a new index called the *root-mean-squared voltage-dip severity index* (RVSI), specifically to evaluate fault-induced delayed voltage recovery severity,

in which some bus voltages undergo a slow recovery that may trigger undervoltage load shedding protection.

$$RVSI = \sqrt{\frac{\sum_{i=1}^{N_b} VSI_i^2}{N_b}} \tag{45}$$

where N_b is the number of buses in the system, and VSI is the voltage-dip severity index which quantifies the voltage recovery. VSI is calculated as:

$$VSI_i = \int_{t_0}^{T} D_{i,t} dt \tag{46}$$

where:

$$D_{i,t} = \begin{cases} |\frac{V_{i,0} - V_{i,t}}{V_{i,0}}|, & \text{if } V_{i,t} \leq \mu V_{i,0} \\ 0, & otherwise \end{cases} \tag{47}$$

t_0 is the time when the fault is cleared, T is the analysis window, $V_{i,t}$ denotes the voltage magnitude of bus i at time t, and $V_{i,0}$ is the prefault voltage magnitude of bus i; and μ is a threshold to define critical voltage-dip magnitude. A value of 90% is used in [110].

2) INPUT DATA AND FEATURE EXTRACTION

In the case of STVS assessment, all of the information utilized is readily available from PMUs (assuming they are present in the system) and associated with busbar quantities. These are summarised in Table 6.

In addition to established feature selection methods, e.g., Relief in [109] and Relief-F in [110], several interesting methods have been developed with respect to the particular characteristics of STVS problems. The time series nature of the voltage response was treated using Time Series Shapelet Classification (TSCC) to extract features from time series data in [102] and [111]. It was argued that this approach provides more interpretable results than other feature selection methods.

An alternative method for feature selection of time series was proposed in [112], which implemented a Symbolic Aggregate approXimation method to transform long time series data into symbolic words as a wordbook. This is more compact and efficient to handle than raw time series data. Similarly, in [108], a symbolic discretization technique was used to transform time series data into symbols to allow for the representation of their temporal relationships in multivariate analysis.

Temporal features were combined with spatial features by combining geospatial information, electrical distances, and time series data with the spatial voltage interpolation method in [114]. Critical features were then extracted from these time series sequences by TSSC.

3) ML ALGORITHMS

Some early work in this area treated STVS assessment as both a classification and a regression problem using a hierarchical system [109]. The hierarchical system was based on Neural Networks with Random Weights with ensemble learning. In the first instance, the STVS assessment is considered a classification task. If the system is deemed stable then the assessment is considered a regression task in the second instance, with the objective of assessing the severity of the dynamic voltage deviation using (43) and (44).

Hierarchical approaches were also presented in [108] and [110], which both considered the classification of the fast voltage collapse but also fault-induced delayed voltage recovery events. In [108], a multi-level (four-state) RF-based classifier was developed, while the framework in [108] consisted of an ensemble of ELMs.

The use of DTs for this task is evident in [102] and [111]. It should be noted that both papers are more focused on the application of the TSSC feature extraction method, and it is acknowledged that other classifiers could have been employed. However, DTs were selected over ANN and SVM as they were deemed easier to interpret. The extension provided in [111] is the detailed consideration of the imbalance learning (class skewness) problem in the context of short-term voltage stability assessment. To this end, a forecasting-based nonlinear synthetic minority oversampling technique was employed in conjunction with cost-sensitive learning to weight the scarce unstable samples. This approach is able to produce new synthetic samples by interpolating between existing samples. Another application of DTs is found in [112] by the same authors, this time using DTs as the classifier in combination with the Symbolic Aggregate approXimation feature selection method.

An interesting application was presented in [103], based on the least squares SVM classifier. However, this technique requires detailed knowledge of load response, i.e., the load composition. For a practical system, the load composition is complex and exhibits spatial and temporal variations, and it may be difficult to obtain accurate information on composite load dynamic models.

The issue of temporal variations was considered in [104], which developed a temporal-adaptive decision-making scheme in which the post-fault response was divided into a number of discrete cycles of varying lengths. The logic behind this approach is to allow the classification assessment decision to be returned as soon as possible, i.e., as a trade-off between cycle length and classification accuracy. This utilized an ensemble of ELM and RVFL to create a so-called hybrid randomized ensemble model.

LSTM models have also been applied for STVS assessment. Reference [106] combined a semi-supervised constraint-partitioning k-means clustering algorithm to classify the system as stable or unstable with a LSTM model to learn the time dependencies between the input and output data.

LSTM models have also been adapted to learn spatial-temporal dependencies in STVS data. In [107], spatial dependencies were incorporated by formulating the inputs to the LSTM as spatial attenuation factors. The scalar output was then converted into a binary classification label. In [115], a LSTM model was combined with a

TABLE 6. Short-term voltage stability assessment input data and machine learning (ML) algorithms.

Ref.	ML Algorithm	Gen Bus P	Gen Bus Q	Other Bus P	Other Bus Q	Other Bus V_{mag}	Other Bus V_{phase}	Other Bus I_{mag}	Addressed Problem	Output (*No. of states*)
[109]	Deep belief network	✓	✓	✓	✓	✓	✓	-	Classification	(*2*)
[102]	DT	-	-	✓	✓	✓	-	✓	Classification	(*2*)
[111]	DT	-	-	✓	✓	✓	-	✓	Classification	(*2*)
[112]	DT	-	-	✓	✓	✓	-	-	Classification	(*2*)
[103]	SVM	-	-	-	-	-	-	-	Regression	Predicted Trajectory
[108]	RF	-	-	-	-	✓	-	-	Classification	(*4*)
[110]	Ensemble (ELM)	-	-	-	-	✓	-	-	Classification	(*4*)
[113]	Ensemble (RVFL)	-	-	-	-	✓	✓	-	Classification	(*2*)
[104]	Ensemble (ELM and RVFL)	-	-	-	-	✓	✓	-	Classification	(*2*)
[105]	Graph convolutional network	-	-	✓	✓	✓	-	-	Classification	(*2*)
[106]	LSTM	-	-	✓	✓	✓	-	-	Classification	(*2*)
[107]	LSTM	-	-	✓	✓	✓	-	-	Classification	(*2*)

Where the output of regression problems is the index utilized and the output of classification problems is given by the number of states (*N*)

GCN to consider both spatial and temporal dependency in the system response. While, in [105], a GCN model was developed to handle both time and spatial dependencies by adopting the one-dimensional convolutional layer to enable the extraction of spatiotemporal features from the hidden states. This method also provides information on the influence of individual buses on the system response.

Finally, the issue of missing data was addressed in [113] by grouping system buses to maintain grid observability using a structure-adaptive ensemble learning model. The ensemble consisted of a set RVFLs, while the system bus clustering algorithm was originally presented in previous work by the authors in [116]. Using the IEEE New England 39-bus network the accuracy was shown to be greater than 90% even with 90% missing data.

The ML algorithms used for STVS assessment are included in Table 6. A comparison with ML algorithms used for other types of DSA is included Table 8.

D. FREQUENCY STABILITY

Frequency stability is the ability of the system to keep a steady frequency during disturbances, where the imbalance between mechanical (which drives the generated electrical power) and electromechanical (driven by the requested electrical power) torque lead to acceleration/deceleration of synchronous generators in the presence of positive/negative imbalance. Since the frequency value is severely restricted, excessive oscillations can trip frequency relays leading to loads and inverter-based power generation disconnection.

1) STABILITY CRITERIA

Different assessment approaches and metrics to quantify frequency stability have been considered in ML techniques. A common and computationally efficient approach to frequency stability assessment is to compare the frequency value using the frequency nadir, i.e., against a minimum threshold value [117], [118]. An upper-frequency limit can naturally also be considered as an important indicator of stability [119].

However, it is also possible to consider the rate of change of frequency (RoCoF), i.e., the time derivative of the power system frequency (df/dt), often in combination with the frequency minimum/maximum values [119], [120], [121], [122].

Further metrics can be included in the analysis, for example, also considering the mean square deviation (from nominal frequency) and integral alongside the nadir and RoCoF [123].

The frequency stability margin (FSM) presented in [124] also introduces the concept of time spent below the nadir value:

$$FSM = (F_{min,i} - F_{cr,i} - k \cdot T_{cr,i}) \times 100\%, \in [-100, 100] \tag{48}$$

where $F_{min,i}$ is the minimum frequency, $F_{cr,i}$ is the critical frequency, and $T_{cr,i}$ is the time that the frequency is below the critical frequency of the bus i during the transient period. k is a user-definable factor. When FSM is greater than zero, the system is stable.

