

## Machine-learned security assessment for changing system topologies

Bellizio, Federica ; Cremer, Jochen; Strbac, Goran

**DOI**

[10.1016/j.ijepes.2021.107380](https://doi.org/10.1016/j.ijepes.2021.107380)

**Publication date**

2022

**Document Version**

Final published version

**Published in**

International Journal of Electrical Power and Energy Systems

**Citation (APA)**

Bellizio, F., Cremer, J., & Strbac, G. (2022). Machine-learned security assessment for changing system topologies. *International Journal of Electrical Power and Energy Systems*, 134, Article 107380. <https://doi.org/10.1016/j.ijepes.2021.107380>

**Important note**

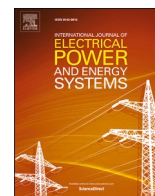
To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



## Machine-learned security assessment for changing system topologies

Federica Bellizio<sup>a,\*</sup>, Jochen L. Cremer<sup>b</sup>, Goran Strbac<sup>a</sup>

<sup>a</sup> Imperial College London, South Kensington, London SW7 2BU, United Kingdom

<sup>b</sup> TU Delft, Mekelweg 5, 2628 CD Delft, the Netherlands

### ARTICLE INFO

#### Keywords:

Topology Changes  
Markov Blanket  
Decision Trees  
Power Systems Operation  
Dynamic Security Assessment

### ABSTRACT

Machine learning has been used in the past to construct predictors, also known as classifiers, for dynamic security assessment. Although accurate classifiers can be trained for a single topology, often they do not work for another. However, the power system topology can change frequently during operation due to maintenance and control actions. At one topological configuration, the system may have a different response to a fault than at another as the underlying distribution of power flows can be completely different. Quantifying the impact of changes in the topology on the predictive models' performance is an important step forward to minimize inaccurate predictions and improve their reliability.

In this paper, for the first time, a metric for quantifying the impact of a topology change on the accuracy of the classification model is proposed. The key novelty is to first select a subset of power flow features with a physically informed feature selection technique and subsequently compute the metric with a novel convex hull-based analysis. In addition, the approach can advise to effectively constructing new training databases that improve the accuracy of new machines trained after high-impact topology changes. Through a case study using transient stability on the IEEE 68-bus system, the use of the proposed metric in real-time operation was demonstrated. 17 high-impact topology changes were successfully detected among 42 studied topological changes. The subsequent effective construction of the training database improved the predictive accuracy by around 10%. An interesting finding is the amount of newly generated data can be reduced by up to 85% as often the generated data is the barrier for data-driven DSA. The proposed workflow significantly reduces data and trains robust classifiers against topological changes marking a fundamental step forward.

### 1. Introduction

The reliable operation of the electric power system is becoming a task of paramount importance worldwide since the massive integration of renewable energy sources exposes the grid to more frequent dynamical phenomena [1,2]. The intermittent nature of renewable energy and the demand-side flexibility makes the operation more uncertain and dynamic than in the past. Novel operating approaches that consider these new dynamics are needed to comply N-1 security standards (operation reduced by one equipment) during all of the hours. Otherwise, investments in redundant grid infrastructure to maintain the system's security become necessary. Hence, an efficient operation of the system is close to its limits with smaller safety margins to increase the utilization of the existing assets, and post-fault corrective control actions are important for the management of the reliability [3]. A power system is reliable when it can supply electricity with high enough probability to the end-users at all times (i.e. adequacy) and withstand sudden

disturbances without major service interruptions in the real-time (i.e. security) [4]. When analysing the security of the system two important components can be distinguished. The static security refers to whether the system subjected to a disturbance settles at a new post-disturbance operating condition that fulfils all physical constraints. This involves the steady-state analysis of the post-disturbance operating condition to verify if the constraints on voltage and equipment ratings are met. The dynamic security refers to whether the system survives the transition from pre-fault to post-fault. Due to the wide range of dynamical phenomena involved in the operations, system's operators can no longer operate in a static paradigm. If the same static paradigm is followed, operators are forced to set very large margins to cope with all the possible dynamical phenomena. Following this static operating paradigm requires significant investments in the infrastructure and is therefore costly and inefficient. In contrast a dynamic paradigm promises to avoid these investments by lowering the operating margins. The initial step toward a dynamic paradigm is to have information available

\* Corresponding author.

E-mail addresses: [f.bellizio18@imperial.ac.uk](mailto:f.bellizio18@imperial.ac.uk) (F. Bellizio), [j.l.cremer@tudelft.nl](mailto:j.l.cremer@tudelft.nl) (J.L. Cremer), [g.strbac@imperial.ac.uk](mailto:g.strbac@imperial.ac.uk) (G. Strbac).

<https://doi.org/10.1016/j.ijepes.2021.107380>

Received 15 February 2021; Received in revised form 30 April 2021; Accepted 15 June 2021

Available online 24 August 2021

0142-0615/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

in real-time about the dynamic performance of the system in response to possible faults [1]. To assess if the system operates dynamically secure, a set of typical dynamical phenomena is studied, mainly relating to the stability in rotor angles, i.e. transient stability, frequency and voltages [2]. For instance, rotor angle stability refers to the ability of synchronous machines to remain in synchronism after being subjected to a disturbance. Each of these phenomena needs to be analysed separately and different analytical techniques are used, e.g. transient stability is evaluated by performing an event-type simulation on a large model involving ordinary differential equations. Therefore, the Dynamic Security Assessment (DSA) generally requires the time-domain simulation of a system's model consisting of complex algebraic and differential equations. Processing these simulations in real-time (or shortly before) is difficult as for each possible fault a separate event-type numerical simulation is required. One could carry out these studies well before real-time operation where more computational power is available. However, this would require carrying out significantly more simulations as the operation is uncertain and more possible operating scenarios need to be studied. This marks a significant challenge for system operations as currently the only viable solution is to increase the static operating margins which is increasingly inefficient.

### 1.1. The machine learning approach to DSA

The idea of data-driven approaches to DSA is to approximate the dynamical system response to faults rather than simulating in the time-domain [5,6]. The key advantage of these approaches is they provide real-time security predictions with almost no computational resources, allowing operators to relax the static margins since dynamics are assessed directly in the real-time operation. Hence, operators can fully use their existing assets in a smarter way and do not need to invest in new system infrastructure, e.g. switches or redundant lines, to satisfy the reliability requirements under contingencies [7].

The first step of these data-driven approaches is to prepare them offline well before real-time operation. This preparation starts with creating a training database that includes operating conditions (OCs) from historical observations and synthetically generated data. For each OC and for each credible contingency, the post-fault security status, i.e. transient stability in this work, is evaluated by running time-domain simulations. Subsequently, a machine learning classifier is constructed using the variables that describe the pre-fault OCs as features and the post-fault status as classification label. The key idea of these approaches is to carry out this preparation periodically and offline and then use the trained data-driven predictors in real-time operations. In this second step, the data-driven approaches can instantly predict/approximate the post-fault security of the real-time OCs and possible OCs that go beyond the training database [6]. Many possible data-driven approaches were studied in the past. Several deep learning models have been proposed for the transient stability assessment showing a very promising performance in terms of accuracy [8–10]. Conversely, Decision Trees (DTs) or ensembles of DTs have been mostly used as they are more understandable [6,11,12].

### 1.2. Topological changes in DSA

System operators change the system topology for maintenance and control purposes in order to better handle faults, their dynamics and the uncertainty surrounding the generation of renewables [13,14]. The system's topology is determined by the status of the switching components responsible for maintaining the connectivity between the components in the network [15]. Therefore, topological changes can be for instance disconnecting lines, as in this work. Topological changes can also be switching on/off generators, shunt components or major (aggregated) batteries, and merging the substations. These frequent changes of equipment in modern power systems pose new challenges to data-driven DSA approaches. The system stability is very sensitive to

system changes [16], and hence different topologies correspond to different power flow distributions [9,17]. The OCs following a topology change may be then very different from those included in the training database as they originated from different power flow distributions. Depending on the type and the size of the change in these distributions, a data-driven classifier trained for a specific system topology of equipment may not work anymore when the equipment is (slightly) different. Thus, changes in the network topology make the data-driven approaches less reliable, ultimately reducing the trust of operators. In the past, researchers approached making data-driven DSA robust against changes in the network topology. These approaches can be classified into three sets:

(i) the first set of approaches generalises the classifier to many topological configurations. The topological variations are considered in the generation of the training database by sampling many different system topologies [18–20]. For large systems, however, the dimension of topological (and equipment) configurations is so large that it is not feasible to consider each configuration in the training. Then, determining if this large training database is representative for a system once the system is changed is challenging [21]. Considering similar topological configurations may also increase the redundancy of the training database, and hence decreases the predictive performance of the classifier [22].

