Imperial College of Science, Technology and Medicine
Department of Electrical and Electronic Engineering
Control and Power Research Group

Topological Changes in Data-Driven Dynamic Security Assessment for Power System Control

Federica Bellizio

Submitted in partial fulfilment of the requirements for the degree of of Doctor of Philosophy in Electrical and Electronics Engineering of the Diploma of Imperial College, April 2022

Abstract

The integration of renewable energy sources into the power system requires new operating paradigms. The higher uncertainty in generation and demand makes the operations much more dynamic than in the past. Novel operating approaches that consider these new dynamics are needed to operate the system close to its physical limits and fully utilise the existing grid assets. Otherwise, expensive investments in redundant grid infrastructure become necessary.

This thesis reviews the key role of digitalisation in the shift toward a decarbonised and decentralised power system. Algorithms based on advanced data analytic techniques and machine learning are investigated to operate the system assets at the full capacity while continuously assessing and controlling security. The impact of topological changes on the performance of these data-driven approaches is studied and algorithms to mitigate this impact are proposed. The relevance of this study resides in the increasingly higher frequency of topological changes in modern power systems and in the need to improve the reliability of digitalised approaches against such changes to reduce the risks of relying on them. A novel physics-informed approach to select the most relevant variables (or features) to the dynamic security of the system is first proposed and then used in two different three-stages workflows. In the first workflow, the proposed feature selection approach allows to train classification models from machine learning (or classifiers) close to real-time operation improving their accuracy and robustness against uncertainty. In the second workflow, the selected features are used to define a new metric to detect high-impact topological changes and train new classifiers in response to such changes.

Subsequently, the potential of corrective control for a dynamically secure operation is investigated. By using a neural network to learn the safety certificates for the post-fault system, the corrective control is combined with preventive control strategies to maintain the system security and at the same time reduce operational costs and carbon emissions. Finally, exemplary changes in assumptions for data-driven dynamic security assessment when moving from high inertia to low inertia systems are questioned, confirming that using machine learning based models will make significantly more sense in future systems. Future research directions in terms of data generation and model reliability of advanced digitalised approaches for dynamic security assessment and control are finally indicated.

Statement of originality and copyright declaration

I hereby declare that this thesis is the result of my own work, and that any ideas, concepts or results from the work of other people, published or otherwise, are appropriately referenced.

Federica Bellizio

Department of Electrical and Electronic Engineering

Imperial College of Science, Technology and Medicine, London, U.K.

Dedication

To the two men in my life, my father Remigio and my boyfriend Alessio.

Acknowledgements

This doctoral degree has been a very intense journey and has significantly changed my life. It would have been a completely different experience without the support of all amazing people around of me.

First of all, my gratitude goes to my supervisor, Prof. Goran Strbac, for providing me with guidance, supervision and this inspiring research project. Special thanks to Dr. Jochen Cremer for discussions and supervision of my research but most importantly for never stopping to believe in me and always being a role model for me. I would also like to thank Al-Amin Bugaje and Alicia Blatiak for their invaluable support during this journey, for the technical collaborations and for the incredible learning experience we shared as 'entrepreneurs'. I also thank my colleagues Haiwei Xie, Yunhe Wei, Dr. Luis Badesa and Dr. Mingyang Sun for their technical contributions, collaborative investigations and insightful discussions.

A big thank you goes to my 'Italian crew' at Imperial College London: Alex, Mirko, Domenico, Omar, Benita, and above all Angelo for still supporting me in my frequent gloomy days. A special thank you goes to Marco who has always supported me (and overestimated me), I would never have started my entrepreneurial journey without you stressing me from the first day we met. Thanks also to my friends in Italy, Anna, Danila, Federica and Rosa, who have always been there for me despite the distance. I wish to thank my family for their continuous support and for always being proud of my achievements. In particular, I am very grateful to my parents who have always pushed me to achieve my dreams, to my sister who has been a role model to me since I was a child, and to my beautiful niece Ludovica, who was born the same day I moved to London for this amazing experience. Finally but not least, my wholehearted thank you goes to Alessio for making me the woman I am today. Thank you for always being by my side and giving me the strength to purse my personal and career goals.

Contents

List of Tables

List of Figures

List of Acronyms

ACOPF Alternating Current Optimal Power Flow

AD Automatic Differentiation

AI Artificial Intelligence

AMB Approximate Markov Blanket

ANN Artificial Neural Network

BN Bayesian Network

CFS Correlation Feature Selection

CI Confidence Interval

CIG Converter Interfaced Generation

DAE Differential Algebraic Equation

DAG Directed Acyclic Graph

DER Distributed Energy Resources

DL Deep Learning

DSA Dynamic Security Assessment

DSO Distribution System Operator

DT Decision Tree

ED Euclidean Distance

ELM Extreme Learning Machine

EV Electric Vehicle

FS Feature Selection

GNN Graph Neural Network

ICT Information and Communication Technologies

ISGA Integral Square Generator Angle

JMI Joint Mutual Information

KL Kullback-Leibler

LEM Local Electricity Market

LQR Linear Quadratic Regulator

MB Markov Blanket

ML Machine Learning

MLP Multi Layer Perceptron

MRMR Minimum Redundancy Maximum Relevance

MTL Multi Task Learning

NERC North American Electric Reliability Council

OC Operating Condition

ODE Ordinary Differential Equation

OPF Optimal Power Flow

PINN Physics Informed Neural Network

PLL Phase Locked Loop

PMU Phasor Measurement Unit

RES Renewable Energy Sources

RF Random Forest

RFE Recursive Feature Elimination

RL Reinforcement Learning

ROA Region of Attraction

SCOPF Security Constrained Optimal Power Flow

SDP Semi Definite Programming

SFS Sequential Forward Selection

SSA Static Security Assessment

SU Symmetrical Uncertainty

SVM Support Vector Machine

TAM Transient Assistive Measure

TAN Tree Augmented Naïve

TSO Transmission System Operator

V2G Vehicle To Grid

Chapter 1

Introduction

1.1 Motivation

The Paris Agreement in 2015 has set out a global and joint effort to reduce the global carbon emission and limit the global warming. This objective was later confirmed in 2021 when nearly 200 countries have agreed to the Glasgow Climate Pact to finalise the outstanding elements of the Paris Agreement. The decarbonisation plan with net-zero target by 2050 involves not only electricity but also heating and transport sectors. Leading countries are now making large strides towards replacing the fossil-fuel generation with renewable energy, including the deployment of heat pumps instead of gas-fired boilers for residential heating and electric vehicles (EVs) instead of fuel-based vehicles for public/private transport. Following these global agreements, power systems are undergoing significant changes that will be even more far-reaching in the next few years. The power system was designed, decades ago, as a centralised unidirectional system of electric power transmission, electricity distribution and demand centers. The energy was traditionally transported from large power generation sites at high voltage levels through transmission lines and distributed to demand centers at low voltage levels through distribution lines. The balancing between generation and demand was easily met at all times, with an increase in demand more fuel was burnt. Therefore, the