Some early work in this area [125] utilized the generation deficiency and frequency decline (GD/FD) ratio:

$$\frac{GD}{FD} = \frac{\frac{\partial P}{P_{sys}}}{\partial f} \approx \frac{\frac{\Delta P}{P_{sys}}}{\Delta f} \left(\frac{1}{Hz}\right) \tag{49}$$

where ΔP is the amount of power lost (MW) and Δf is the initial decline of frequency (Hz).

2) INPUT DATA AND FEATURE EXTRACTION

The type and number of input data used for data-driven frequency stability assessment can vary considerably from one approach to another. A summary of the input data used is provided in Table 7.

The relative infancy of this research area is evident when compared with the more homogeneous data utilized in other types of DSA. This also reflects the need to better understand this complex system response to lead to better feature selection. Several of the required data are not trivial to ascertain and/or calculate, e.g., load damping coefficients and generator governor response times. One approach also considers the output of a continuously

calculated model-driven system frequency response (SFR) value as input to the prediction [126]. Furthermore, the output of specific generation types and/or reserve capacity is often requested, suggesting that the trained models are very sensitive to the specific network configuration and operating conditions.

Unlike the other types of DSA, the feature selection algorithms discussed in Section III-B have not been widely considered in frequency stability assessment methods. Rather, feature selection is directly handled using deep learning ML algorithms.

3) ML ALGORITHMS

Early work in this area was based on regression trees and focused on the system frequency response to a specific event, i.e.a generator outage [125]. The GD/FD (49) was used and returned an error of 8-9% when applied to events simulated using the Taipower system. In the same year, frequency nadir and RoCoF were combined in a DT binary classification problem in [122].

Early examples of FNN approaches for frequency stability assessment can be found in [120] and [124]. In [124], an extreme learning machine-based predictor for the real-time frequency stability assessment based on the FSM (48) was developed. The frequency stability of isolated island power systems with significant amounts of renewable generation was the subject addressed in [120]. This two-stage approach based on FNN and backpropagation algorithm considered the frequency nadir as the first indicator of frequency stability, with the RoCoF evaluated when the frequency nadir falls a defined threshold value. An additional feature of the developed approach was the output of the optimal load-shedding response to maintain stability.

Another work that also considers subsequent control actions to alleviate any forecast stability issues is found in [127]. In this hybrid approach, the SFR is determined by a model-driven methodology and then ELM is used to fit and correct errors in the model-driven analysis for frequency prediction and control.

ELM models were also applied in the learning methodology proposed in [121], which combined DNNs and a stacked ELM for fast online transient frequency stability assessment. The DNN was used to obtain the network parameters using input-output feature data, while the stacked ELM was used to further reduce the prediction error. Several metrics can be simultaneously predicted by the developed methodology, including the maximum RoCoF, frequency nadir, time to reach frequency nadir and quasi-steady-state frequency deviation.

As is the case with other types of DSA, the incorporation of time series features has also been shown to be important for frequency stability assessment. As such, this topic has been addressed by methodologies proposed in recent research. Reference [128] used an LSTM model to predict the time series frequency response using historical grid data of disturbances recorded over a three-year period. However, data of the type of disturbances analyzed was not discussed in detail in the paper. Reference [126] also used an LSTM model, here in combination with a physical-based model of the system frequency response, to predict the system state. An alternative approach was developed in [117], which considered the impact of uncertainty in wind generation on frequency stability using a physics-guided gated recurrent unit (PG-GRU) neural network as the foundation of the frequency stability assessment. The PG-GRU was able to handle time series data and performs well even when the training set is small.

Time-spatial dependencies were the subject of [118]. This work proposed a deep learning model combining CNN and LSTM models to account for the spatial and temporal dependencies in power system operation. A transfer learning process was also implemented to overcome the challenge of insufficient data and changes in power system operation conditions. Furthermore, an optimal load-shedding strategy was implemented to avoid frequency collapse if necessary.

An interesting, and important, case study considering an interconnected AC-DC was presented in [119]. The online frequency security assessment was built on the Deep Belief Network (DBN) method and applied to a large power grid model comprised of northern and central China. The performance of data-driven DSA in assessing such systems will become increasingly important in future power systems.

Although not directly related to frequency stability in response to a disturbance, [123] is an interesting work. It was devoted to the interpretability of machine learning model outputs using SHapley Additive exPlanation (SHAP) values to analyze the impact of system conditions on frequency stability. Shapely values represent the effect of individual features on model outputs and were applied to analyze the impact of generator ramping conditions on the maximum frequency deviation, rate of change of frequency, mean square deviation, and the integral of the frequency of three European synchronous areas (Continental Europe, Great Britain, and the Nordic area).

The ML algorithms used for frequency stability assessment are included in Table 7. A comparison with ML algorithms used for other types of DSA is included Table 8.

VI. RESEARCH GAPS AND MOST PROMISING RESEARCH TRENDS

A. NEW FORMS OF STABILITY

As shown in Fig. 5, it has recently been proposed to extend the classification of DSA to include *converter-driven stability* and *resonance stability* to account for the growing penetration of converter interfaced equipment, particularly generation, and their impact on power dynamics [7].

The response of inverter based resources (IBR) to small and large disturbances is determined by the specific control algorithms utilized within the converter. However, it can be expected that the response is sufficiently different

TABLE 7. Frequency stability assessment input data and machine learning (ML) algorithms.

Ref.	ML Algorithm	Inertia	Gen reserves	N_{units}	P_{elec}	P_{mech}	Damping	Load P	V_{mag}	V_{phase}	Other	Addressed Problem	Output Type (*No. of States*)
[117]	ANN (physics-guided gated recurrent unit)	✓	-	-	✓	✓	✓	✓	-	-	-	Classification	(*2*)
[126]	LSTM	-	-	-	✓	-	-	✓	✓	✓	-	Regression	Frequency response model
[120]	FNN	✓	✓	✓	✓	-	-	✓	-	-	-	Regression	Min frequency frequency variation
[118]	CNN+LSTM	-	-	-	✓	-	-	✓	✓	✓	-	Regression	System Response Trajectory
[121]	Two-stage (DNN and stacked ELM)	✓	✓	-	-	-	✓	-	-	-	-	Regression	Generator response times
[119]	Deep belief network	✓	✓	-	✓	-	✓	✓	-	-	DC capacity, interconnection capacity	Prediction	ROCOF
[122]	Probabilistic ANN	-	✓	-	✓	-	-	✓	-	-	Specific generation mix	Classification	(*2*)

Where the output of regression problems is the index utilized and the output of classification problems is given by the number of states (*N*)

TABLE 8. Machine learning (ML) algorithms applied to different areas of dynamic stability assessment.

ML Class	ML algorithm	RS	LTVS	STVS	FS
ANN	FNN	[84]	[86]–[89], [96], [99], [100]	[109]	[120]
ANN	FFNN	[65]			
ANN	Deep belief network	[71]			[119]
ANN	Ensemble ELM			[110]	
ANN	Random vector functional link (Ensemble)			[113]	
ANN	Graph convolutional network			[105]	
ANN	LSTM			[106], [107]	[126], [128]
ANN	ELM	[80]	[90], [94]		[124], [127]
ANN	CNN	[73], [78], [79]	[92]		
ANN	Tri-Net FNN		[95]		
ANN	Probabilistic FNN				[122]
ANN	Ensemble (CNN,LSTM)				[118]
ANN	Physics-guided GRU NN				[117]
ANN	GCN+LSTM	[74]		[115]	
ANN	Graph attention network	[82]			
ANN	Generative adversarial network	[85]			
Tree-based method	DT	[66]–[70]	[97]	[102], [111], [112], [114]	[125]
Tree-based method	RF		[96], [98], [99]	[108]	
Support vector machine		[67]	[75], [83]	[91], [96]	[103]
Hybrid Ensemble	ELM and RVFL	[81]		[104]	
Hybrid Ensemble	CNN + SVM	[77]			
Two-stage	DNN and stacked ELM				[121]
Other	Multi-label learning	[72]			
Other	LASSO		[93]		
Other	Discernibility Matrix	[67]			

Where: RS = rotor stability; LTVS = long term voltage stability; STVS = short term voltage stability; FS = frequency stability.

from the response of traditional synchronous generators to warrant particular consideration in DSA in the future. This phenomenon is defined as *converter-driven stability*. This is divided into *fast* and *slow* interactions to account for the different response times of the different control circuits/loops present in individual equipment. It has been reported that these will interact with the system and neighboring equipment resulting in oscillations in the kHz range [129].