(ii) the second set of approaches considers updating the training database and the classifier at fixed time intervals (e.g. daily or hourly). These approaches periodically generate small portions of new data and consider these new data by automatically modifying the classifiers. For instance, DTs or the weights in an ensemble of DTs are periodically updated in [23,24]. The key concept of these updates is to include in the training database more recent information of renewable energy generation and demand, and the latest system topology. However, such periodic updates can be inefficient. For instance, updates may be done at times when actually no update would be needed or the small portions of new generated data may be insufficient to provide a good representation of the new system topology [22].

(iii) the third set of approaches updates the training database and re-trains a new prediction model in the real-time operation using a trigger (instead of using a fixed time-interval), as shown in Fig. 1. This training is triggered either every time the network topology changes [13] or as soon as the model performs poorly by tracking the accuracy in real-time [25]. These approaches are computationally impractical as the frequency of topology changes is increasing and it would require data generations and training very often. At each training only a few new data would be generated due to limited computational resources and, as for approaches (ii), these may be insufficient to provide a good representation of the new system topology. New approaches are building upon the approaches (iii) and make them computationally more practical by identifying new methods to trigger the online updates. In [26], an early-warning system composed by an ensemble of extreme learning machines is proposed to detect risky OCs, and in [27] such ensemble learning model adapts its structure according to the availability of the input features which results in higher tolerance against missing data or topology changes. A new variation of particle swarm optimization is used in [28] to quickly identify the security border close to the current operating condition and update the border to reflect changes in the system. Similarly, the confidence of the prediction is used in [29] for a

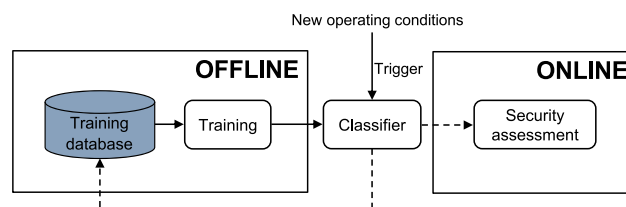


Fig. 1. Data-driven workflow for DSA based on triggered re-training.

conditional Bayesian deep auto-encoder to trigger new training when the prediction is of low confidence and cannot be trusted. Another data-driven based index is proposed in [30] to evaluate the confidence level of the security states one minute ahead of real-time operation and trigger updates of the DTs when the confidence level is low. This index uses a testing set of OCs in real-time to validate whether retraining is required. However, this index is based only on comparing data and is not informed by the type of the topological change itself. This can have a severe impact on system operator as a DSA classifier may be still being used although the accuracy dropped by orders of magnitudes without detecting such a drop.

Overall, it is challenging to know when a classifier requires retraining following a topology change, and it is challenging to effectively perform the subsequent data generation and retraining. For each retraining, a sufficient amount of training data needs to be considered to represent the new topological configuration, otherwise the classifier may not detect the change. The first challenge is to quantify the impact of a topology change on the performance of the classifier in the real-time operation. This quantification in real-time cannot involve time-domain simulations as only limited computational resources are available. The second challenge is the retraining after a high-impact topological change. The newly generated data are often redundant to the information content of the training database and the training database often contains information not relevant anymore to the new system topology.

### 1.3. Proposed approach

In this paper, a novel approach is proposed to address frequent changes in the network topology (Fig. 2). Firstly, for the first time, a metric for quantifying the impact of a topology change is proposed that considers the physical changes of the system instead of purely comparing the changes in the data. This is an informed approach that takes information into account that would have been missed-out by purely data-driven approaches. These additional information on system's physics make the proposed approach more robust against variations of the training database size, resulting in a lower sensitivity to the amount of training data needed to represent each topological change. In the proposed approach, the physical changes in the system are combined with the changes in the data by proposing a causality based feature selection (FS) approach able to capture the dependency between the system's transient stability and the network topology. Selecting such features that best represent the security states and do not miss any relevant information to prediction is necessary in large-scale systems, such as power systems, to guarantee high accuracy performance. In [21], energy function terms are used as a set of preprocessed meaningful input features, whereas the input features are selected using Fisher's

discriminant distance in [28]. The proposed FS approach uses the system's physics and data to discover the causal structure between features and then identifies highly relevant features by learning the approximate Markov Blanket (MB) on this causal structure. The novelty of the proposed FS approach is to physically inform the MB search for identifying dependencies and this goes beyond the original concept of the MB for feature selection that purely identified dependencies from data as in [31]. Therefore, changes in the probability distributions of selected features are good estimates of the impact of the topology change on the performance of the classifier, and hence good predictors of potential cascading failures. In practice, the proposed metric indirectly estimates how much a change in the system topology impacts the transient stability of the system itself. To subsequently create an efficient training database in response to a high-impact topology change, a convex hulls-based approach is proposed. This approach is used to assess the relevance of OCs based on their similarity. OCs from a previous database that are still relevant to the new topology are selected and new OCs are added that are not similar to those previously selected. These new OCs require then to carry out time-domain simulations to compute the security labels. The proposed workflow reduces the amount of newly generated data by making use of already existing data and filters out irrelevant information. It is more efficient in identifying high-impact topology changes and respond to them in order to make the data-driven DSA workflows more robust.

A case study on the IEEE 68 bus system considering transient stability is used to demonstrate the performance of the proposed workflow. First, the proposed metric is compared to existing methods for DSA to deal with topology changes. Then, the accuracy of classifiers trained on the databases constructed through the proposed approach is tested. The rest of the paper is structured as follows. In Section 2, the Markov Blanket-based FS approach is summarized. Thereafter, in Section 3, the proposed metric for quantifying the impact of a topology change and the method for the efficient construction of training databases are described. Subsequently, the case study is presented in Section 4 and conclusions are finally drawn in Section 5.

## 2. A Markov Blanket based feature selection approach

In this section, a Markov Blanket based FS approach that allows to capture the interactions between the system's dynamic stability and the network topology is described. This section sequentially introduces the final feature selection approach by starting with the graph model for voltages, and gradually introducing assumptions and modifications of the methods investigated.

### 2.1. Graphical model for power system voltages

The power network is defined as a physical graph  $G(V, \epsilon)$ , where  $V$  and  $\epsilon$  represent the buses and lines, respectively. Each bus  $i$  is associated to a random variable  $v_i$  [32], the voltage measurements. These voltage measurements are characterized by some conditional properties that make them suitable for an efficient representation of the grid topology through graphical models [33]. The variables considered are the voltage magnitudes of the buses. Hence, it is assumed that a state-estimation was performed beforehand [34]. The probabilistic relationships among voltage measurements are then described through a joint probability distribution:

$$p(\mathbf{v}) = p(v_1, v_2, \dots, v_n) = p(v_1)p(v_2|v_1)\dots p(v_n|v_1, \dots, v_{n-1}) \quad (1)$$

where  $v_i$  represents the voltage measurements at bus  $i$  and  $n$  is the number of buses. Evaluating this joint probability distribution is computationally expensive as the computational cost would be  $\mathcal{O}(m^{n-1})$  if  $m$  different voltage measurements are available. To reduce this cost,  $p(\mathbf{v})$  can be approximated by a simplified distribution  $p_a(\mathbf{v})$  while guaranteeing that the information loss is minimized.