generation was controllable and followed the demand in time. Nowadays, the rapid integration of Renewable Energy Sources (RES) and Distributed Energy Resources (DERs) at low and medium voltage levels and the increasing electricity demand pose new threats to modern power system operations. On one hand, the intermittent nature of renewable sources makes the energy generation less controllable and highly unpredictable, hence the balancing between generation and demand may be compromised. For example, solar power does not depend only on day/night hours but also on clouds. On the other hand, the new DERs at low voltage levels, such as rooftop solar panels or batteries, which store energy, increase the power flows and enable bi-directional flows (from demand centers to generation sites), making the demand highly flexible.

This increasing uncertainty that surrounds modern operations exposes the grid to new kinds of faults [11] and operators may not be able to comply with the security standards at all times. During normal operations, system operators must ensure that the equipment is not overloaded, meaning that the physical limits of the system are fulfilled at all times. At the same time, operators are required to immediately respond to potential disturbances to avoid emergency situations. Typical disturbances could be small, such as changes in load levels, that are generally addressed by automatic controllers, or major and sudden disturbances, such as equipment failures due to external events (e.g. weather) or internal events (e.g. line overloading). In the past operators were able to ensure a continuous and manageable service by properly designing the grid in the planning stage and using preventive control actions, however this is no longer sufficient. Redundancy has been mostly used as principle for proper planning designs; whenever one asset fails, a redundant asset continues operating maintaining the system in a secure operation. Therefore, most of the assets are not used at full capacity and this is highly cost-inefficient. If the same principle of redundancy is followed in the future, a number of expensive investments would be needed to cope with increasing operational uncertainty. An efficient operation of the system would include being close to its limits with smaller safety margins to increase the utilization of the existing assets and post-fault corrective control actions would be important for the reliability of the system [12]. New operating tools that allow for an efficient utilisation of existing assets rather than investing in new grid infrastructure are

investigated in this thesis. These operating tools open up the opportunity to take advantage of the new flexibility provided by end-users. The end-users, triggered by legislation and economic incentives, have started to invest in distributed generation, energy storage systems or more generally DERs, leading to a decentralisation of the generation and a flexible demand-side. Since DERs can balance electricity locally, they can either avoid network congestion and exposure to cyber-attacks or offer flexibility to the network to alleviate RES uncertainty [13]. Through Local Electricity Markets (LEMs), the end-users optimise their costs by trading energy locally and maximise their economic revenues by offering their flexibility to the grid. These local users, organised in smartgrids and microgrids, can also go further and operate off-grid, meaning that they are completely energy-independent from the grid and can self-sustain themselves with local generation and storage.

These local DERs can be controlled in a centralised or decentralised way. In centralised operation paradigms, a central controller employs network-level optimisation and provides system-wide optimal set-points for all controlled DERs. In decentralised approaches, only local information is used to dictate the DER response yielding to sub-optimal solutions compared to centralised methods, but resulting in a much easier implementation, reduced costs and higher robustness against cyber-attacks. For both approaches, an advanced Information and Communication Technology (ICT) network is needed to provide local and central controllers with reliable real-time data to be rapidly transmitted over the network. The recent advances in communication, above all wireless communication, and the large collection of data from Phasor Measurement Units (PMUs) have been the key drivers towards a full digitalisation of the power system. This digitalisation has opened up the opportunity to operate the system in real-time in a decentralised and 'smarter' way [14]. Therefore, the vision of future systems includes three interactive, smart blocks integrated each other through an advanced communication network and new digital technologies: control centers, transmission networks and substations [15]. According to this vision, such a complex system employs a unique digital platform that allows a more reliable, secure and cost-efficient operation using the real-time measurements from PMUs [16]. The development of this digital platform and of innovative control tools should be carried out

in parallel. During recent years, several techniques have been investigated for the development of these 'smart' controllers. Machine learning and game theory have been frequently used to accomplish the decision-making process in power system planning, operation and energy markets [17, 18]. Simultaneously, digital technologies such as Internet-of-Things, cloud computing and peer-to-peer applications, have significantly improved the connectivity and communication between the different distributed players [19–21]. The advent of blockchain, in particular, has enabled a more cyber-secure distribution of the data, while overcoming the typical problems affecting distributed databases, such as privacy or synchronisation issues [22].

In such a challenging and developing context, this thesis aims to investigate novel operating and control tools for a secure power system operation, which would enable a more efficient utilisation of the existing grid assets. These novel methods could be able to quickly assess the security of the system when a fault occurs and respond to this fault by changing the pre-/post-fault operation. The objective is to comply certain security criteria, for example N-1 security standards (operation reduced by one piece of equipment), at all times to avoid catastrophic events such as power blackouts. If these standards were easily met in the past, this is no longer the case as new dynamics and hence new kinds of faults, need to be assessed and controlled continuously, otherwise investments in redundant infrastructure are necessary to maintain the system within adequate security levels [11]. However, this is highly inefficient; a more efficient operation would be to operate the system close to its limits with smaller safety margins and actively use the post-fault corrective control actions for the management of the reliability, rather than using them as a backup strategies when preventive control fails [12]. Traditionally, numerical approaches were used to simulate the system's response to a fault and based on these simulations, operators assessed the dynamic security of the post-fault system in advance according to certain standard criteria. Following the Dynamic Security Assessment (DSA) of the post-fault system, operators determined the cost-optimal preventive control actions to adopt in order to maintain adequate security levels [23]. However, due to the high uncertainty surrounding Operating Conditions (OCs) and the large scale of modern systems, a vast number of potential OCs and faults would need to be simulated in advance which is impractical in real-

time operation. The computational complexity of the dynamic simulations is a key bottleneck of real-time DSA.