Resonance stability is the term applied to all cover all sub-synchronous resonances. This can occur between the series compensation present in the system and either the electrical impedance of the generator or the mechanical torsional frequencies of the turbine-generator shaft (defined as electromechanical resonance stability). Clearly, the impact of electromechanical oscillations is manifest as a threat to mechanical structure integrity, while electrical oscillations could lead to large currents and voltages (which may damage equipment) or changes in electrical torque. Further details of the phenomena and its impact can be found in the recent review [130], with a review of specific wind turbine considerations available in [131].

These phenomena are typically faster than traditional power system dynamics, with inverter-based control operating in the microsecond region. Furthermore, these cover a wide range of response times, covering ''wave'', ''electromagnetic'' and ''electromechanical'' phenomena at times scales from 10^{-7} to 10^{1} s. Therefore, to include these emerging aspects in online SA, ML models will require very short computation times, also accounting for the data transfer speed from measurement devices. An interesting aspect to address is the possible need of assessing multiple time scales to determine the stability status of the system. However, before arriving at this point, a better understanding of the phenomena is required in order to develop sufficiently accurate models, and thus generate suitable training data. A particularly important point will be in the development

of classification or regression-based indices, as these issues can be particularly sensitive to local network conditions. The application of digital twin models may be useful for supporting both online analysis and also in the generation of training data [132]. Quantum computing technology, discussed further in Section VI-H, is another emerging area that may support these aspects.

B. NETWORK CONFIGURATION

Network configuration plays an important role as system operators seek to incorporate increasing numbers of low-carbon technologies, whilst simultaneously maximizing the utilization of existing assets. Data-driven models are inherently trained for a given topology but it is important to consider their applicability when applied to topology changes that can occur during fault conditions.

Considering multiple topology variations and developing a model for each scenario is one option. However, this is likely to result in a combinatorial explosion as the number of scenarios increases, especially when considering the impact of embedded distributed resources on transmission systems. A more attractive, and methodologically robust, approach would be to apply transfer learning techniques, e.g., [73], [81], [95], and [118].

In the case of considering the dynamic response of distribution networks on transmission system stability, the distribution network becomes an Active Distribution Network Cell (ADNC) [133] or aggregated dynamic equivalent model, in which the aggregated response is modeled as a single component (discussed further in Section VI-C).

Another important aspect of network configuration is the operation of microgrids. Microgrids can operate in grid-connected and/or islanded mode and can be formed at either low or medium voltage levels [134]. As they are characterized by smaller groups of generation and load, the dynamic characteristics are considerably different from larger, interconnected systems, and exhibit high uncertainty [135]. Furthermore, the transition from grid-to-islanded operation and islanded-to-grid operation may result in some controllers changing from voltage control to current control mode [136] (i.e., grid-forming to grid-following in the context of IBR [137]). Stability issues associated with converter controllers, e.g., as discussed in Section VI-A, may be of increasing interest in the SA of such systems. A comprehensive review of microgrid stability issues is available in [138] while grid-forming and grid-following stability aspects are addressed in detail in [139].

C. COMPONENT MODELS

Naturally, the overwhelming majority of data utilized in the training of ML models for power system SA comes from simulations. To account for the ongoing massive integration of converter interfaced generation (and equipment in general), accurate models of these components, and their interaction with the system and neighboring equipment are vital.

The models used to represent such equipment in stability analysis were traditionally based on either detailed switching models (i.e., time domain circuit models, equivalent or otherwise), state-space equation analysis, or impedance-based model techniques [140]. However, machine learning models can themselves be trained to model the dynamic response of the system or system components, with early work in this area demonstrating the use of ANN to develop dynamic load models [141], [142], [143]. Such applications appear to have value to reduce the model complexity and simulation time of individual components, and the overall system.

Although a recent review of HVDC modeling for power system stability assessment [144] made no mention of ML model applications, one example has recently been presented in [145]. Furthermore, the use of data-driven modeling techniques and machine learning methods for the development of dynamic models of microgrids is advocated in [146]. Applying ML methods to model ADNC has received considerable attention: [147] used an unsupervised learning method for feature selection, [148] used ANN to estimate the parameters of a generic measurement-based equivalent model (interestingly, grid state parameters, i.e., load and generation mix, form part of the model development, and, hence, implementation), while [149] developed an LSTM RNN to directly model, i.e. a black box model of the ADNC.

D. INPUT FEATURES FOR GRIDS WITH HIGH PENETRATION OF RENEWABLE ENERGY

A high penetration of renewable energy can compromise system reliability. For example, under favorable weather conditions, installed wind and PV photovoltaic or wind power generator systems can cause overgeneration, leading to power curtailment to prevent the violation of transmission line thermal limits [150]. Overgeneration side-effects are exacerbated by the presence of low-meshed grids, where the contingency of one line can overload neighboring lines, risking the intervention of protection systems and potentially causing large cascading events that compromise system operation, leading to blackouts [151]. A significant portion of low-inertia renewable power plants can also compromise system stability, violating dynamic security requirements. Therefore, accurately assessing system security in the presence of high shares of renewable energies is crucial. However, few studies address the challenges posed by a large share of renewable power generators. Some examples include [152], which attempts to generate and cluster a high number of scenarios rapidly due to the variability of non-programmable renewable power profiles for DSA, and [117], which incorporates wind power data to consider wind power uncertainty in frequency stability assessment. Time-series analysis is essential for understanding the temporal variation of wind power generation over multiple time

scales to analyze system stability, as demonstrated by [153] in proposing a fast-scanning method for the current stability analysis of the system. However, the use of weather data in data-driven SSA/DSA tasks is still rare. For instance, [59] proposed considering weather prediction to assess security margins. The underutilization of these features is due to the definition of power system security, which depends on the operating state, given contingencies, and generator types, rather than weather variables. Nevertheless, weather-related variables could play a crucial role in selecting the most critical contingencies or deploying severe event forecasters.

E. THE ROLE OF SPATIO-TEMPORAL FEATURES IN DATA-DRIVEN-BASED SSA/DSA METHODS

The advances in ML, particularly in its ability to effectively interpret spatio-temporal dependencies between variables, have garnered significant attention from researchers in enhancing data-driven SSA/DSA tools. However, it is important to note that spatial features are typically already considered as input data. These spatial features include power generation and demand, voltage phase and magnitude, as well as line power flow profiles, all of which inherently provide spatial information.

In data-driven-based SSA, there is a growing trend toward deploying algorithms that are better equipped to capture spatial information, such as CNN. However, the utilization of temporal features is still relatively rare. An exception is the work presented in [59], where past system states and weather data were employed to forecast the future secure/insecure state of power system areas.

Spatio-temporal features hold significant promise, yet only a handful of SSA works have explored this aspect. One example is the modeling of spatio-temporal dependencies between wind power and load, which uses copulas to probabilistically assess the overloading of transmission lines, as proposed by [154].

On the other hand, in DSA, temporal relationships come into play when dealing with time series data. For instance, works like [103], [106], and [107] focus on STVS assessment. References [103] and [106] employ LSTM networks, specifically designed to capture temporal dependencies in time series data, whereas another approach is presented in [107], which uses a Recurrent Neural Network (RNN) to process both static network information and dynamic system responses.

Innovative research, exemplified by [155], introduces DeepONet-grid-UQ, an advanced neural network, which predicts post-fault trajectories of a system in the presence of a fault by taking on-fault trajectories collected by sensor units distributed across the grids as input data.

F. IMPACT OF STABILITY CRITERIA ON THE PERFORMANCE OF DATA-DRIVEN-BASED SSA/DSA METHODS

This section summarizes the impact of stability criteria on the accuracy of data-driven SA. It is essential to emphasize that stability criteria, in any form they are presented, are established independently of the type of ML workflow.

Particularly, as discussed in Section III, any ML algorithm learns from an input-output map, where the output is represented by a status or an index score. Hence, the potential impact on accuracy and reliability depends on the type of problem being addressed, whether it is classification or regression.

Specifically, if the stability criteria are challenging to violate, there will be few instances where the algorithm can learn an insecure state within the training set. This situation is well-documented in the literature as classification in the presence of imbalanced data sets [152].

Other challenges arise from the data splitting performed by several classification algorithms to partition input data based on labels. Nonlinear relationships are more challenging to learn and may require precise feature transformations using techniques such as kernel tricks or tools like PCA or independent component analysis [4]. Recently, more advanced feature selection methods have been proposed to enhance the effectiveness of selecting features in the presence of highly imbalanced datasets [156].