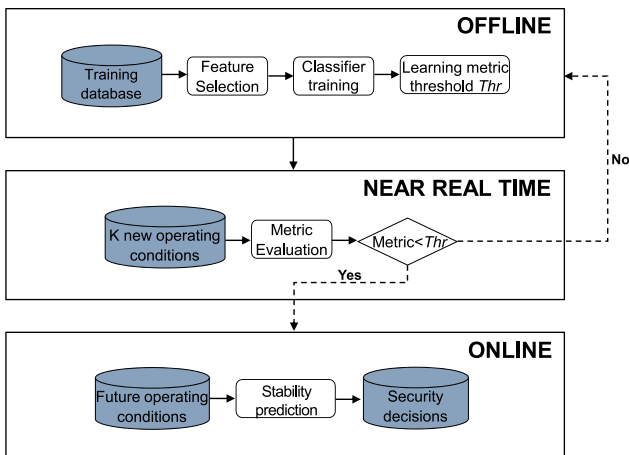


Fig. 2. Data-driven workflow for DSA to deal with topology changes.

## 2.2. Bayesian network

This section describes the construction of a Bayesian Network (BN) model where the correlation structure indirectly describes the grid topology. The tree-dependent probabilistic graphical model  $p_a(\mathbf{v})$  can be used as approximation of  $p(\mathbf{v})$  in Eq. (1):

$$p_a(\mathbf{v}) = \prod_{i=1}^n p(v_i | v_{pa(i)}) \quad (2)$$

where  $v_{pa(i)}$  is the direct predecessor, known as parent node, of  $v_i$ . In this model, voltages are conditionally independent given their parent nodes' voltage information if the current injections are independent [32]. In a transmission network, current injections can be approximated as independent as the voltages generally remain within the nominal range and the loads can be assumed as being independent. The Kullback–Leibler (KL) divergence is used to represent the difference of information contained in  $p(\mathbf{v})$  and those contained in  $p_a(\mathbf{v})$  about  $p(\mathbf{v})$ :

$$D(p||p_a) = E_{p(\mathbf{v})} \log \frac{p(\mathbf{v})}{p_a(\mathbf{v})} \quad (3)$$

The KL divergence should be minimal to minimize the information loss in approximating  $p(\mathbf{v})$  with  $p_a(\mathbf{v})$ . According to the well-known Chow-Liu algorithm [35], this minimization is optimized by constructing the maximum spanning tree in which branches of successively higher values of mutual information are selected and branches which involve loops are rejected. The resulting model is a Directed Acyclic Graph (DAG), known as Bayesian Network (BN) [36], where the correlation structure indirectly describes the grid topology.

## 2.3. Modified tree-augmented Naïve (TAN) bayes model

The learning approach for the BN is extended to learn the maximum likelihood Tree-Augmented Naïve (TAN) Bayes model as its causal dependence structure supports the search of an approximate Markov Blanket (AMB) of the classification target [37]. This extension of the BN is done by comparing the conditional mutual information between each  $v_i$  and the classification target  $C$  corresponding to the post-fault security status [38]. In this model each feature has as parents  $C$  and at most one other feature. However, the conditional dependencies of loops are neglected in such a model resulting in low accuracy performances when applied to highly meshed topologies such as in transmission networks. Therefore, considering the conditional dependencies of loops is particularly important to obtain a faithful model for transmission networks. At the same time, considering potential loops implies losing the causal dependence between features and causality is very relevant to the proposed MB-based approach as the MB can be identified only in a causal model. This issue is solved by the proposed modification of introducing an auxiliary variable  $e_l$  with fixed value for each loop  $l$ . The introduction of  $e_l$  does not change the probabilistic relationship between the parent nodes as it is clamped to a fixed value. Hence it results in

$$p(e_l) = 1 \Rightarrow p(v_{pa(e_l),1}, v_{pa(e_l),2}, e_l) = p(v_{pa(e_l),1}, v_{pa(e_l),2}) \quad (4)$$

where  $\{pa(e_l), 1\}, \{pa(e_l), 2\}$  represent the two parent nodes which are linked by the loop  $l$ . Then, for a general loop  $l$ , the probability distribution  $p_a$  is

$$p_a(\mathbf{v}) = p(e_l | v_{pa(e_l),1}, v_{pa(e_l),2}) \prod_{i=1}^n p(v_i | v_{pa(i)}) \quad (5)$$

Thus, by replacing Eq. (5) into Eq. (3), the divergence measure between  $p(\mathbf{v})$  and  $p_a(\mathbf{v})$  is

$$\begin{aligned} D(p||p_a) &= \sum p(\mathbf{v}) \log \frac{p(\mathbf{v})}{p_a(\mathbf{v})} = \sum p(\mathbf{v}) \log p(\mathbf{v}) - p(\mathbf{v}) \log p_a(\mathbf{v}) = \\ &= \sum p(\mathbf{v}) \log p(\mathbf{v}) - \sum p(\mathbf{v}) \log \prod_i^n p(v_i | v_{pa(i)}) \\ &= \sum p(\mathbf{v}) \log p(\mathbf{v}) - \sum p(\mathbf{v}) \log p(e_l | v_{pa(e_l),1}, v_{pa(e_l),2}) = \\ &= \sum p(\mathbf{v}) \log p(\mathbf{v}) - \sum p(\mathbf{v}) \sum_{i=1}^n \log \frac{p(v_i, v_{pa(i)})}{p(v_{pa(i)})} \\ &= \sum p(\mathbf{v}) \log p(\mathbf{v}) - \sum p(\mathbf{v}) \log \frac{p(e_l, v_{pa(e_l),1}, v_{pa(e_l),2})}{p(v_{pa(e_l),1})p(v_{pa(e_l),2})} \end{aligned} \quad (6)$$

By adding  $p(v_i)$  and  $p(e_l)$  inside the denominator

$$\begin{aligned} D(p||p_a) &= \sum p(\mathbf{v}) \log p(\mathbf{v}) - \sum p(\mathbf{v}) \sum_{i=1}^n \log p(v_i) - \\ &= \sum p(\mathbf{v}) \sum_{i=1}^n \log \frac{p(v_i, v_{pa(i)})}{p(v_i)p(v_{pa(i)})} - \sum p(\mathbf{v}) \log \frac{p(e_l, v_{pa(e_l),1}, v_{pa(e_l),2})}{p(e_l)p(v_{pa(e_l),1}, v_{pa(e_l),2})} \\ &= \sum p(\mathbf{v}) \log p(e_l) \end{aligned} \quad (7)$$

The last term of Eq. (7) is zero as  $p(e_l) = 1$ . Thus, the following equality holds:

$$D(p||p_a) = \sum_{i=1}^n H(V_i) - H(V_1, \dots, V_n) - \sum_{i=1}^n I(V_i; V_{pa(i)}) - I(e_l; v_{pa(e_l),1}, v_{pa(e_l),2}) \quad (8)$$

where  $I$  and  $H$  indicate the mutual information and entropy, respectively. The first two terms are both independent of the dependence tree, whereas the last two terms represent the branch weights. Minimizing the divergence measure is equivalent to maximize the total branch weight for both directed edges and loops.

## 2.4. Feature selection based on modified TAN model

The derived, modified TAN model includes directed edges and loops, and can be used for FS by taking advantage of its causal dependence structure to search for the approximate MB of the classification target  $AMB(C)$ . By performing pairwise comparison between each parent and children nodes, the AMB-based FS algorithm discards features which are irrelevant to classification. Algorithm 1 shows in detail how the identification of the AMB of  $C$  is performed. More specifically,  $\varepsilon$  is the set of directed edges, for which  $v_i$  is the parent of  $v_j$  and  $\tilde{\varepsilon}$  is the set of loops.  $AMB(C)$  and  $SP(C)$  are the approximate MB and spouse set of  $C$ , respectively. Firstly, all features with a higher relevance to  $C$  over the directed edges are included in the  $AMB(C)$ . Then, some of these features are removed by comparing the correlation to  $C$  over the loops. The final power flow features included in the  $AMB(C)$  have the highest relevance to classification for the given topology and hence, they provide the best representation of the interactions between the system's dynamic stability and the network topology. Thus, a change in the network topology necessarily impacts on the probability distributions of the selected features.