The recent digitalisation of the power system may play a key role in overcoming the computational bottleneck of DSA. The processing and analysis of historical observations (or data) may provide new insights to speed up the assessment and control tools and make them available in real-time operation. Approaches based on data, or partially based on data, are closely related to Machine Learning (ML) and Artificial Intelligence (AI) and have been widely investigated to address a number of power system challenges [17], including DSA and security control. However, due to a number of factors, such as privacy, customer sensitivities, fragmentation of data ownership and network regulation, the full potential of such data has not yet been exploited in the energy sector, compared to other economic sectors [14].

1.1.1 The machine learning approach to DSA

ML techniques have attracted attention as they may provide Transmission System Operators (TSOs) with a scalable way of managing uncertainty to make the system more reliable [9]. The key concept of the ML-based approaches is to first generate operating conditions similar to the ones in real-time and subsequently carry out the training process of ML models in an offline manner some days before operation. For instance, in the case of DSA, a vast number of potential operating conditions are simulated in the offline stage to model the residual uncertainties, e.g. possible network's configurations or different contingencies, between the day the classifier is trained and the day the classifier is used. Then, these predictors can be used in real-time operation to instantly infer the post-fault security status of unseen operating conditions [9, 24]. Due to the ability of classifiers to instantly infer the post-fault status, these approaches are promising for predicting various security phenomena under small and large disturbances, e.g. short-term voltage stability [25, 26], or dynamic voltage stability [27]. Various classification models have been investigated, such as Support Vector Machines (SVM) for transient stability analysis [28, 29], or k-nearest neighbours [30]. Different methodologies for modeling an

Artificial Neural Network (ANN) which allows the assessment of transient stability after a disturbance are discussed in [25, 31, 32]. Although ANN models perform better in terms of accuracy, Decision Trees (DTs) [33–35] or DT ensembles [4,36,37] have been mostly adopted as they provide a promising trade-off between accuracy performance, computational complexity and interpretability [24, 38, 39]. In particular, interpretability is a key requirement of data-driven security assessments as it ensures the ability to understand how a classifier predicts a particular OC with little inspection, allowing operators to be still involved in the control loop.

Although the ML-based approaches have shown promising performance when used for real-time operation [24, 35, 36], several challenges to address in order to improve their applicability to day-to-day operations still exist. These challenges mainly relate to: i) training of the ML-based model, ii) quality of the training data, iii) applicability to real-time operation. When training the classifier, operators must trade-off/prioritise several objectives, for instance high prediction accuracy irrespective of system's change, high interpretability to be able to mitigate unsuccessful cases [40] and minimisation of training time as the computational resources are limited. Similarly, the quality of the training data has to be very high to ensure high prediction performance. However, historical data are often imbalanced as typically there are many more samples from the secure class than from the insecure class [41–43]. A number of parameters are assumed to be fixed when classifying an operating condition as secure or insecure. However these parameters, e.g. fault clearance, can affect the security of the system, therefore the training data needs to take into account the sensitivity to such parameters. Also, missed alarms are more important to be avoided than false alarms, and this imbalance in the cost of wrong predictions needs to be considered [44]. Finally, when using these classifiers in the operations, the residual risk of relying on these 'black boxes' for either prediction or control needs to be managed [37, 45]. Hence, a shift from worst-case considerations of security toward a paradigm that allows for probabilistic security approaches is also needed [46].

1.1.2 The challenge of topological changes

A key challenge in data-driven DSA for system control is to make ML-based approaches consider changes in the network topology. Generally, topology changes refer to a set of actions, e.g. disconnecting lines, switching off generators or major (aggregated) batteries, or merging the substations [47,48], which are used for maintenance and control purposes. For instance, in cases of contingency, operators can use modifications of the network topology to re-direct power flows and protect the grid equipment. The line disconnection following a short-circuit is an example of a topology change to protect the grid assets. Only line disconnections are considered as topological changes in this thesis, however more complex topological changes can nowadays be performed as remedial actions. Similarly, only three-phase line faults are considered as contingencies. These faults only affect single lines and do not span to multiple lines or regions at the same time. The quality of these faults are full-line faults corresponding to an abrupt change of the full capacity of energy transmitted to a zero capacity being transmitted. However, due to the increasing operational uncertainty, new and more disruptive events can nowadays impact the grid, e.g. lost of generation or cascading failures.

The security of the power system is very sensitive to these changes [49] that are becoming very frequent in modern power systems. The reason is that the operating conditions following a topology change may be very different from those included in the training database as they originate from a different topology that is no longer representative and corresponds to a different power flow distribution [50]. The challenge is that if topological changes are not considered and security rules are trained only for one topology, then the assessment of the security using these rules may be wrong when a topology change occurs. From a machine learning perspective, the differences between the operating conditions can be seen as discrepancies between the training and testing probability distributions (covariate shift) [51]. Importance sampling is a useful approach to cope with these covariate shifts. The strategy of importance sampling is to maximize the weights of the data samples in the training set that have higher densities in the testing set. However, the estimation of these densities may not be a viable option for

applications to stability assessment and control as the frequency of probability changes in the power flow distributions is high as much as the frequency of equipment changes is high.

Several research efforts have been proposed in power systems to identify the relationship of the power flow distributions to the topology changes, and hence enhance the machine working workflow for DSA by considering changes in the system topology. These approaches can be classified in three sets [52]. (i) The first set of approaches considers updating the training database and the classifier periodically (e.g. daily or hourly). These approaches generate small portions of new data periodically and consider this new data by automatically modifying classifiers. The key concept of these updates is to include in the training database more recent information regarding the uncertainty deriving from the increasing integration of renewables. For instance, an ensemble of DTs is periodically updated in [33, 41, 53]. These approaches have two main problems. Firstly, the updates of both the training database and the classification model are undertaken very close to real-time operation, i.e. a few hours before operation. In this near real-time stage, more recent information regarding the residual uncertainty deriving from the increasing integration of renewables are included in the training database. The computational intensity of the updates in the near real-time stage is high as the system size requires a non-linear increase of dynamic simulations and this increase (more data) leads to a slower training process. To decrease the computational time in the training of the model, the dimension of the attributes can be reduced by applying Feature Selection (FS) techniques as a pre-processing step [50, 53]. The second problem is that the updates do not consider whether and how much the topological changes require updating being a steady periodic repetition. Hence, these approaches are inefficient as updating requirements are erratic. (ii) The second set of approaches generalizes the classifier to many topological configurations. The topological variations are considered in the generation of the training database by sampling many different system topologies [34, 41]. However, the dimension of topological (and equipment) configurations is so large that it is not feasible to consider each configuration in the training. (iii) The third set of approaches retrains a new prediction model in the real-time operation. This training is triggered either every time the network topology changes [47] or as soon as the model

performs poorly by tracking the accuracy in real-time [54]. The accuracy of these approaches is high; however, they are computationally impractical as the frequency of topology changes is increasing and it would require data generations and training very often.