Similarly, a low number of samples falling within a certain index score range can result in reduced prediction accuracy when tackling regression problems.

G. EMERGING APPLICATIONS

SA primarily functions as a tool aimed at scrutinizing system behavior in response to contingencies, often at the transmission grid level. In essence, it delves into the potential outcomes when one or more elements are removed from the system, regardless of the cause. Considering a large list of contingencies, especially for DSA may lead to costly and timely analyses. Data-driven methods can reduce the high costs and time of conventional techniques.

Further application of ability of data-driven models to select the most credible contingencies [157] should be encouraged. Also, risk-based SA can lead to forecasting of potential outages [158]. Clearly, in such cases, the forecasting horizon must align with the scheduling timing of the SA process. Contingency forecasting could be strategic to anticipate critical conditions in power systems increasing system security, some examples are wind power ramps classification/forecasting [159] and voltage excursion prediction [160]

Furthermore, data-driven methods in power systems could be related to distribution systems, such as fault prediction and location, grid topology reconstruction, and other tasks where available data are unknown or scarce [161]. Several factors contribute to this scarcity, ranging from uncertainties associated with demand and distributed/embedded energy resource profiles to the sparse or entirely absent presence of measurement sensors. Nevertheless, the advent of Smart Grid technologies holds the potential to empower Distribution System Operators, enabling them to conduct these

analyses at the distribution level. Some works such as [162] and [163] extended the SSA and the transmission congestion management to the distribution grids, respectively.

Finally, emerging algorithms such as multi-label classification algorithms [164] could be employed when, for example, labelling multiple instability occurrences, for the given scenarios. Federated learning, which trains a number of local models using local (i.e., not shared) datasets and then shares the local model parameters to create a global model, has particular advantages when considering data privacy and robustness in the presence of the growing number of cyber threats; further investigation into using this approach for SA is recommended. In general, system vulnerability analysis, although beyond the scope of this article, will form an increasingly important part of SA in order to identify security weaknesses before they can be exploited and ensure that the system continues to operate in the face of a range of conditions, including malicious attacks.

H. QUANTUM COMPUTING APPLIED TO SSA/DSA

Quantum Computing (QC) is a novel computational paradigm based on the principles of quantum mechanics. Its advantages compared to classical computing include high-speed complex problem-solving, massive parallelism, and the application of complex operators such as entanglement. However, QC is currently limited by the lack of specific algorithms, commercial quantum-based processors, and the need to deploy algorithms to reduce quantum errors. Particularly, the high cost and complexity of the technology (i.e. advanced cooling systems to reduce the information losses in qubits) could inhibit the large spread of this calculation paradigm in the near future. Presently, there are only limited examples that utilize QC for power system security assessment, e.g., [165]. However, applications to solve static and dynamic power system problems such as power flow study [166], [167] and electromagnetic transient modeling [168] show promising results. Based on these findings, it could be reasonable to imagine that QC, due to its tremendous computational performance, could significantly reduce the time necessary to perform physics-based SSA/DSA. These reduced computation times could be exploited by data-driven SSA/DSA to allow for the generation of input data closer to real time, i.e., based on actual network operating conditions, and the analysis of faster stability phenomena, e.g., converter-driven stability and resonance stability.

I. OPEN ISSUES

From this review, the following key challenges associated with data-driven methods for power system SA emerge:

i The predominant use of synthetic data to train models. Such data are unaffected by any measurement errors which may be present in real world systems, e.g., SCADA and PMU data, and the impact of errors/missing data should be more deeply considered;
ii The use of aggregated data, which may be strategic when system SAs are used to validate market clearing results (i.e., where generation and demand profiles are provided for a given area rather than each busbar), has not been widely explored;
iii The default usage of a decision threshold of 0.5 in classification algorithms, with scarce information provided about the chosen thresholds. Tuning this threshold could better align system state predictions with risk thresholds;
iv The absence of investigation into the use of asymmetric indexes. For instance, not predicting an insecure state (false negative) is more critical than a false positive (predicting a secure state as insecure);
v The need for research on the most interpretable input features, whose variations have the greatest impact on the system state, and further exploration of the use advanced feature selection techniques, e.g., autoencoders and causality-based feature selection;
vi The necessity to increase industry-oriented applications, as the majority of developed algorithms have been validated in offline environments.

VII. CONCLUSION

Research into machine learning (ML)-based methodologies for the important task of power system security assessment (SA) is continuing to increase. This paper has provided a comprehensive and systematic review of this fast-moving research area and covers data-driven-based methodologies deployed in both static and dynamic SA. To extend beyond existing reviews, particular attention was paid to recent trends, such as the use of spatiotemporal feature selection algorithms and the increasing research activity in short-term voltage stability and frequency stability. The presented research developments, obtained in the space of only a few years, clearly demonstrate the growing potential of applying ML-based methods in power system SA.

VIII. ABBREVIATIONS AND ACRONYMS

The following abbreviations are used in this manuscript:

ADNC	Active Distribution Network Cell.
AUC	Area Under the Curve.
ANN	Artificial Neural Network.
CNN	Convolutional Neural Network.
DBN	Deep Belief Network.
DNN	Deep Neural Network.
DSA	Dynamic Security Assessment.
DT	Decision Tree.
ELM	Extreme Learning Machine.
FNN	Feedforward Neural Network.
FFNN	Fuzzy Feedforward Neural Network.
FSA	Frequency stability assessment.
FSM	Frequency Stability Margin.
GCN	Graph Convolutional Network.
GRU	Gated Recurrent Unit.

IBR	Inverter-Based Resources.
kNN	k-Nearest Neighbors.
HVDC	High Voltage Direct Current.
LASSO	Least Absolute Shrinkage and Selection Operator.
LDA	Linear Discriminant Analysis.
LSTM	Long short-term memory.
LTVS	Long-term voltage stability.
ML	Machine Learning.
mRMR	minimum Redundancy Maximum Relevancy.
PCA	Principal Component Analysis.
PG	Physic Guided.
PMU	Phasor Measurement Unit.
QC	Quantum Computing.
RBF	Radial Basis Function.
ROC	Receiver Operating Characteristic.
RoCoF	Rate of Change of Frequency.
RC	Random Classifier.
RF	Random Forest.
RNN	Recurrent Neural Network.
RVFL	Random Vector Functional Link.
SA	Security Assessment
SBS	Sequential Backward Selection
SFR	System Frequency Response
SFS	Sequential Forward Selection
SHAPE	SHapley Additive exPlanation
sl-CONE	saturated linear Couple Neurons
SOM	Self-Organizing Map
SSA	Static Security Assessment
STVS	Short-term voltage stability
SVM	Support Vector Machine
TSCC	Time Series Shapelet Classification
VSM	Voltage Stability Margin

REFERENCES

[1] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, T. Van Cutsem, and V. Vittal, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387–1401, May 2004.

[2] P. W. Sauer, K. L. Tomsovic, and V. Vittal, *Power System Stability and Control*, L. L. Grisby, Ed. CRC Press, 2007, ch. 15, pp. 421–430.

[3] C. Fu and A. Bose, "Contingency ranking based on severity indices in dynamic security analysis," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 980–985, Aug. 1999.

[4] F. De Caro, A. Andreotti, R. Araneo, M. Panella, A. Rosato, A. Vaccaro, and D. Villacci, "A review of the enabling methodologies for knowledge discovery from smart grids data," *Energies*, vol. 13, no. 24, p. 6579, Dec. 2020.

[5] L. Duchesne, E. Karangelos, and L. Wehenkel, "Recent developments in machine learning for energy systems reliability management," *Proc. IEEE*, vol. 108, no. 9, pp. 1656–1676, Sep. 2020.

[6] O. A. Alimi, K. Ouahada, and A. M. Abu-Mahfouz, "A review of machine learning approaches to power system security and stability," *IEEE Access*, vol. 8, pp. 113512–113531, 2020.

[7] N. Hatziargyriou, J. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal, and P. Pourbeik, "Stability definitions and characterization of dynamic behavior in systems with high penetration of power electronic interfaced technologies," IEEE, Piscataway, NJ, USA, Tech. Rep. PES-TR77, 2020.

[8] N. Hatziargyriou, J. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal, and P. Pourbeik, "Definition and classification of power system stability—Revisited & extended," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3271–3281, Jul. 2021.

[9] M. Köppen, "The curse of dimensionality," in *Proc. 5th Online World Conf. Soft Comput. Ind. Appl.*, vol. 1, 2000, pp. 4–8.