### Algorithm 1. AMB TAN based feature selection

**Initialization:**  $AMB(C) = \{\emptyset\}, SP(C) = \{\emptyset\}$   
**Output:**  $AMB(C)$   
1. **for**  $(i, j) \in \varepsilon$  **do** ▷ Pairwise comparisons over directed edges  
2.   **if**  $I(v_j; C) \geq I(v_i; C)$  &  $I(v_j; v_i | C) \geq I(v_i; C)$  **then**  
3.      $AMB(C) = AMB(C) \cup \{v_j\}$   
4.      $SP(C) = SP(C) \cup \{v_i\}$   
5.   **else**  
6.      $AMB(C) = AMB(C) \cup \{v_i\}$   
7.   **end if**  
8. **end for**

(continued on next page)

(continued)

```

9. for (i, j) ∈  $\bar{e}$  do ▷Pairwise comparisons over loops
10.   if  $I(v_j; C) \geq I(v_i; C)$  then
11.      $AMB(C) = AMB(C) - \{v_i\}$ 
12.   else if  $I(v_i; C) \geq I(v_j; C)$  then
13.      $AMB(C) = AMB(C) - \{v_j\}$ 
14.   end if
15. end for
16. for  $v_{pa(i)} \in SP(C)$  do ▷Deletion from spouses
17.   if  $v_i \notin AMB(C)$  then
18.      $SP(C) = SP(C) - \{v_{pa(i)}\}$ 
19.   end if
20. end for
21.  $AMB(C) = AMB(C) \cup SP(C)$  ▷Final AMB evaluation

```

### 3. Dealing with high-impact topology changes

In this section the proposed workflow for dealing with topology changes, i.e. the metric for quantifying the impact of a topology change and the method for creating a training database in response to a high-impact topological change, is described.

#### 3.1. Metric for detection of high-impact topology changes

This section describes how the metric to detect the high-impact topology changes using only input data is defined. A classifier is trained on the features selected through the MB-based FS. Since the selected features are the best representation of the interactions between the dynamic stability and the network topology, the change in their probability distributions, and hence in the probability distribution of the OCs, may provide an estimate of the classifier’s performance after changes in the system topology. However, choosing the most suitable metric to quantify these changes in the probability distributions is not trivial. Distance metrics, e.g Euclidean distance (ED), have been widely used to measure the similarities between probability distributions [30,39]. The ED between two OCs  $x_i$  and  $x_j$  originated from two distributions is calculated as follows:

$$ED(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{in} - x_{jn})^2} \quad (9)$$

where  $n$  is the number of features. However, in classification problems, the closeness in terms of ED does not necessarily correspond to similarities in terms of information contents. Two OCs may be close in terms of ED but belong to different classification regions. As shown in Fig. 3,  $x_i$  and  $x_j$  are closer than  $x_j$  and  $x_k$  in terms of ED but are more different in terms of information content as they belong to two different classification regions,  $C_0$  and  $C_1$ . To overcome the drawback of distance metrics in classification problems, the probability distributions of the OCs are

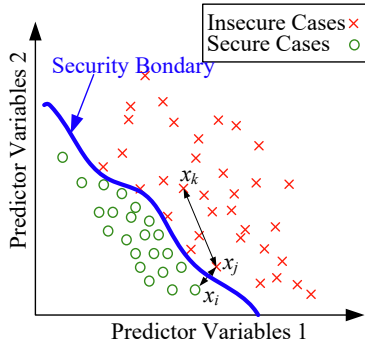


Fig. 3. The drawback of the Euclidean Distance [30]. The ED between  $x_i$  and  $x_j$  is lower than the ED between  $x_j$  and  $x_k$  although  $x_i$  and  $x_j$  are on different sides of the security boundary.

compared between subsets  $S_i$  and  $S_j$  containing the OCs with  $S_i \subseteq C_0$  and  $S_j \subseteq C_1$ . For example, the terminal nodes (or leaves) of the DT are used as subsets in this work. Then, the mean value over all the subsets is considered as metric. To make the comparison more accurate against small changes in the probability distribution of the OCs, convex hulls containing the OCs that end up in each leaf node are defined [40]. This means that two convex hulls containing the OCs are defined for each leaf node, one before the topology change occurs and the second one immediately after. Subsequently, these two convex hulls,  $X$  and  $Y$ , are compared through a well-known distance metric between the vertices, i. e. the Hausdorff distance  $d_H(X, Y)$ :

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\} \quad (10)$$

The two terms to calculate the Hausdorff distance are shown in Fig. 4. The aforementioned distance can be easily described as the greatest of all the distances from a point in set  $X$  to the closest point in set  $Y$ . Thus,  $d_H$  is calculated between the two convex hulls containing the OCs ending up in each leaf node of the classifier before and after the topological change occurs. Finally, the metric to quantify the impact of a topology change on the performance of the trained classifier is defined as the mean value of  $d_H$  over all leaves. The metric threshold is evaluated and calibrated for the trained contingency in the offline stage, and then used in the real-time operation.

#### 3.2. A generic metric for multiple contingencies

To calibrate the metric threshold in the offline stage, intensive time-domain simulations for different topologies should be performed. The whole process should be then repeated for several contingencies. In this section, a spectral clustering-based approach is used to identify similar electrical regions. An electrical region defines a set of buses, physically connected or not, that are electrically correlated. Then, the metric threshold of a trained contingency  $c_0$  can be generalized to other contingencies  $c_i$  not part of the training based on the regions to which the features of the  $AMB(C)$  of  $c_i$  belong.

**Definition 1:** If the  $AMB(C)$  of contingency  $c_i$  not part of the training includes features that are in the same electrical region of  $c_0$ , then the metric threshold trained for  $c_0$  can be still used for  $c_i$ .

A spectral clustering approach based on the admittance matrix is used to identify the electrical regions [41,42]. According to this approach, given the system admittance matrix  $Y_{bus}$ , the absolute values of the elements of the inverse matrix  $Y_{bus}^{-1}$  are used as a measure of the electrical distance  $D$ :

$$D = |Y_{bus}^{-1}| \quad (11)$$

with

$$D_{ij} = \begin{cases} d_{ij} & \text{if } j \text{ adjacent to } i \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

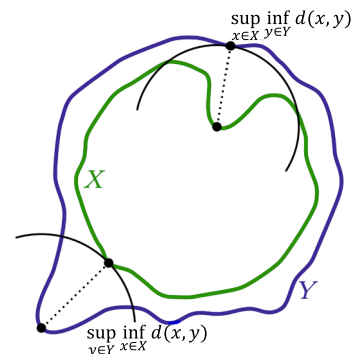


Fig. 4. The Hausdorff distance between the two convex hulls in green and blue.

To each bus  $i$  in the power network, the closest adjacent bus  $j$  in terms of  $d_{ij}$  is associated. Then, all the pairs that share common buses are grouped into regions. For each region  $S_k$ , the maximum contained  $d_{ij}$  is set as threshold  $r_k$ . Finally, for each pair of regions  $S_k$  and  $S_t$ , if  $d_{ij} \leq r_k$  with  $i \in S_k$  and  $j \in S_t$ , then  $S_k$  and  $S_t$  are merged. Hence, the final electrical regions of the network are identified through the spectral clustering based-approach and Definition 1 defines whether the generalization of the metric threshold is accurate or not.

### 3.3. Effective training database following topology changes

In this section, a convex-hulls based approach to efficiently construct new training databases after high-impact topology changes making use of the available database is described. Two convex hulls,  $H_0$  and  $H_i$ , containing OCs from a previous database and new OCs respectively, are defined. Algorithm 2 shows in detail how the new training database  $Tr$  is constructed, where  $i$  represents a general OC. Firstly, the OCs contained in  $H_0$  which are also contained in  $H_i$ , and hence are still relevant, are included in  $Tr$ . Secondly, the OCs contained in  $H_i$  which are not contained in  $H_0$ , and hence no similar to those already available, are included in  $Tr$ . Only for these last OCs, time-domain simulations are performed to compute the security labels.

#### Algorithm 2. Training database construction

Initialization:  $Tr = \{\emptyset\}$

Output:  $Tr$

1. for  $i \in H_0$  do
2.   if  $i \in H_i$  then
3.      $Tr = Tr \cup i$
4.   end if
5. end for
6. for  $i \in H_i$  do
7.   if  $i \notin H_0$  then
8.      $Tr = Tr \cup i$
9.   end if
10. end for

## 4. Case study

Several studies were undertaken to demonstrate the benefits of the proposed workflow to address frequent topology changes in data-driven DSA approaches. A case study on the IEEE 68-bus system where the security assessment involved transient stability was first used to study the effectiveness of the proposed metric for quantifying the impact of topology changes on DSA performance. Then, the performance of the proposed construction method for new training databases was investigated by comparing the prediction accuracy of the newly trained classifiers against conventional approaches for DSA. Finally, the computational savings of using the proposed workflow on larger systems were investigated.