Similar challenges relate to ML-based approaches for power system dynamic control. An operating condition that is classified as secure may be later found to be insecure when the contingency occurs. Therefore, control strategies that are designed for the pre-fault system may not be equally effective for the post-fault system. However, several approaches that use grid topology reconfiguration for control purposes have been proposed. For instance, transmission switching is nowadays used by operators to overcome various situations, such as voltage violations, line overloads [55], line losses and cost reduction [56], or system security [57]. Topology controllers have also been designed in [58, 59] by considering transmission switching within an Optimal Power Flow (OPF) to identify the optimal topology that minimises the operational costs and meets the security requirements. Because of its non-linear and combinatorial nature, no existing OPF solver can easily tackle this problem that is becoming more and more multidimensional. The challenge of high-dimensionality can be overcome using Reinforcement Learning (RL) to learn these topology controllers [60]. However, such RL-based controllers do not consider the dynamic response of the system, and therefore can only satisfy static security margins.

1.2 Objectives

In light of the above considerations arising from relevant research into state-of-the-art of ML-based approaches for the future need of power system operators to continuously monitor and control the grid [61], we pose the following research questions:

1. How can we inform ML-based DSA for power system control with the knowledge of network configuration in order to improve its robustness against increasingly frequent topological changes?

2. How can we use corrective control to actively enhance the system's dynamic security and reduce the operating costs?

3. Can we assume that state-of-the-art ML-based DSA for power system control will still perform well in future low inertia systems?

In the following section, we briefly illustrate how we try to answer these questions/challenges in this work.

1.3 Contributions

The first contribution of this work is addressing the challenge of the low robustness of ML-based approaches for DSA and control against topological changes. Our contribution is a novel approach to select the most relevant variables (or features in machine learning) to the classification task, which combines the causality concept and the physical knowledge of the system. Two combinations of this novel approach for feature selection with a three-stages workflow for DSA are then proposed as additional contributions that fulfil the objective of generalising the machine learning based approaches to different network configurations and making them more reliable for real-time applications.

The second contribution is a novel control approach based on the combination of preventive and corrective control strategies. Our contribution is to learn the safety certificates for the post-fault system by using a neural network and then incorporate such certificates within an optimal power flow considering both preventive and corrective control tools. The main results of this approach are significant steps forward towards more important and critical objectives such as operating cost reductions, carbon emission reductions and stability guarantees. The additional storage capacity from electric vehicles with vehicle-to-grid (V2G) technology is used for corrective control in this thesis.

Finally, the third contribution is a detailed analysis of the performance of machine learning-based approaches for DSA in future low inertia systems. The results provide an answer to the question whether data-driven approaches can still be used in the future, showing that they will be even more promising and their adoption will be even more crucial for secure and cost-efficient system operation.

1.4 Structure of the Thesis

Chapter 2 provides background on the stability and security assessment of the power system operation and finishes with fundamentals of the machine learning-based approach to DSA and control. Subsequently, Chapter 3 provides a detailed description of the novel approach for feature selection in combination with a three-stage workflow to increase the robustness of machine learning based approaches to DSA against uncertainty. Then, Chapter 4 describes the application of the proposed approach for feature selection to detect and deal with high-impact topological changes. In Chapter 5, the proposed control approach for dynamic security based on the combination of preventive and corrective strategies is described. Finally, in Chapter 6, the changes in the assumptions and conclusions for data-driven DSA when moving from high inertia to low inertia system are questioned. Chapter 7 concludes this thesis and provides outlooks for future research directions.

1.5 List of Publications

Some of the material presented in this work is also featured in the following works:

- Y.Wei, A. A. Bugaje, F.Bellizio, & G. Strbac. Reinforcement Learning Based Optimal Load Shedding for Transient Stability. IEEE PES Innovative Smart Grid Technologies, Europe (2022). *Accepted*

- F. Bellizio, J. L. Cremer, & G. Strbac. Transient Stable Corrective Control Using Neural Lyapunov Learning. IEEE Transactions on Power Systems (2022). *Accepted*

- A. Blatiak, F. Bellizio, L. Badesa, & G. Strbac. Value of Optimal Trip and Charging Scheduling of Commercial Electric Vehicle Fleets with Vehicle-to-Grid in Future Low Inertia Systems. Sustainable Energy, Grids and Networks 31 (2022): 100738.

- F. Bellizio, W. Xu, D. Qiu, Y. Ye, D. Papadaskalopoulos, J. L. Cremer, F. Teng, & G. Strbac. Transition to Digitalized Paradigms for Security Control and Decentralized Electricity Market. Proceedings of the IEEE (2022).

- F. Bellizio, A. A. Bugaje, J. L. Cremer, & G. Strbac. Verifying Machine Learning conclusions for securing Low Inertia systems. Sustainable Energy, Grids and Networks (2022): 100656.

- R. Segundo, et al. State-of-the-art of data collection, analytics, and future needs of transmission utilities worldwide to account for the continuous growth of sensing data. International Journal of Electrical Power & Energy Systems 137 (2022): 107772.

- F. Bellizio, J. L. Cremer, & G. Strbac. Machine-learned security assessment for changing system topologies. International Journal of Electrical Power & Energy Systems 134 (2022): 107380.

- F. Bellizio, J. L. Cremer, M. Sun, & G. Strbac. A causality based feature selection approach for data-driven dynamic security assessment. Electric Power Systems Research 201 (2021): 107537.

Chapter 2

Background

2.1 Introduction

Since modern societies rely on a constant supply of electricity, a reliable and continuous functioning of the power system is nowadays crucial. A power system is reliable when it can supply electricity with high enough probability to end-users at all times (i.e. adequacy) and withstand sudden disturbances without major service interruptions (i.e. security) [9]. The adequacy is evaluated over months or years through performance metrics such as expected energy not supplied or loss of load probability. On the other hand, security focuses on the operation of the system while it undergoes state transitions in the real-time. Typical disturbances may have different origins, for example natural disasters, equipment failures or human errors. When a fault occurs, the system can undergo five high-level operating states, that are shown in Fig. 2.1, according to the severeness of the fault: normal, alert, emergency, in extremis and restorative states [1].