[10] V. Kumar, "Feature selection: A literature review," *Smart Comput. Rev.*, vol. 4, no. 3, pp. 211–229, Jun. 2014.

[11] P. Pudil and J. Novovičová, "Novel methods for feature subset selection with respect to problem knowledge," in *Feature Extraction, Construction and Selection: A Data Mining Perspective*. New York, NY, USA: Springer, 1998, pp. 101–116, doi: 10.1007/978-1-4615-5725-8.

[12] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[13] P. Xanthopoulos, P. M. Pardalos, T. B. Trafalis, P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Linear discriminant analysis," in *Robust Data Mining*. New York, NY, USA: Springer, 2013, pp. 27–33, doi: 10.1007/978-1-4419-9878-1.

[14] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Informat.*, vol. 85, pp. 189–203, Sep. 2018.

[15] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, Aug. 1987.

[16] G. Bontempi, *Statistical Foundation of Machine Learning: The Handbook*. Bruxelles: Univ. Libre de Bruxelles, 2022.

[17] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, Apr. 2016.

[18] J. De Stefani and G. Bontempi, "Factor-based framework for multivariate and multi-step-ahead forecasting of large scale time series," *Frontiers Big Data*, vol. 4, Sep. 2021, Art. no. 690267.

[19] K. Yu, X. Guo, L. Liu, J. Li, H. Wang, Z. Ling, and X. Wu, "Causality-based feature selection: Methods and evaluations," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–36, Sep. 2021.

[20] P. A. D. Castro and F. J. Von Zuben, "Learning Bayesian networks to perform feature selection," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 467–473.

[21] B. De Ville, "Decision trees," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 5, no. 6, pp. 448–455, 2013.

[22] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the Gini index and information gain criteria," *Ann. Math. Artif. Intell.*, vol. 41, no. 1, pp. 77–93, 2004.

[23] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, Dec. 2002.

[24] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification: Thinking With Examples for Effective Learning*. Berlin, Germany: Springer, 2016, pp. 207–235.

[25] B. Schölkopf, "The kernel trick for distances," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000, pp. 1–7.

[26] S. R. Samantaray, P. K. Dash, and G. Panda, "Distance relaying for transmission line using support vector machine and radial basis function neural network," *Int. J. Electr. Power Energy Syst.*, vol. 29, no. 7, pp. 551–556, Sep. 2007.

[27] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," in *Multiple Classifier Systems*. Berlin, Germany: Springer, 2005, pp. 278–285.

[28] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 1255–1260.

[29] G.-B. Huang, Y.-Q. Chen, and H. A. Babri, "Classification ability of single hidden layer feedforward neural networks," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 799–801, May 2000.

[30] A. D. Rasamoelina, F. Adjailia, and P. Sincák, "A review of activation function for artificial neural network," in *Proc. IEEE 18th World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Jan. 2020, pp. 281–286.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[32] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, and J. Cai, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[33] P. V. de Campos Souza, "Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature," *Appl. Soft Comput.*, vol. 92, Jul. 2020, Art. no. 106275.

[34] M. A. A. Albadra and S. Tiuna, "Extreme learning machine: A review," *Int. J. Appl. Eng. Res.*, vol. 12, no. 14, pp. 4610–4623, 2017.

[35] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1249, 2018.

[36] M. A. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, Oct. 2022, Art. no. 105151.

[37] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, Dec. 2020.

[38] G. Ejebe and B. Wollenberg, "Automatic contingency selection," *IEEE Trans. Power App. Syst.*, vol. PAS-98, no. 1, pp. 97–109, Jan. 1979.

[39] S. Kalyani and K. Shanti Swarup, "Classification and assessment of power system security using multiclass SVM," *IEEE Trans. Syst., Man, Cybern., C*, vol. 41, no. 5, pp. 753–758, Sep. 2011.

[40] T. Venkatesh and T. Jain, "Synchronized measurements-based wide-area static security assessment and classification of power systems using case based reasoning classifiers," *Comput. Electr. Eng.*, vol. 68, pp. 513–525, May 2018.

[41] R. Sunitha, R. K. Sreerama, and A. T. Mathew, "A composite security index for on-line steady-state security evaluation," *Electric Power Compon. Syst.*, vol. 39, no. 1, pp. 1–14, Jan. 2011.

[42] D. Niebur and A. J. Germond, "Unsupervised neural net classification of power system static security states," *Int. J. Electr. Power Energy Syst.*, vol. 14, nos. 2–3, pp. 233–242, Apr. 1992.

[43] C. I. F. Agreira, C. M. M. Ferreira, J. A. D. Pinto, and F. P. M. Barbosa, "The rough set theory applied to the steady-state contingency classification," in *Proc. 39th Int. Universities Power Eng. Conf. (UPEC)*, vol. 2, Bristol, U.K., 2004, pp. 1140–1144.

[44] D. S. Javan, H. R. Mashhadi, and M. Rouhani, "Static security assessment using radial basis function neural networks based on growing and pruning method," in *Proc. IEEE Electr. Power Energy Conf.*, Aug. 2010, pp. 1–6.

[45] D. Seyed Javan, H. Rajabi Mashhadi, and M. Rouhani, "A fast static security assessment method based on radial basis function neural networks using enhanced clustering," *Int. J. Electr. Power Energy Syst.*, vol. 44, no. 1, pp. 988–996, Jan. 2013.

[46] P. Sekhar and S. Mohanty, "Classification and assessment of power system static security using decision tree and random forest classifiers," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 29, no. 3, pp. 465–474, May 2016.

[47] I. S. Saeh, M. W. Mustafa, Y. S. Mohammed, and M. Almaktar, "Static security classification and evaluation classifier design in electric power grid with presence of PV power plants using C-4.5," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 283–290, Apr. 2016.

[48] M. Ramirez-Gonzalez, F. R. S. Sevilla, and P. Korba, "Convolutional neural network based approach for static security assessment of power systems," in *Proc. World Autom. Congr. (WAC)*, Aug. 2021, pp. 106–110.

[49] N. D. Hatziargyriou, G. C. Contaxis, and N. C. Sideris, "A decision tree method for on-line steady state security assessment," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 1052–1061, May 1994.

[50] K. S. Swarup and P. B. Corthis, "ANN approach assesses system security," *IEEE Comput. Appl. Power*, vol. 15, no. 3, pp. 32–38, Jul. 2002.

[51] S. Weerasooriya, M. A. El-Sharkawi, M. Damborg, and R. J. Marks, "Towards static-security assessment of a large-scale power system using neural networks," *IEE Proc. C Gener., Transmiss. Distrib.*, vol. 139, no. 1, pp. 64–70, 1992.

[52] A. Dhandhia and V. Pandya, "Multi classification of static security assessment using teaching learning based optimization enhanced support vector machine," in *Proc. 8th Int. Conf. Power Syst. (ICPS)*, Dec. 2019, pp. 1–6.

[53] Y. Du, F. Li, and C. Huang, "Applying deep convolutional neural network for fast security assessment with N-1 contingency," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2019, pp. 1–5.

[54] P. Sekhar and S. Mohanty, "An online power system static security assessment module using multi-layer perceptron and radial basis function network," *Int. J. Electr. Power Energy Syst.*, vol. 76, pp. 165–173, Mar. 2016.

[55] S. Chauhan and M. Dave, "Ann for transmission system static security assessment," *Int. J. Electr. Power Energy Syst.*, vol. 24, no. 10, pp. 867–873, 2002.

[56] R. Sunitha, R. S. Kumar, and A. T. Mathew, "Online static security assessment module using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4328–4335, Nov. 2013.

[57] Y. Du, F. Li, J. Li, and T. Zheng, "Achieving 100x acceleration for N-1 contingency screening with uncertain scenarios using deep convolutional neural network," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3303–3305, Jul. 2019.

[58] Y. Du, F. Li, T. Zheng, and J. Li, "Fast cascading outage screening based on deep convolutional neural network and depth-first search," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2704–2715, Jul. 2020.

[59] T.-E. Huang, Q. Guo, H. Sun, C.-W. Tan, and T. Hu, "A deep spatial–temporal data-driven approach considering microclimates for power system security assessment," *Appl. Energy*, vol. 237, pp. 36–48, Mar. 2019.

[60] Y. Li, Y. Li, and Y. Sun, "Online static security assessment of power systems based on lasso algorithm," *Appl. Sci.*, vol. 8, no. 9, p. 1442, Aug. 2018.