### 4.1. Test system and assumptions

The IEEE 68-bus system (Fig. 5) was used as one of the three test systems [43]. A set of 20,000 OCs and sets of 10,000 OCs were generated for the reference topology and other 42 different topologies, respectively. The reference topology is the system topological configuration as shown in Fig. 5 and the other 42 topological configurations have each one disconnected line. All the potential topology changes that may occur in real-time, need to be considered offline to validate the approach. The OCs were generated by sampling the active loads from a multivariate Gaussian distribution with a Pearson's correlation coefficient  $c = 0.75$ . Then, by using the method of inverse transformation, the active loads were converted to a marginal Kumaraswamy distribution with the probability density function

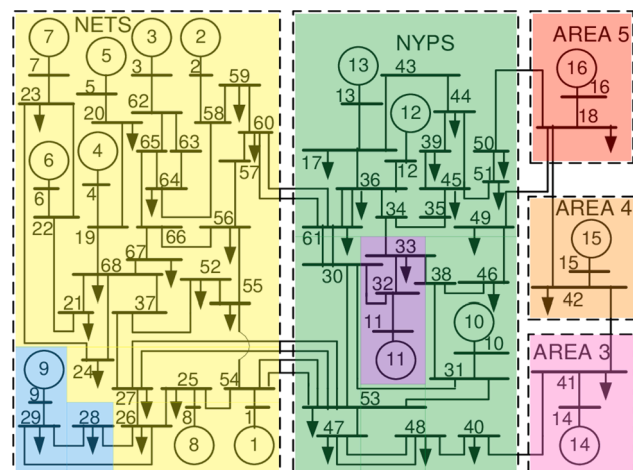


Fig. 5. The IEEE 68-bus system [43]. In colour the seven areas with similar characteristics identified through spectral clustering analysis.

$$f(x) = abx^{a-1}(1-x)^{b-1} \quad (13)$$

where  $a = 1.6$ ,  $b = 2.8$  are shape parameters and  $x \in [0, 1]$ . Finally, the active loads were scaled to be within  $\pm 50\%$  of the nominal values. The reactive powers follow the active powers proportionally as constant impedances were assumed. Subsequently, power factors were sampled i.i.d in the range of  $[0.95, 1]$  for each generator. Then, the full AC model was considered in a mathematical optimization problem to minimize the absolute differences to these power factors. Feasible OCs with set-points of active and reactive powers of the generators were obtained from this optimization. The optimization problem was implemented in Python 3.5.2 and Pyomo package and solved with IPOPT 3.12.4. The transients of three-phase faults over 9 different lines were simulated ( $k = 1, \dots, 9$ ) for all 42 topologies. A fault clearance time of 0.1s was used. If within 10 s simulation time all the differences between each two phase angles of the generators were less than  $180^\circ$ , then the OC  $i$  was considered transient stable  $Y_{i,k} = 1$ , otherwise unstable  $Y_{i,k} = 0$  and with that the security label was computed. Simulations were performed in Matlab R2016b Simulink. The resulting datasets for the various contingencies and topological configurations have class imbalances between 30% and 70%.

A second system, the IEEE 39-bus system, was used to investigate whether the low accuracies resulted from extreme operating scenarios considered in the training database. Therefore, as in the IEEE 68-bus system, a hard training problem with high load uncertainties was considered to generate the training data for 22 contingencies [44]. Finally, the French transmission system, corresponding to 1,955 transmission lines, 1,886 buses and 411 generators, was used to estimate potential benefits of computational savings when applying the proposed approach to larger systems. A set of 7,000 OCs and smaller sets of 1500 OCs were assumed to be available for each of the 1,000 potential topological changes in the offline and online (real-time operation) stages, respectively.

The machine learning workflow considered voltage magnitudes as features as mentioned in Section 2. Hence, for the IEEE-68 bus system, the pre-fault data for each OC were the values of the voltage magnitudes of the 68 buses. No modeling or simulation errors were considered, therefore the training OCs were assumed to be accurate. Subsequently, the AMB TAN approach was applied as pre-processing and CART was used to train the DT-based classifiers (as implemented in *scikit-learn*). In Table 1, the mean accuracy performance across all contingencies using DTs was compared against more advanced classification models to show that selecting DTs as models did not impact the final accuracies which are not very high as extreme operating scenarios are considered for the

**Table 1**  
Accuracy performance using different classification models.

	Classification model				
	DT	SVM	AdaBoost	XGBoost	ANN
Mean Accuracy	0.89	0.90	0.89	0.92	0.90

training. DT with depth equal to 3, SVM with linear kernel [45], AdaBoost and XGBoost with 50 estimators [46,47], and single layer feed-forward ANN with 10 neurons [8] were used. It resulted that all testing accuracies of these approaches were very similar, therefore DTs with maximum depth equal to the number of selected features through AMB TAN FS were preferred as they are more interpretable. All the following studies were repeated 10 times with different combinations of training/testing data at a split of 70%/30%. One DT was learned for each of the 9 contingencies based on the data of the reference topology. The DT was then tested against the 42 topology changes and new DTs were trained after high-impact topology changes. Finally, the case study on the IEEE 39-bus system was used to show that the low accuracies related to the extreme operating scenarios considered for the training. One DT was learned for each of the 22 contingencies. These DTs resulted in a low mean accuracy of 91% with a minimum value of 66%.

4.2. Detection of high-impact topology changes

In this study, the performance of three DSA approaches was tested under high-impact topology changes. (i) The first approach was a two-stages workflow where the machine was trained on the reference topology and then used in the real-time operation irrespective on any topological changes. (ii) The second approach was the three-stages workflow where a new machine was trained periodically by including 1, 500 new OCs in the training database. Once a topology change occurred, the operator selected randomly (uniformly) whether the classifier and database were being updated. (iii) The third approach is the proposed approach, the three-stages workflow where the proposed metric was used to decide whether retraining was needed. A new machine was trained only when the proposed metric was higher than the threshold. The other approaches mentioned in this paper were not tested as they were computationally not feasible. All classification models were trained offline on the generated 20,000 OCs for the reference topology and then the three approaches were tested against 42 topology changes on 1,500 new OCs generated from the power flow distribution following the system's change. The line contingency between bus 31 and 38 was randomly chosen to illustrate the benefits of the proposed workflow (iii) compared to the two existing approaches (i-ii). In the proposed approach the metric was calibrated offline using the generated 20,000 OCs for the reference topology and 7,000 OCs for each topology change.

The AMB TAN FS was used as pre-processing step in the offline stage for all approaches as it resulted in high prediction performance. To test whether the MB based FS approach selected the best predictors, two classifiers were trained and tested against the 42 topological changes: one classifier was trained using only the selected features, the other using all the features. The two classifiers resulted in the same accuracy performance for varying training database sizes. Thus, the MB based FS approach did not miss any relevant information to the prediction of potential cascading failures.