All system variables are within normal ranges in normal operation and the system is able to withstand a fault when it occurs. The system enters the alert state when the security is below acceptable levels, and a potential disturbance that overloads the equipment places the system in

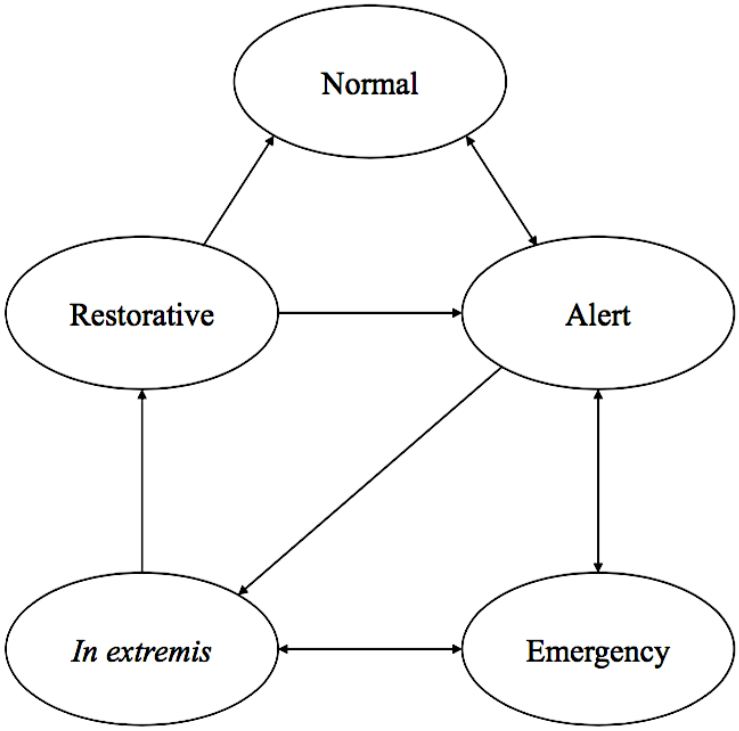

Figure 2.1: Power system operating states [1].

the emergency state. In the emergency state, some equipment operate at their emergency rating and voltage buses are out of their operating limits. In such conditions, operators apply some emergency measures to restore the alert state, for instance fast-valving, tripping generators, load curtailments or excitation controllers. If these measures are not effective, the system directly enters the in extremis state. The result of this condition are cascading outages and partial blackouts. Some control actions are then taken in the restorative state to restore the normal or alert state from the in extremis state. Typical control actions are load shedding and controlled system separation, which aim to reconnect all the facilities.

In this chapter, a review on the definitions of stability, security in power systems are first provided. Then, the concept of security assessment is introduced and the ML-approaches for security assessment and control are described. The objective of this chapter is to provide the basics to better understand the final aim of this work, that is DSA for power system control.

2.2 Power System Stability

Power system stability is the ability of the system to remain in an equilibrium state in normal operating conditions and to regain an acceptable state after being subjected to a disturbance. The disturbances may be small or large and understanding how the system responds to such disturbances is key in evaluating stability. For small disturbances, such as load changes, the system generally adjusts itself to the changing condition. On the other hand, large disturbances, such as short-circuits on a transmission line or loss of large generators, generally involve much of the equipment. For example, a short-circuit fault may trigger a sequence of cascading events: i) the isolation of the faulted component by protective relays causes variations in power transfers, machine rotor speeds and bus voltages; ii) the voltage and speed variations actuate the voltage regulators and the prime mover governors, respectively; iii) the changes in voltage and frequency affect the loads. However, a disturbance may be severe and result in instability even if it involves only few components. Traditionally, the instability was mainly related to the loss of synchronism between the different machines. This is no longer the case as the collapse of voltage and frequency can also make the system unstable, as in the example above. There are therefore different forms of instability that are classified according to the dominating phenomenon involved. These phenomena mainly surround the stability in the rotor angles, the stability in the interplay of power and angles, and the stability in voltages.

2.2.1 Rotor angle stability

The stability in the rotor angles relates to the synchronism of all synchronous machines in the power system. A synchronism machine has two main components, which are the field and the armature. In normal operations, the field is on the rotor and the armature is on the stator. In the case of a turbine as rotor, the rotor winding excited by a direct current generates a magnetic field that induces alternating voltages in the windings of the stator. If a load is connected to the stator, a current will flow in the windings. The frequency of this current depends on the speed

of the rotor and results in a rotating magnetic field. The rotor and stator fields react with each other resulting in an electromagnetic torque. In a generator, the mechanical torque must be applied continuously to the rotor via the prime mover as the electromagnetic torque opposes rotation of the rotor. When controlling the applied mechanical torque, the angular separation between the rotor and the revolving magnetic field changes, changing the power output of the generator. Under steady-state operating conditions, the rotor field and the revolving field of the stator have the same speed. Conversely, in a synchronous motor, the electromagnetic torque sustains rotation while mechanical load opposes rotation.

When connecting a generator to a synchronous motor, the power transfer is a function of the angular separation between the rotors of the two machines. The power transfer increases proportionally with the angle. However, for angles larger than 90°, the power transfer starts decreasing again. Similarly, when there are more than two machines, the power interchange depends on their relative angular displacements. However, in this case, an angular separation of 90° between any two machines has no particular significance as the limiting values of power transfers (and angular separation) are a much more complex function of generation and load distribution than when connecting only two machines. In a network with multiple machines, an operating condition is in steady-state if there is an equilibrium between the input mechanical torque and the output electrical torque of each machine, and their speeds remain constant. If the steady-state condition is disturbed, some machines will either accelerate or decelerate. If one machine runs faster than another one, the angular position between the rotors of the faster and slower machine will increase, resulting in a transfer of part of the load from the slower machine to the faster one. This transfer adjusts the angular difference and reduces the speed of the faster machine. However, beyond a certain limit, an increase in angular separation is accompanied by a decrease in power transfer that further increases the angular separation and leads to instability. This shift between the stator and rotor fields generates large fluctuations of power output, current, and voltage. Subsequently, the protection system reacts by isolating the unstable machine. The same loss of the synchronism can also involve sub-networks of the system with more than one machine.