[61] G. Ravikumar and S. A. Khaparde, "Taxonomy of PMU data based catastrophic indicators for power system stability assessment," *IEEE Syst. J.*, vol. 12, no. 1, pp. 452–464, Mar. 2018.

[62] H. S. Salama and I. Vokony, "Voltage stability indices—A comparison and a review," *Comput. Electr. Eng.*, vol. 98, Mar. 2022, Art. no. 107743.

[63] R. Shah, N. Mithulananthan, R. C. Bansal, and V. K. Ramachandaramurthy, "A review of key power system stability challenges for large-scale PV integration," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 1423–1436, Jan. 2015.

[64] C. Pang, F. Prabhakara, A. El-abiad, and A. Koivo, "Security evaluation in power systems using pattern recognition," *IEEE Trans. Power App. Syst.*, vol. PAS-93, no. 3, pp. 969–976, May 1974.

[65] C.-W. Liu, M.-C. Su, S.-S. Tsay, and Y.-J. Wang, "Application of a novel fuzzy neural network to real-time transient stability swings prediction based on synchronized phasor measurements," *IEEE Trans. Power Syst.*, vol. 14, no. 2, pp. 685–692, May 1999.

[66] S. R. Samantaray, I. Kamwa, and G. Joos, "Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame," *IET Gener., Transmiss. Distrib.*, vol. 4, no. 12, p. 1334, 2010.

[67] B. Li, J. Xiao, and X. Wang, "Feature reduction for power system transient stability assessment based on neighborhood rough set and discernibility matrix," *Energies*, vol. 11, no. 1, p. 185, Jan. 2018.

[68] F. Bellizio, J. L. Cremer, M. Sun, and G. Strbac, "A causality based feature selection approach for data-driven dynamic security assessment," *Electric Power Syst. Res.*, vol. 201, Dec. 2021, Art. no. 107537.

[69] 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, May 2022, Art. no. 107738.

[70] S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "Decision trees for real-time transient stability prediction," *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp. 1417–1426, 1994.

[71] S. Wu, L. Zheng, W. Hu, R. Yu, and B. Liu, "Improved deep belief network and model interpretation method for power system transient stability assessment," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 1, pp. 27–37, Jan. 2020.

[72] C. Ren, H. Yuan, Q. Li, R. Zhang, and Y. Xu, "Pre-fault dynamic security assessment of power systems for multiple different faults via multi-label learning," *IEEE Trans. Power Syst.*, vol. 38, no. 6, pp. 5501–5511, Nov. 2023.

[73] H. Cui, Q. Wang, Y. Ye, Y. Tang, and Z. Lin, "A combinational transfer learning framework for online transient stability prediction," *Sustain. Energy, Grids Netw.*, vol. 30, Jun. 2022, Art. no. 100674.

[74] J. Huang, L. Guan, Y. Su, H. Yao, M. Guo, and Z. Zhong, "Recurrent graph convolutional network-based multi-task transient stability assessment framework in power system," *IEEE Access*, vol. 8, pp. 93283–93296, 2020.

[75] S. Kalyani and K. Swarp, "Transient security assessment and classification using support vector machine," *J. Elect. Syst.*, 2009.

[76] I. Kamwa, S. R. Samantaray, and G. Joos, "Development of rule-based classifiers for rapid stability assessment of wide-area post-disturbance records," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 258–270, Feb. 2009.

[77] A. B. Mosavi, A. Amiri, and H. Hosseini, "A learning framework for size and type independent transient stability prediction of power system using twin convolutional support vector machine," *IEEE Access*, vol. 6, pp. 69937–69947, 2018.

[78] A. Gupta, G. Gurrala, and P. S. Sastry, "An online power system stability monitoring system using convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 864–872, Mar. 2019.

[79] R. Yan, G. Geng, Q. Jiang, and Y. Li, "Fast transient stability batch assessment using cascaded convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2802–2813, Jul. 2019.

[80] R. Liu, G. Verbič, and J. Ma, "A new dynamic security assessment framework based on semi-supervised learning and data editing," *Electric Power Syst. Res.*, vol. 172, pp. 221–229, Jul. 2019.

[81] C. Ren and Y. Xu, "Transfer learning-based power system online dynamic security assessment: Using one model to assess many unlearned faults," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 821–824, Jan. 2020.

[82] R. Zhang, W. Yao, Z. Shi, L. Zeng, Y. Tang, and J. Wen, "A graph attention networks-based model to distinguish the transient rotor angle instability and short-term voltage instability in power systems," *Int. J. Electr. Power Energy Syst.*, vol. 137, May 2022, Art. no. 107783.

[83] J. Liu, H. Sun, Y. Li, W. Fang, and S. Niu, "An improved power system transient stability prediction model based on mRMR feature selection and WTA ensemble learning," *Appl. Sci.*, vol. 10, no. 7, p. 2255, Mar. 2020.

[84] D. J. Sobajic and Y.-H. Pao, "Artificial neural-net based dynamic security assessment for electric power systems," *IEEE Trans. Power Syst.*, vol. 4, no. 1, pp. 220–228, 1989.

[85] 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 Trans. Power Syst.*, vol. 34, no. 6, pp. 5044–5052, Nov. 2019.

[86] A. A. El-Keib and X. Ma, "Application of artificial neural networks in voltage stability assessment," *IEEE Trans. Power Syst.*, vol. 10, no. 4, pp. 1890–1896, Nov. 1995.

[87] B. Jeyasurya, "Artificial neural networks for on-line voltage stability assessment," in *Proc. IEEE PES Summer Meeting*, vol. 4, Jul. 2000, pp. 2014–2018.

[88] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, "Online monitoring of voltage stability margin using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1566–1574, Aug. 2010.

[89] S. Hashemi and M. R. Aghamohammadi, "Wavelet based feature extraction of voltage profile for online voltage stability assessment using RBF neural network," *Int. J. Electr. Power Energy Syst.*, vol. 49, pp. 86–94, Jul. 2013.

[90] R. Zhang, Y. Xu, Z. Y. Dong, P. Zhang, and K. P. Wong, "Voltage stability margin prediction by ensemble based extreme learning machine," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2013, pp. 1–5.

[91] K. S. Sajan, V. Kumar, and B. Tyagi, "Genetic algorithm based support vector machine for on-line voltage stability monitoring," *Int. J. Electr. Power Energy Syst.*, vol. 73, pp. 200–208, Dec. 2015.

[92] S. Li and V. Ajjarapu, "Real-time monitoring of long-term voltage stability via convolutional neural network," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2017, pp. 1–5.

[93] S. Li, V. Ajjarapu, and M. Djukanovic, "Adaptive online monitoring of voltage stability margin via local regression," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 701–713, Jan. 2018.

[94] W. M. Villa-Acevedo, J. M. López-Lezama, and D. G. Colomé, "Voltage stability margin index estimation using a hybrid kernel extreme learning machine approach," *Energies*, vol. 13, no. 4, p. 857, Feb. 2020.

[95] T. Wu, Y. A. Zhang, and H. Wen, "Voltage stability monitoring based on disagreement-based deep learning in a time-varying environment," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 28–38, Jan. 2021.

[96] V. Malbasa, C. Zheng, P.-C. Chen, T. Popovic, and M. Kezunovic, "Voltage stability prediction using active machine learning," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 3117–3124, Nov. 2017.

[97] R. Diao, K. Sun, V. Vittal, R. J. O'Keefe, M. R. Richardson, N. Bhatt, D. Stradford, and S. K. Sarawgi, "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 832–839, May 2009.

[98] H.-Y. Su and T.-Y. Liu, "Enhanced-online-random-forest model for static voltage stability assessment using wide area measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6696–6704, Nov. 2018.

[99] K. D. Dharmapala, A. Rajapakse, K. Narendra, and Y. Zhang, "Machine learning based real-time monitoring of long-term voltage stability using voltage stability indices," *IEEE Access*, vol. 8, pp. 222544–222555, 2020.

[100] H. Cai, H. Ma, and D. J. Hill, "A data-based learning and control method for long-term voltage stability," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3203–3212, Jul. 2020.

[101] A. Boricic, J. L. R. Torres, and M. Popov, "Comprehensive review of short-term voltage stability evaluation methods in modern power systems," *Energies*, vol. 14, no. 14, p. 4076, Jul. 2021.

[102] 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, Mar. 2016.

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

[104] C. Ren, Y. Xu, Y. Zhang, and R. Zhang, "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, Jun. 2020.

[105] Y. Luo, C. Lu, L. Zhu, and J. Song, "Data-driven short-term voltage stability assessment based on spatial–temporal graph convolutional network," *Int. J. Electr. Power Energy Syst.*, vol. 130, Sep. 2021, Art. no. 106753.