The results of the study were as follows: In the first approach (i), the reference DT classifier is used for all topological changes and no high-impact topological change can be detected. This approach resulted in accuracies lower than 92% for 14 of 42 topologies as presented in Fig. 6 (a) for the line contingency between bus 31 and 38. In the second approach (ii) based on periodic updates, in the best case, all high-impact topology changes were detected, however, in the worst case none of them were detected. No guarantees can be obtained as the results were random as new classifiers were trained for randomly selected topology changes. In the proposed third approach (iii), the relationship of metric

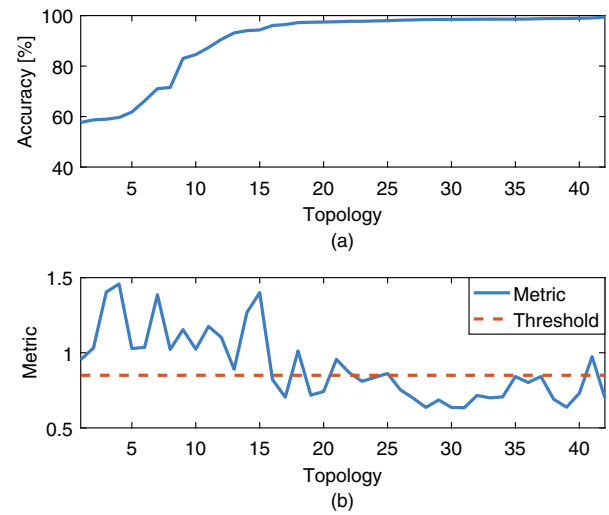


Fig. 6. The accuracy and metric of classifier corresponding to contingency on line 31 – 38 for 42 topologies. The topologies (x-axis) were ordered according to increasing accuracies.

and accuracy for the line contingency between bus 31 and 38 is shown in the two Figs. 6 (a)-(b). These results show that higher values of the metric corresponded to lower accuracies. A threshold equal to 0.85 corresponding to an unnormalized accuracy of 92% was defined and then used in the real-time operation to detect high-impact topology changes. The defined threshold depends on the training data but it is also related to the dependency between the system's stability and the reference topology. In the real-time operation, 17 high-impact topology changes were detected. However, for 3 topology changes the proposed metric was higher than the threshold even if the accuracy was higher than 92%, resulting in unnecessary training of new machines. For these 3 topologies, the classifier trained on the reference topology performed slightly better in accuracy than the reference topology itself, resulting in small differences in the power flow distributions that are detected by the metric. This issue is solved if the classifier trained on the reference topology has very high accuracy. However, the unnecessary training of only 3 machines is still acceptable as the cost of training a new machine in vain is significantly lower than providing unreliable security rules. The metric was then evaluated with varying sizes of the training database, i.e.  $|\Omega| = 1000, 4000$  (Fig. 7). The trained metric was nearly invariant in the studied range of database sizes, even when the training database size was reduced by 85%. Therefore, in the studied case, the metric and approach were robust against variations in the amount of training data needed to represent each topological configuration.

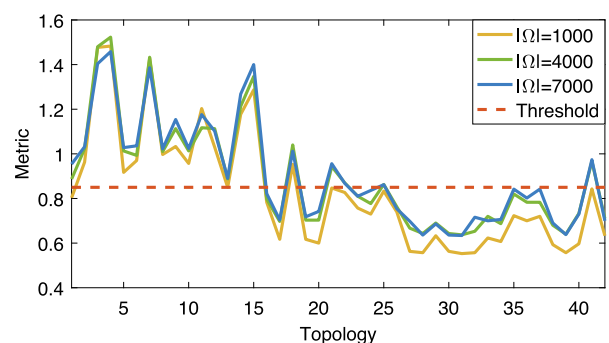


Fig. 7. The metric for classifier corresponding to contingency on line 31 – 38 for 42 topologies with varying training database sizes.



### 4.3. Detection for unseen contingencies

This section investigates the main benefit of the proposed approach to indicate whether the classifier and the metric threshold trained on one contingency can be used for other contingencies not part of the training. In the proposed approach, the MB is computed for the trained contingency (Section 3.1) and electrical regions are computed using the spectral clustering approach (Section 3.2). These 7 regions are highlighted in Fig. 5. Subsequently, if another contingency has common features to the MB of the trained contingency, then the classifier can still be used. If another contingency has most features of the MB in the same electrical region where the trained contingency occurs, then the metric threshold can still be used. The advantage of this approach is to not require any labels of the contingencies and uses only the features, hence no time-domain simulations are needed to validate the continued use of the machine learning workflow.

In this study, the contingency  $c_0$  between bus 31 and 38 was used as the reference, i.e. trained contingency. Subsequently, the tested contingencies showed a correlation between accuracy and common features in the MBs (Fig. 8). Then, the metric threshold of  $c_0$  was used for the tested contingencies. Across 8 tested contingencies, 5 of them had more than 3 features (i.e. half of the total number of features included in the largest considered MB) of their MBs located in the same electrical region of fault  $c_0$ . Only 20% of high-impact topology changes were missed for these contingencies compared to almost 50% of missed high-impact topology changes for the other contingencies.

### 4.4. Construction of new training databases

This study investigates whether the proposed approach can be used for the efficient construction of a new training database following a topology change. The proposed approach uses two sources, the existing knowledge database and newly generated data. These two sources are studied separately for the same line contingency (between bus 31 and 38).

#### 4.4.1. Utilization of the initial knowledge base

To utilize the existing database for training new machines reduces the number of time-domain simulations required in the real-time operation. Two approaches for constructing the training databases were compared in terms of predictive accuracy: (a) The database included all 20,000 OCs from the knowledge base and 1,500 new OCs. This approach used all available information and the created training database had 21,500 OCs. (b) The database included a selection of OCs (offline) from the knowledge base and 1,500 new OCs. This approach uses the proposed convex-hull approach to select OCs based on their power flow values. The results are shown in Table 2. The first approach resulted in an accuracy of 83%. The approach (b) increased the accuracy to 85%. Therefore, the proposed selection of high-quality data from the existing database corresponded to an improvement in accuracy of up-to 13% and average of 2%. The approach was able to select only OCs from the training database that were relevant for the new topology.

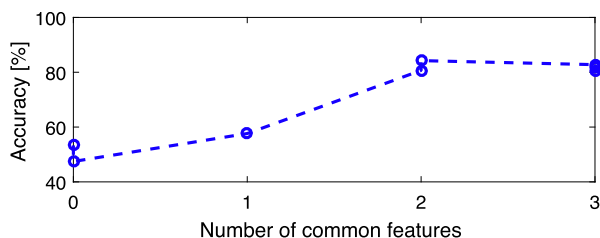


Fig. 8. The correlation between accuracy of the reference classifier  $c_0$  for different contingencies  $c_i$  and similarities in their MBs.

Table 2

The mean accuracy of the trained classifier under high-impact topology changes for three approaches (a)-(c) to construct new training databases.

	Approach		
	(a) All knowledge	(b) Relevant knowledge	(c) Generation of new OCs
$N^a$ offline OCs	20,000	4,307	4,307
$N^a$ new OCs	1,500	1,500	684
Accuracy	83.26%	85.54%	85.39%

#### 4.4.2. Generation of new data

An effective approach to generating new OCs is to considering only OCs that add knowledge to the database. In other words, OCs that are redundant are not generated. The proposed approach (c) uses convex-hulls to make this selection effective. This is the final proposed approach in this paper. The results in Table 2 show that the number of time-domain simulations were significantly reduced by 55% from 1,500 to 684. The other 816 OCs were considered redundant to the existing knowledge database. The result shows that the mean accuracy decreased only slightly by 0.15%. In the best case, only 224 OCs over 1,500 were selected. Hence, the time needed for time-domain simulations in real-time operation was strongly reduced and is promising in making the training computationally feasible.

#### 4.5. Training strategy

The training strategy of the proposed approach for DSA was compared to existing approaches (i-iii) introduced in Section 4.2, and additionally to a three-stages workflow using the proposed metric where the training is exclusively based on the new OCs (iv). Although this approach is computationally inefficient, it provides a theoretical upper limit for the accuracy achievable from new OCs. All approaches were tested on 1,500 new OCs generated from the power flow distribution following the topology changes. The results of the test accuracy are shown in Table 3 for the 17 high-impact topology changes. The following benefits of the proposed approach can be observed: an improvement in accuracy of up-to 52% and average of 9% in comparison to the two-stage workflow; a 2% average improvement in comparison to the three-stage workflow with periodic updates. However, if these periodic updates detected only the low-impact topology changes, the improvements was high at 9% in average and up-to 53%, and hence the proposed approach provided more robust results. The proposed approach also outperformed when only new training conditions are selected by 1.5%, hence using previous data improved the accuracy. In addition, the proposed approach reduced unnecessary time-domain simulations needed by 50% by selecting the new OCs that provide new knowledge to the database. The proposed workflow compared to existing approaches for DSA allowed operators to train new machines in real-time operation and enhanced the reliability of the security rules against frequent changes in the system topology.