Rotor angle stability can be classified in small-signal and transient/dynamic stability. Small-signal stability is the ability of the power system to maintain synchronism under small disturbances, such as small variations in load and generation. These disturbances may result either in a steady increase of the rotor angle due to lack of sufficient synchronizing torque, or in rotor oscillations due to lack of damping torque. The instabilities depends on a number of factors, such as the initial operating condition, the transmission system or the type of generator excitation controllers used.

Transient or dynamic stability is the ability of the power system to maintain synchronism when subjected to a severe disturbance and reach a new post-fault equilibrium that differs from that prior to the disturbance (pre-fault). The response of the system to such severe disturbances involves large excursions of generator rotor angles with highly non-linear relationship between power and rotor angles. However, the system is designed to be stable for a vast list of contingencies that usually occur on transmission lines, e.g. phase-to-ground, phase-to-phase-to-ground, or three-phase faults. When a fault occurs, the appropriate circuit breaker opens and isolates the faulty equipment. Fig. 2.2 shows the response of a single synchronous machine for stable and unstable operations. In Case 3, which is the stable case, the rotor angle increases to a maximum but then decreases with always smaller oscillation amplitudes. In Case 1, the rotor angle continues to increase until synchronism is lost. Finally, in Case 2 the system is stable in the first swing but becomes unstable as a result of growing oscillations as the end state is approached. This occurs when the post-fault steady-state condition is locally unstable. Therefore, transient instability may occur after several swings as a combination of several modes of oscillations. This means that long transient times need to be analysed to assess if a system is transient stable or not. For small systems, the study period of interest is usually limited to 3 to 5 seconds following the disturbance but this needs to be extended to 10 seconds for larger systems.

While transient stability assesses whether the system subjected to a disturbance can reach a new post-fault equilibrium, the steady-state stability only verifies whether the post-fault

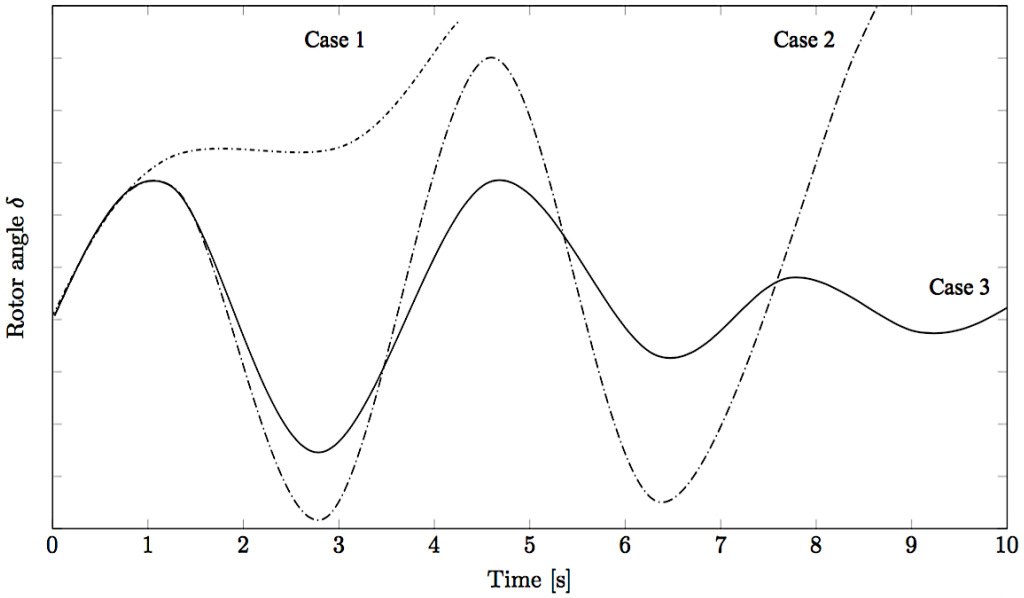

Figure 2.2: Rotor angle response to different transient disturbances [1]. Case $1-2$ are unstable, while Case 3 is transient stable.

equilibrium exists or not. This corresponds to verify whether the energy balance under the steady-state constraints of the network can be maintained. The difference between dynamic and steady-state stability is illustrated in Fig. 2.3, where the yellow and red areas represent the stability and instability region, respectively. Here, three different faults on the pre-fault steady-state equilibrium condition (a) lead to three types of post-fault conditions corresponding to stability/instability. The two dimensions describe the operational states that are represented using power levels of generation at two buses in this case. However, each operating condition is typically described using active and reactive power generation levels, active and reactive power loads and voltages magnitudes and phase angles at each system bus. More precisely, a different load scenario is considered for each operating condition and then static state estimation is used to evaluate power injections and power flows across the systems assuming a full system observability [62]. Under the assumption of pre-fault equilibrium conditions, the state estimation provides a unique solution to each input load condition. (b) and (b") are two steady-state post-fault equilibrium conditions, so they are both steady-state stable. However (b) is transient stable, while (b") is transient unstable as it cannot be reached. Finally, (b') is a transient and steady-state unstable as it is not a post-fault equilibrium and cannot be reached.

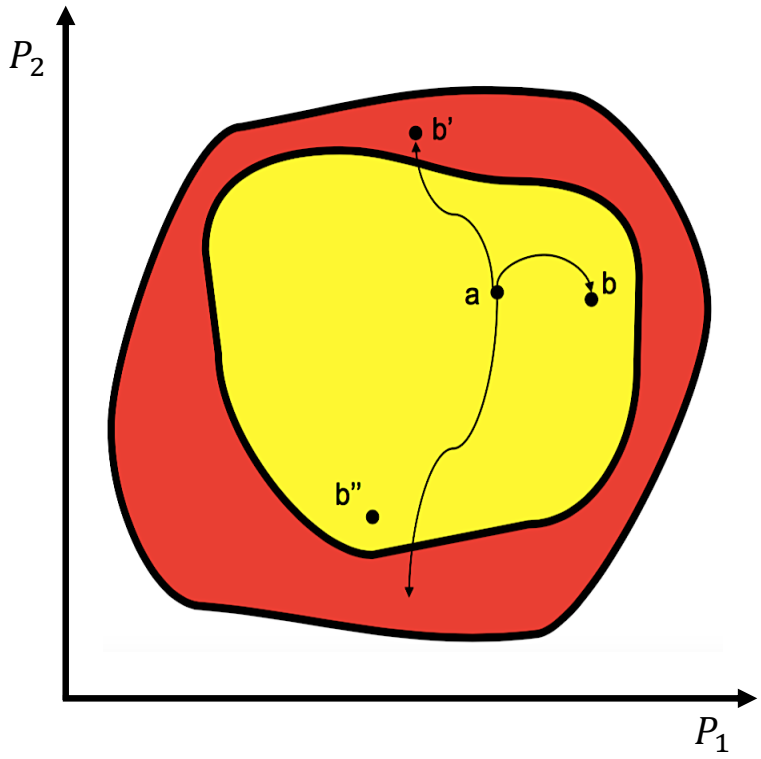

Figure 2.3: The difference between steady-state and transient stability [2].