[106] M. Zhang, J. Li, Y. Li, and R. Xu, "Deep learning for short-term voltage stability assessment of power systems," *IEEE Access*, vol. 9, pp. 29711–29718, 2021.

[107] L. Zhu, D. J. Hill, and C. Lu, "Intelligent short-term voltage stability assessment via spatial attention rectified RNN learning," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 7005–7016, Oct. 2021.

[108] J. D. Pinzón and D. G. Colomé, "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, Jun. 2019.

[109] Y. Xu, R. Zhang, J. Zhao, Z. Y. Dong, D. Wang, H. Yang, and K. P. Wong, "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, Aug. 2016.

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

[111] 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, Oct. 2017.

[112] L. Zhu, C. Lu, Y. Liu, W. Wu, and C. Hong, "Wordbook-based light-duty time series learning machine for short-term voltage stability assessment," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 18, pp. 4492–4499, Dec. 2017.

[113] 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, Sep. 2019.

[114] L. Zhu, C. Lu, I. Kamwa, and H. Zeng, "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, Mar. 2020.

[115] G. Wang, Z. Zhang, Z. Bian, and Z. Xu, "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, May 2021, Art. no. 106647.

[116] Y. Zhang, Y. Xu, and Z. Y. Dong, "Robust ensemble data analytics for incomplete PMU measurements-based power system stability assessment," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1124–1126, Jan. 2018.

[117] H. Zhang, X. Li, X. Fu, C. Wang, and T. Bi, "Data-driven frequency stability assessment of power systems considering wind power uncertainty," in *Proc. 7th Int. Conf. Power Renew. Energy (ICPRE)*, Sep. 2022, pp. 815–820.

[118] J. Xie and W. Sun, "A transfer and deep learning-based method for online frequency stability assessment and control," *IEEE Access*, vol. 9, pp. 75712–75721, 2021.

[119] C. Ma, L. Wang, C. Gai, D. Yang, P. Zhang, H. Zhang, and C. Li, "Frequency security assessment for receiving-end system based on deep learning method," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia*, Jul. 2020, pp. 831–836.

[120] Y.-K. Wu, "Frequency stability for an island power system: Developing an intelligent preventive-corrective control mechanism for an offshore location," *IEEE Ind. Appl. Mag.*, vol. 23, no. 2, pp. 74–87, Mar. 2017.

[121] Y. Wen, R. Zhao, M. Huang, and C. Guo, "Data-driven transient frequency stability assessment: A deep learning method with combined estimation-correction framework," *Energy Convers. Econ.*, vol. 1, no. 3, pp. 198–209, Dec. 2020.

[122] A. E. Gavoyiannis, E. M. Voumvoulakis, and N. D. Hatziargyriou, "On-line supervised learning for dynamic security classification using probabilistic neural networks," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Jun. 2005, pp. 2669–2675.

[123] J. Kruse, B. Schäfer, and D. Witthaut, "Revealing drivers and risks for power grid frequency stability with explainable AI," *Patterns*, vol. 2, no. 11, Nov. 2021, Art. no. 100365.

[124] Y. Xu, Y. Dai, Z. Y. Dong, R. Zhang, and K. Meng, "Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems," *Neural Comput. Appl.*, vol. 22, nos. 3–4, pp. 501–508, Mar. 2013.

[125] R.-F. Chang, C.-N. Lu, and T.-Y. Hsiao, "Prediction of frequency response after generator outage using regression tree," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2146–2147, Nov. 2005.

[126] L. Xu, L. Li, M. Wang, X. Wang, Y. Li, W. Li, and K. Zhou, "Online prediction method for power system frequency response analysis based on swarm intelligence fusion model," *IEEE Access*, vol. 11, pp. 13519–13532, 2023.

[127] Q. Wang, F. Li, Y. Tang, and Y. Xu, "Integrating model-driven and data-driven methods for power system frequency stability assessment and control," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4557–4568, Nov. 2019.

[128] S. Lei, L. Wang, W. Cui, M. Wozniak, and D. Polap, "Research on a LSTM based method of forecasting primary frequency modulation of grid," *J. Internet Technol.*, vol. 21, no. 3, pp. 791–798, 2020.

[129] X. Wang, F. Blaabjerg, and W. Wu, "Modeling and analysis of harmonic stability in an AC power-electronics-based power system," *IEEE Trans. Power Electron.*, vol. 29, no. 12, pp. 6421–6432, Dec. 2014.

[130] R. N. Damas, Y. Son, M. Yoon, S.-Y. Kim, and S. Choi, "Subsynchronous oscillation and advanced analysis: A review," *IEEE Access*, vol. 8, pp. 224020–224032, 2020.

[131] J. Shair, X. Xie, L. Wang, W. Liu, J. He, and H. Liu, "Overview of emerging subsynchronous oscillations in practical wind power systems," *Renew. Sustain. Energy Rev.*, vol. 99, pp. 159–168, Jan. 2019.

[132] Z. Song, C. M. Hackl, A. Anand, A. Thommessen, J. Petzschmann, O. Kamel, R. Braunbehrens, A. Kaifel, C. Roos, and S. Hauptmann, "Digital twins for the future power system: An overview and a future perspective," *Sustainability*, vol. 15, no. 6, p. 5259, Mar. 2023.

[133] *Modelling and Aggregation of Loads in Flexible Power Networks*, CIGRE WG C4.605, CIGRE, Paris, France, 2013.

[134] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids," *IEEE Power Energy Mag.*, vol. 5, no. 4, pp. 78–94, Jul./Aug. 2007.

[135] M. Farrokhabadi, C. A. Canizares, J. W. Simpson-Porco, E. Nasr, L. Fan, P. A. Mendoza-Araya, R. Tonkoski, U. Tamrakar, N. Hatziargyriou, D. Lagos, and R. W. Wies, "Microgrid stability definitions, analysis, and examples," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 13–29, Jan. 2020.

[136] M. H. J. Bollen, "What is power quality?" *Electric Power Syst. Res.*, vol. 66, no. 1, pp. 5–14, Jul. 2003.

[137] D. Pattabiraman, R. H. Lasseter., and T. M. Jahns, "Comparison of grid following and grid forming control for a high inverter penetration power system," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2018, pp. 1–5.

[138] IEEE PES Task Force on Microgrid Stability Analysis and Modeling, *Microgrid Stability Definitions, Analysis, and Modeling*, IEEE, Piscataway, NY, USA, 2018.

[139] X. Wang, M. G. Taul, H. Wu, Y. Liao, F. Blaabjerg, and L. Harnefors, "Grid-synchronization stability of converter-based resources—An overview," *IEEE Open J. Ind. Appl.*, vol. 1, pp. 115–134, 2020.

[140] Q. Zhang, M. Mao, G. Ke, L. Zhou, and B. Xie, "Stability problems of PV inverter in weak grid: A review," *IET Power Electron.*, vol. 13, no. 11, pp. 2165–2174, Aug. 2020.

[141] R. J. Thomas and B.-Y. Ku, "Approximations of power system dynamic load characteristics by artificial neural networks," in *Proc. 1st Int. Forum Appl. Neural Netw. Power Syst.*, 1991, pp. 178–182.

[142] B.-Y. Ku, R. J. Thomas, C.-Y. Chiou, and C.-J. Lin, "Power system dynamic load modeling using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1868–1874, 1994.

[143] M. Bostanci, J. Koplowitz, and C. W. Taylor, "Identification of power system load dynamics using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1468–1473, 1997.

[144] T. Abedin, M. S. H. Lipu, M. A. Hannan, P. J. Ker, S. A. Rahman, C. T. Yaw, S. K. Tiong, and K. M. Muttaqi, "Dynamic modeling of HVDC for power system stability assessment: A review, issues, and recommendations," *Energies*, vol. 14, no. 16, p. 4829, Aug. 2021.

[145] S. Cao, V. Dinavahi, and N. Lin, "Machine learning based transient stability emulation and dynamic system equivalencing of large-scale AC–DC grids for faster-than-real-time digital twin," *IEEE Access*, vol. 10, pp. 112975–112988, 2022.

[146] M. Naderi, Y. Khayat, Q. Shafiee, F. Blaabjerg, and H. Bevrani, "Dynamic modeling, stability analysis and control of interconnected microgrids: A review," *Appl. Energy*, vol. 334, Mar. 2023, Art. no. 120647.