The French transmission system was then used to estimate potential benefits of computational savings to larger systems. In this estimation, the same reductions of 85% for the offline training database size (Section 4.2), of 60% for new machines to be trained (Section 4.2) and of 55% for new data to be generated following high-impact topological changes (Section 4.4), as in the IEEE 68 bus system, were assumed. Subsequently, to estimate the computational benefits, the proposed approach (iii) was compared to the three-stages workflow (v) where the training is triggered at each topological change and is based on all the available new OCs. This approach is computationally inefficient as simulating the new OCs at each topological change would significantly increase the time span in which no accurate security assessment is in place. The results were summarised in Table 4. The estimation showed that the proposed approach has the potential to reduce the time for data generation by up-to 85% from 472 h to 71 h.

**Table 3**

Comparison between DSA workflows in terms of mean accuracy over high-impact topology changes.

	Approach			
	(i) Offline training	(ii) Periodic updates	(iii) Proposed workflow	(iv) Upper Limit
$N^{\text{a}}$ offline OCs	20,000	20,000	4,307	0
$N^{\text{a}}$ new OCs	0	1,500	684	1,500
Accuracy	76.62%	76.61% (w.c)	85.39%	83.85%

**Table 4**

Estimation of computational times for data generation for a large system using different DSA workflows.

	Approach	
	(iii) Proposed workflow	(v) Frequent regular updates
$N^{\text{a}}$ offline OCs	1,000,000	7,000,000
$N^{\text{a}}$ new near real-time machines	400	1,000
$N^{\text{a}}$ simulations per new machine	680	1,500
Simulation time per OC	0.2s	0.2s
Total	71h	472h

#### 4.6. Discussion

The proposed metric in combination with the construction method for new training databases showed promising results for online DSA applications, resulting in a maximal accuracy improvement of 52% compared against the conventional two-stages workflow for DSA. The metric detected all the 17 high-impact topology changes for the line contingency between bus 31 and 38. Thus, new machines needed to be trained only 17 times rather than training new machines for all 42 topology changes. Then, the number of time-domain simulations to be performed was more than halved as relevant information were first selected from the existing database by the proposed approach. Additionally, the proposed approach improved the mean accuracy by 9% against the two-stages workflow and by 9% and 2% against the worst and best case of the three-stages workflow based on periodic updates.

A few key limitations in designing data-driven DSA approaches for dealing with frequent system's changes still exist. A lot of data are still required as different topology changes can be considered, e.g. disconnected lines or switched off generators, and they may happen simultaneously, i.e.  $N-k$  contingencies. The robustness of the proposed metric against variations of the training database size and the high relevance of the selected features to failure prediction (Section 4.2) cannot be concluded for all types of failures, topology changes or system's settings. In the case of  $N-k$  contingencies, the approach can be extended by considering  $k$  faults in the offline time-domain simulations. Additional analysis should be conducted in this research direction to improve the applicability of the proposed approach. Moreover, to evaluate the metric in the real-time operation, it is necessary to guarantee that the measurements are sufficiently accurate for an efficient graphical modelling of the grid. The machine learning approaches that are used along the proposed workflow were selected based on their relevance in the literature and their choice does not affect the performance of the proposed workflow, e.g. classification models different from DTs can be used (Table 1). Although the metric is based on designing convex hulls for the terminal nodes of the DT, a similar approach can be extended to any prediction models. The proposed workflow should be also tested against other stability metrics to validate their use. The benefits in terms of computational savings for large systems that were estimated for the French transmission system, should be verified on a real test system to assess the scalability of the approach. Relying on machine learning based DSA workflows rather than investing in new grid infrastructure has a risk that should be considered in the decision making process.

#### 5. Conclusion

The challenges of dealing with high-impact topology changes for real-time DSA were investigated, showing that machine learning based DSA suffers from changes in the system topology. Neglecting these changes results in high inaccuracies of DSA classifiers, and low operational reliability. In response, a metric is proposed to identify topology changes that highly impact the classification accuracies. The key advancement of the approach is that the metric does not need any dynamic simulations, but only investigates the changes in the power flow features. This metric uses a causality-based feature selection approach for selecting features based on capturing the dependency between the system's transient stability and the network topology. Subsequently, the metric uses a convex hull-based approach to identify changes of data within the selected features.

The IEEE 68 bus system and transient stability were used to study the proposed approach. The metric correctly detected the 17 highest-impacting topology changes from a set of 42. Subsequently, only these 17 triggered retraining of the classifier, whereas other uninformed approaches would need to retrain 42 classifiers. The proposed approach improves the predictive accuracy by around 10% in average and up to 50% and, reduces the required training data by up to 85% which is the key finding. This approach allows to consider varying system topologies and marks a significant step forward to include dynamics in machine learning supported real-time DSA. The vision is promising as the system's operation closer to the physical (stability) limitations is more efficient. In the future, this workflow can be proposed for other stability phenomena, and in a control framework.

#### CRedit authorship contribution statement

**Federica Bellizio:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing - original draft. **Jochen L. Cremer:** Formal analysis, Resources, Supervision, Visualization, Writing - review & editing. **Goran Strbac:** Funding acquisition, Project administration, Supervision, Writing - review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work is supported by the Engineering and Physical Sciences Research Council under the Integrated Development of Low-Carbon Energy Systems programme. We are thankful to colleagues from Réseau de Transport d'Electricité who provided expertise that greatly assisted the research.