2.2.2 Voltage stability

Voltage stability refers to the ability of a power system to maintain voltages at all buses within acceptable levels after being subjected to a disturbance. Typical disturbances that result in voltage instabilities are increases in loads or changes in the system condition that cause the voltage to drop. The origin of such instability is the voltage drop that occurs when active and reactive power flow through the inductive reactance of the network. The system is voltage stable if for every bus in the system, the increase of the bus voltage magnitude corresponds to an increase of the reactive power at the same bus. Voltage instability is generally a local event, however a voltage collapse, which is a voltage drop in a significant part of the system, may also occur as a result of a sequence of events.

According to the size of the disturbance, small and large disturbance voltage stability can be distinguished. The large disturbance stability refers to the ability of the system to maintain adequate voltage levels following large disturbances, such as loss of generation or line contin-

gencies. The assessment of such stability requires the analysis of the full system's dynamical response to the disturbance. Therefore, the period of interest could last from several seconds to tens of minutes. The small disturbance voltage stability refers to the ability of the system to control the voltages following small perturbations, such as load changes, and mainly depends on the load characteristics of the local controllers in place.

2.3 The Concept of Reliability and Security

A power system is reliable when it can supply electricity with high enough probability to end-users at all times, i.e. adequacy, and withstand sudden disturbances without major service interruptions in the real-time, i.e. security [9].

The adequacy is used in long-term planning applications as it focuses on the system dimensioning to accommodate the variability of the demand side. It is generally evaluated over long time periods that can last up to several years. The security is defined as "the ability of the bulk power electric system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system components" according to the North American Electric Reliability Council (NERC) [63]. The concept of security is narrower than stability as it requires that both the system remains stable during the transient response and all the operating limits are met pre- and post-fault. Security relates to a specified list of contingencies and depends on the operating limits that define the security criteria. The difference between the system's security and stability is illustrated in Fig. 2.4. As in Fig. 2.3, the two dimensions refer to the power levels of two generators. Each operating condition is described using both the load inputs and the system states and assuming that a static state estimation is performed beforehand [62]. The green area is the input subspace containing the operating conditions that are secure against a given list of contingencies, and hence conditions for which post-fault operating limits remain. The yellow region contains conditions that are steady-state stable but insecure as operating limits do not remain in post-fault, as for the shown trajectory starting from (a). The red region

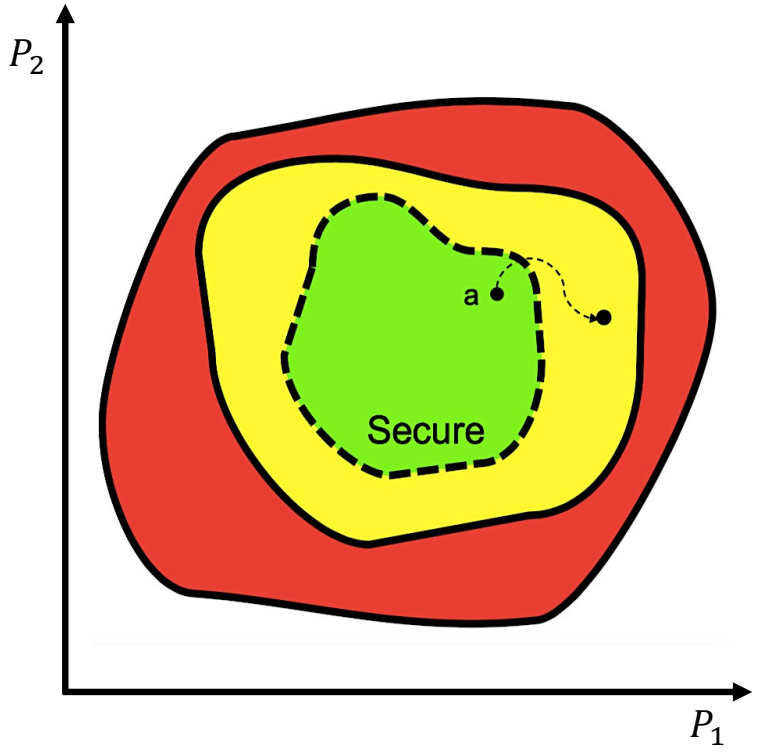

Figure 2.4: The difference between system's security and stability [2].

contains unstable and insecure operating conditions. In particular, the boundary between stable/unstable subspaces is a straight line being a feature of the system itself. On the other hand, the boundary between secure and stable subspaces is a dashed line as it can change according to the security criteria used and the list of contingencies considered. The list of contingencies generally includes all contingencies that are more likely to occur and can lead to high-impact failures in the operation. Nowadays, most worldwide transmission operators comply N-1 security standards at all times, that means operators must ensure a secure operation following each failure occurring in every single asset of the system.

The system security differentiates into static and dynamic security [9]. The static security refers to the system subjected to a disturbance fulfilling all physical constraints in the post-fault steady-state. However, the static security does not include whether the system survives the transition from pre-fault to post-fault. Such a transition is instead considered to be a dynamic security.

2.3.1 Security assessment

System's operators need to be aware of the current and future level of security (situational awareness) in real-time to decide on the optimal operating strategies and measures to adopt. The following security criteria for normal operation were established [63]:

- Pre- and post-fault flows on transmission lines must be within normal (or at most emergency) ratings;
- Pre- and post-fault bus voltages must remain within certain limits, which are normally ±5% of the nominal voltage level;
- Transient flows must be within certain transient limits obtained by running offline transient simulations and adding safety margins.

These requirements correspond in verifying whether the system survives the transient response and ends up in a steady-state post-fault condition for which all the assets remain within their operating limits. The analysis of the transients is termed the Dynamic Security Assessment (DSA), whereas the Steady-State Analysis is the Static Security Assessment (SSA or SA).