[147] G. Mitrentsis and H. Lens, "Data-driven dynamic models of active distribution networks using unsupervised learning techniques on field measurements," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 2952–2965, Jul. 2021.

[148] E. O. Kontis, T. A. Papadopoulos, M. H. Syed, E. Guillo-Sansano, G. M. Burt, and G. K. Papagiannis, "Artificial-intelligence method for the derivation of generic aggregated dynamic equivalent models," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2947–2956, Jul. 2019.

[149] 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, Sep. 2019.

[150] J. Beyza and J. M. Yusta, "The effects of the high penetration of renewable energies on the reliability and vulnerability of interconnected electric power systems," *Rel. Eng. Syst. Saf.*, vol. 215, Nov. 2021, Art. no. 107881.

[151] A. Stanković, K. Tomsovic, F. De Caro, M. Braun, J. Chow, N. Äukalevski, I. Dobson, J. Eto, and B. Fink, "Methods for analysis and quantification of power system resilience," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 4774–4787, Sep. 2023.

[152] K. P. Guddanti, B. Vyakaranam, K. Mahapatra, Z. Hou, P. Etingov, N. Samaan, T. Nguyen, and Q. Nguyen, "Dynamic security analysis framework for future large grids with high renewable penetrations," *IEEE Access*, vol. 11, pp. 8159–8171, 2023.

[153] R. Liu, G. Verbic, J. Ma, and D. J. Hill, "Fast stability scanning for future grid scenario analysis," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 514–524, Jan. 2018.

[154] S. R. Khuntia, J. L. Rueda, and M. A. M. M. Meijden, "Risk-based security assessment of transmission line overloading considering spatio-temporal dependence of load and wind power using vine copula," *IET Renew. Power Gener.*, vol. 13, no. 10, pp. 1770–1779, Jul. 2019.

[155] C. Moya, S. Zhang, G. Lin, and M. Yue, "DeepONet-grid-UQ: A trustworthy deep operator framework for predicting the power grid's post-fault trajectories," *Neurocomputing*, vol. 535, pp. 166–182, May 2023.

[156] X.-Y. Jing, X. Zhang, X. Zhu, F. Wu, X. You, Y. Gao, S. Shan, and J.-Y. Yang, "Multiset feature learning for highly imbalanced data classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 139–156, Jan. 2021.

[157] C. Crozier, K. Baker, Y. Du, J. Mohammadi, and M. Li, "Data-driven contingency selection for fast security constrained optimal power flow," in *Proc. 17th Int. Conf. Probabilistic Methods Appl. to Power Syst. (PMAPS)*, Jun. 2022, pp. 1–6.

[158] E. Ciapessoni, D. Cirio, A. Pitto, and M. Sforna, "A probabilistic risk-based security assessment tool allowing contingency forecasting," in *Proc. IEEE Int. Conf. Probabilistic Methods Appl. to Power Syst. (PMAPS)*, Jun. 2018, pp. 1–6.

[159] F. De Caro, D. Astolfi, and A. Vaccaro, "Experimental assessment of data-driven methods for detecting wind power ramps," in *Proc. Int. Conf. Clean Electr. Power (ICCEP)*, Jun. 2023, pp. 492–497.

[160] F. De Caro, A. J. Collin, and A. Vaccaro, "Evaluation of classification models for forecasting critical voltage events in power systems," *Smart Grids Sustain. Energy*, vol. 8, no. 1, 2023, doi: 10.1007/s40866-022-00159-6.

[161] R. Dashti, M. Daisy, H. Mirshekali, H. R. Shaker, and M. H. Aliabadi, "A survey of fault prediction and location methods in electrical energy distribution networks," *Measurement*, vol. 184, Nov. 2021, Art. no. 109947.

[162] O. F. Avila, J. A. P. Filho, and W. Peres, "Steady-state security assessment in distribution systems with high penetration of distributed energy resources," *Electric Power Syst. Res.*, vol. 201, Dec. 2021, Art. no. 107500.
[163] K. Tang, M. Ge, S. Dong, J. Cui, and X. Ma, "Transmission contingency analysis based on data-driven equivalencing of radial distribution networks considering uncertainties," *IEEE Access*, vol. 8, pp. 227247–227254, 2020.
[164] W. Liu, H. Wang, X. Shen, and I. W. Tsang, "The emerging trends of multi-label learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7955–7974, Nov. 2022.
[165] Y. Zhou and P. Zhang, "Noise-resilient quantum machine learning for stability assessment of power systems," *IEEE Trans. Power Syst.*, vol. 38, no. 1, pp. 475–487, Jan. 2023.
[166] F. Feng, Y. Zhou, and P. Zhang, "Quantum power flow," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3810–3812, Jul. 2021.
[167] B. Sævarsson, S. Chatzivasileiadis, H. Jóhannsson, and J. Østergaard, "Quantum computing for power flow algorithms: Testing on real quantum computers," 2022, *arXiv:2204.14028*.
[168] Y. Zhou, F. Feng, and P. Zhang, "Quantum electromagnetic transients program," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3813–3816, Jul. 2021.

FABRIZIO DE CARO (Member, IEEE) was born in Benevento, Italy, in 1992. He received the B.Sc. and M.Sc. degrees in energy engineering and the Ph.D. degree in information technologies for engineering from the University of Sannio (UniSannio), Benevento, in 2014, 2016, and 2020, respectively. He is currently a Postdoctoral Researcher with the Power System Research Group (PSRG), Engineering Department, UniSannio, and a Scientific Collaborator with the Machine Learning Group (MLG), Université Libre de Bruxelles, Belgium. His research interests include effective renewable energy sources integration in smart grids, wind power forecasting, artificial intelligence in power systems, decision-making tools in the presence of uncertainty, electricity markets, and resilience of power systems. He serves as an associate editor and a reviewer for several international journals.

ADAM JOHN COLLIN (Senior Member, IEEE) was born in Edinburgh, U.K., in April 1985. He received the B.Eng. degree (Hons.) in electrical and electronic engineering from The University of Edinburgh, Edinburgh, in 2007, the M.Sc. degree in renewable energy and distributed generation from Heriot-Watt University, Edinburgh, in 2008, and the Ph.D. degree from The University of Edinburgh, in 2013. He was a Postdoctoral Researcher with The University of Edinburgh, until 2017. From 2017 to 2021, he was with the University of Campania "Luigi Vanvitelli," Aversa, Italy. He is currently a Senior Researcher/Lecturer with the University of Sannio, Benevento, Italy. His research interests include power system harmonics, load modeling, and distribution system analysis.

GIORGIO MARIA GIANNUZZI received the degree in electrical engineering from the University of Rome. Until December 2000, he was with ABB, where he was in charge of network studies, protection, and control applications, with special reference to RTU apparatus and data engineering issues. Since 2001, he has been with Terna Rete Italia SpA as an Expert in defense plans/systems, dynamic studies, protection, telecontrol, and substation automation. From 2004 to 2011, he coordinated the study, design, and activation of wide-area defense systems (including an interruptible customers system) and wide-area monitoring systems. In addition, under guidance, the main security energy management systems were designed and coded. He supervised the revision of the main Italian grid code technical enclosures.

COSIMO PISANI was born in Benevento, Italy, in 1985. He received the M.Sc. degree (Hons.) in energy engineering from the University of Sannio, Benevento, in 2010, and the Ph.D. degree in electrical engineering from the University of Naples "Federico II," Naples, Italy, in 2014. From May 2014 to March 2016, he was a Research Fellow with the University of Sannio. Since March 2016, he has been with Terna. He is currently a Senior Power System Engineer with the Dispatching and Switching Department Stability and Network Calculations. His research interests include the applications of dynamic stability of power systems, wide area monitoring, and protection systems, high-voltage direct current systems, and power system restoration.

ALFREDO VACCARO (Senior Member, IEEE) received the M.Sc. (Hons.) degree in electronic engineering from the University of Salerno, Salerno, Italy, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada. From 1999 to 2002, he was an Assistant Researcher with the Department of Electrical and Electronic Engineering, University of Salerno. From March 2002 to October 2015, he was an Assistant Professor of electric power systems with the Department of Engineering, University of Sannio, Benevento, Italy, where he is currently a Full Professor of electrical power systems. His research interests include soft computing and interval-based methods applied to power system analysis, and advanced control architectures for diagnosis and protection of distribution networks. He is an Associate Editor of IEEE TRANSACTIONS ON POWER SYSTEMS and IEEE TRANSACTIONS ON SMART GRIDS.

Open Access funding provided by 'Università degli Studi del Sannio' within the CRUI CARE Agreement