## References

- [1] Panciatici P, Bareaux G, Wehenkel L. Operating in the Fog: Security Management Under Uncertainty. *IEEE Power Energy Mag* 2012;10(5):40–9. <https://doi.org/10.1109/MPE.2012.2205318>.
- [2] Kundur P, Paserba J, Ajarapu V, Andersson G, Bose A, Canizares C, Hatziaargyriou N, Hill D, Stankovic A, Taylor C, Van Cutsem T, Vittal V. Definition and Classification of Power System Stability IEEE/CIGRE Joint Task Force on Stability Terms and Definitions. *IEEE Trans Power Syst* 2004;19(3):1387–401. <https://doi.org/10.1109/TPWRS.2004.825981>.
- [3] Strbac G, Hatziaargyriou N, Lopes JP, Moreira C, Dimeas A, Papadaskalopoulos D. Microgrids: Enhancing the Resilience of the European Megagrid. *IEEE Power Energy Mag* 2015;13(3):35–43. <https://doi.org/10.1109/MPE.2015.2397336>.
- [4] Duchesne L, Karangelos E, Wehenkel L. Recent Developments in Machine Learning for Energy Systems Reliability Management. *Proc IEEE* 2020;108(9):1656–76. <https://doi.org/10.1109/JPROC.2020.2988715>.
- [5] Sun M, Konstantelos I, Strbac G. A Deep Learning-Based Feature Extraction Framework for System Security Assessment. *IEEE Trans Smart Grid* 2019;10(5):5007–20. <https://doi.org/10.1109/TSG.2018.2873001>.
- [6] Konstantelos I, Jamgotchian G, Tindemans SH, Duchesne P, Cole S, Merckx C, Strbac G, Panciatici P. Implementation of a Massively Parallel Dynamic Security Assessment Platform for Large-Scale Grids. *IEEE Trans Smart Grid* 2017;8(3):1417–26. <https://doi.org/10.1109/TSG.2016.2606888>.
- [7] Wehenkel LA. *Automatic Learning Techniques in Power Systems*. Norwell, MA, USA: Kluwer Academic Publishers; 1998.
- [8] Zhang Y, Xu Y, Dong ZY, Zhang R. A Hierarchical Self-Adaptive Data-Analytics Method for Real-Time Power System Short-Term Voltage Stability Assessment. *IEEE Trans Industr Inf* 2019;15(1):74–84. <https://doi.org/10.1109/TII.2018.2829818>.
- [9] Zhou Y, Guo Q, Sun H, Yu Z, Wu J, Hao L. A novel data-driven approach for transient stability prediction of power systems considering the operational variability. *Int J Electr Power Energy Syst* 2019;107:379–94. <https://doi.org/10.1016/j.ijepes.2018.11.031>.
- [10] Chan KW, Edwards A, Dunn R, Daniels A. On-line dynamic security contingency screening using artificial neural networks. *IEE Proc-Generat Transmiss Distrib* 2000;147(6):367–72.
- [11] Cremer JL, Konstantelos I, Strbac G. From Optimization-based Machine Learning to Interpretable Security Rules for Operation. *IEEE Trans Power Syst* 2019;34(5):3826–36. <https://doi.org/10.1109/TPWRS.2019.2911598>.
- [12] Liu C, Sun K, Rather ZH, Chen Z, Bak CL, Thøgersen P, Lund P. A Systematic Approach for Dynamic Security Assessment and the Corresponding Preventive Control Scheme Based on Decision Trees. *IEEE Trans Power Syst* 2014;29(2):717–30. <https://doi.org/10.1109/TPWRS.2013.2283064>.
- [13] Donnot B, Donon B, Guyon I, Liu Z, Marot A, Panciatici P, Schoenauer M. LEAP nets for power grid perturbations. In: *ESANN 2019 - 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 2019.
- [14] Dehghanian P, Kezunovic M. Probabilistic Decision Making for the Bulk Power System Optimal Topology Control. *IEEE Trans Smart Grid* 2016;7(4):2071–81. <https://doi.org/10.1109/TSG.2016.2544794>.
- [15] Kezunovic M. Monitoring of Power System Topology in Real-Time. In: *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 10. IEEE; 2006. p. 244sb–244b.
- [16] Loparo KA, Abdel-Malek F. A Probabilistic Approach to Dynamic Power System Security. *IEEE Trans Circ Syst* 1990;37(6):787–98. <https://doi.org/10.1109/31.55036>.
- [17] He M, Zhang J, Vittal V. Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning. *IEEE Trans Power Syst* 2013;28(4):4089–98. <https://doi.org/10.1109/TPWRS.2013.2266617>.
- [18] Xu Y, Dong ZY, Zhao JH, Zhang P, Wong KP. A Reliable Intelligent System for Real-Time Dynamic Security Assessment of Power Systems. *IEEE Trans Power Syst* 2012;27(3):1253–63. <https://doi.org/10.1109/TPWRS.2012.2183899>.
- [19] Gomez FR, Rajapakse AD, Annakkage UD, Fernando IT. Support Vector Machine-Based Algorithm for Post-Fault Transient Stability Status Prediction Using Synchronized Measurements. *IEEE Trans Power Syst* 2010;26(3):1474–83.
- [20] Hashiesh F, Mostafa HE, Khatib A-R, Helal I, Mansour MM. An Intelligent Wide Area Synchrophasor Based System for Predicting and Mitigating Transient Instabilities. *IEEE Trans Smart Grid* 2012;3(2):645–52.
- [21] Geeganage J, Annakkage U, Weekes T, Archer BA. Application of Energy-Based Power System Features for Dynamic Security Assessment. *IEEE Trans Power Syst* 2014;30(4):1957–65.
- [22] Krishnan V, McCalley JD, Henry S, Issad S. Efficient Database Generation for Decision Tree Based Power System Security Assessment. *IEEE Trans Power Syst* 2011;26(4):2319–27. <https://doi.org/10.1109/TPWRS.2011.2112784>.
- [23] Sun K, Likhate S, Vittal V, Kolluri VS, Mandal S. An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees. *IEEE Trans Power Syst* 2007;22(4):1935–43. <https://doi.org/10.1109/TPWRS.2007.908476>.
- [24] He M, Zhang J, Vittal V. A Data Mining Framework for Online Dynamic Security Assessment: Decision Trees, Boosting, and Complexity Analysis. In: *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*; 2012. p. 1–8. <https://doi.org/10.1109/ISGT.2012.6175766>.
- [25] Diao R, Sun K, Vittal V, O'Keefe RJ, Richardson MR, Bhatt N, Stradford D, Sarawgi SK. Decision Tree-Based Online Voltage Security Assessment Using PMU Measurements. *IEEE Trans Power Syst* 2009;24(2):832–9. <https://doi.org/10.1109/TPWRS.2009.2016528>.
- [26] Zhang Y, Xu Y, Dong ZY, Xu Z, Wong KP. Intelligent Early Warning of Power System Dynamic Insecurity Risk: Toward Optimal Accuracy-Earliness Tradeoff. *IEEE Trans Industr Inf* 2017;13(5):2544–54.
- [27] Zhang Y, Xu Y, Zhang R, Dong ZY. A Missing-Data Tolerant Method for Data-Driven Short-Term Voltage Stability Assessment of Power Systems. *IEEE Trans Smart Grid* 2018;10(5):5663–74.
- [28] Kassabalidis IN, El-Sharkawi MA, Marks RJ, Moulin LS, Da Silva AA. Dynamic Security Border Identification Using Enhanced Particle Swarm Optimization. *IEEE Trans Power Syst* 2002;17(3):723–9.
- [29] Zhang T, Sun M, Cremer JL, Zhang N, Strbac G, Kang C. A Confidence-Aware Machine Learning Framework for Dynamic Security Assessment. *IEEE Trans Power Syst* 2021.
- [30] Liu C, Tang F, Leth Bak C. An Accurate Online Dynamic Security Assessment Scheme Based on Random Forest. *Energies* 2018;11(7). <https://doi.org/10.3390/en11071914>.
- [31] Koller D, Sahami M. Toward Optimal Feature Selection. *Tech. Rep.*, Stanford InfoLab; 1996.
- [32] Weng Y, Liao Y, Rajagopal R. Distributed Energy Resources Topology Identification via Graphical Modeling. *IEEE Trans Power Syst* 2017;32(4):2682–94. <https://doi.org/10.1109/TPWRS.2016.2628876>.
- [33] Wainwright MJ, Jordan MI. *Graphical Models, Exponential Families, and Variational Inference*. Now Publishers Inc; 2008.
- [34] Zhao J, et al. Roles of Dynamic State Estimation in Power System Modeling, Monitoring and Operation. *IEEE Trans Power Syst* 2020;1. <https://doi.org/10.1109/TPWRS.2020.3028047>.
- [35] Chow C, Liu C. Approximating discrete probability distributions with dependence trees. *IEEE Trans Inf Theory* 1968;14(3):462–7. <https://doi.org/10.1109/TIT.1968.1054142>.
- [36] Pearl J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier; 2014.
- [37] Yu L, Liu H. Efficient Feature Selection via Analysis of Relevance and Redundancy. *J Machine Learn Res* 2004;5:1205–24.
- [38] Friedman N. Building Classifiers Using Bayesian Networks. In: *Proceedings of the thirteenth national conference on artificial intelligence*. AAAI Press; 1996. p. 1277–84.
- [39] Donnat C, Holmes S. Tracking Network Dynamics: A Survey Using Graph Distances. *Ann Appl Stat* 2018;12(2):971–1012. <https://doi.org/10.1214/18-AOAS1176>.
- [40] Nalbantov G, Groenen P, Bioch C. Nearest Convex Hull Classification. *Tech. Rep.*, 2006.
- [41] Murray A, Keyessa M, Schurr P, Çakmak HK, Hagenmeyer V. A Comparison of Partitioning Strategies in AC Optimal Power Flow, CoRR abs/1911.11516.
- [42] Blumsack S, Hines P, Patel M, Barrows C, Sanchez EC. Defining Power Network Zones from Measures of Electrical Distance. In: *2009 IEEE Power & Energy Society General Meeting*; 2009. p. 1–8. <https://doi.org/10.1109/PES.2009.5275353>.
- [43] Pal B, Chaudhuri B. *Robust control in power systems*. Springer Science & Business Media; 2006.
- [44] Cremer JL, Konstantelos I, Strbac G, Tindemans SH. Sample-Derived Disjunctive Rules for Secure Power System Operation. In: *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*; 2018. p. 1–6. <https://doi.org/10.1109/PMAPS.2018.8440373>.
- [45] Moulin LS, da Silva APA, El-Sharkawi MA, Marks RJ. Support Vector Machines for Transient Stability Analysis of Large-Scale Power Systems. *IEEE Trans Power Syst* 2004;19(2):818–25. <https://doi.org/10.1109/TPWRS.2004.826018>.
- [46] Hastie T, Rosset S, Zhu J, Zou H. Multi-class AdaBoost, *Statistics and its Interface* 2009;2(3):349–60.
- [47] Chen M, Liu Q, Chen S, Liu Y, Zhang C, Liu R. XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System. *IEEE Access* 2019;7:13149–58. <https://doi.org/10.1109/ACCESS.2019.2893448>.