The static security can be assessed in real-time operation by modelling the energy balances for the post-fault static state with the faulted power system model, or directly the pre-fault for the N-1 system in a Security-Constrained Optimal Power Flow (SCOPF) problem, e.g. [64]. The OPF problem minimises the generation costs subject to the network constraints including line flow limits, the balance of the network nodes and the physical feasibility (Kirchhoff's laws). However, the associated transients of a contingency may be unacceptable, even if the operating condition is secure in the steady-state, as shown in Fig. 2.4.

Assessing the wide range of transients involved in the operation is challenging. Hence, the dynamic security is often not considered, instead operators place large safety margins on their operating equipment to ensure that the dynamic response is secure. This results in economic

loss as the system's assets are underutilised for most of the time and operators invest in new redundant infrastructure. To relax these static margins and utilise the grid assets in a more efficient manner, evaluating the dynamic security of the system is necessary. The consideration of dynamic security increases the situational awareness and allows to operate closer to the operating limits, reducing the new investments. To assess the dynamic security, all the dynamical phenomena, i.e. rotor angle, voltage and frequency stability, need to be studied separately [65]. However, these studies involve computationally intensive simulations. For instance, transient stability is evaluated by performing an event-type simulation on a large model involving ordinary differential equations. Such a simulation requires numerical integration that requires long computational times. Hence, assessing the dynamic security of a large electrical system is computationally intensive and would require significant online computational resources, specifically if a vast number of potential operating conditions and contingencies are to be considered.

The number of contingencies to consider in DSA is a critical issue. The more contingencies are considered, the higher the situational awareness. In addition to single faults, cascading faults should be also taken into accounts. However, since these cascading events are less likely to occur, most operators nowadays comply with the N-1 security standards. According to how operators deal with the vast number of potential contingencies, two approaches for security assessment can be distinguished: deterministic and probabilistic security assessment [66]. In the deterministic security assessment, operators only consider a short-list of severe and most likely contingencies against which only static security is assessed [67]. The list is generally decided according to the system's configuration, or the operator experience and knowledge of the grid. For example, N-1 criterion is also a deterministic security assessment approach where the list includes all the single failures of the assets. The problem with the deterministic approach is that the list may not consider those contingencies that are unlikely to occur but that correspond to very high risk. Furthermore, the risk and the impact of a contingency are time-dependent as the asset degrades, the weather and system's network change over the years. When using a deterministic approach, operators accept the potential impact of non-short listed contingencies, nevertheless this is the the current practice of most operators.

In the probabilistic security assessment, the risk of all contingencies are evaluated according to their probability and impact, and the highest-impact risks are considered in the assessment of the security [68,69]. Since the risks are time-dependent, their estimation is a very difficult task. For instance, the impacts are evaluated according to the violation of the operational limits in [70], or the probabilities are estimated according to the actual weather predictions in [71]. The probabilistic approach was mainly studied for SSA, however is not widely used in practice.

2.4 Machine Learning Based Approaches

Most current real-time software tools are limited to only the assessment of the static security of a shortlist of system faults. With current software, the assessment of the dynamic security is infeasible in real-time as they require time-domain simulations which have long computational times as they rely on numerical integration [65, 72]. In a similar way, the system also requires fast control operating tools that can be applied in real-time to control the system's dynamic security following potential disrupting faults. However, this would require implementing the system's dynamic response within an Alternating Current OPF (ACOPF). This implementation is very challenging and takes long computational times when conventional approaches are used, for instance Forward Euler or Runge-Kutta methods [73, 74].

ML techniques have attracted attention as they may provide TSOs with a scalable way of managing reliability [9]. The key concept of ML based approaches is to first generate operating conditions similar to the ones in real-time and carry out the training process of the ML models in an offline manner some days before operation. Subsequently, the offline models are used in real-time to assess (dynamic) security of many possible fault scenarios and control the operating conditions. The key advantage of such approaches is they provide real-time security predictions and control decisions 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

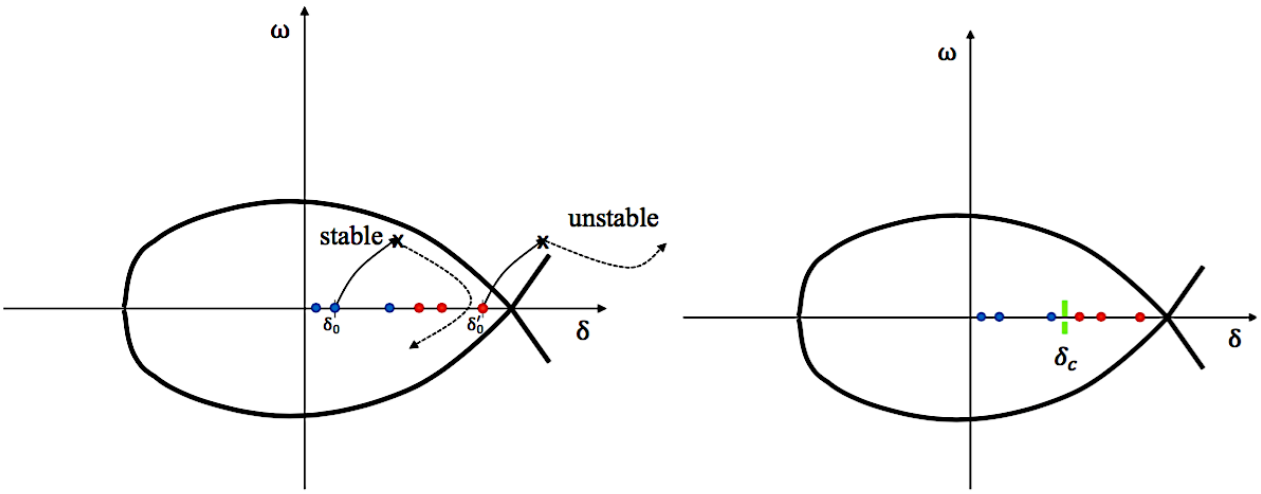

Figure 2.5: The offline workflow of machine learning approach to DSA [2].

infrastructure, e.g. switches or redundant lines, to satisfy the reliability requirements under contingencies [17]. As an example, the basic idea of using machine learning for system's security is shown in Fig. 2.5 using a "one-machine-infinite-bus" system. In Fig. 2.5 (a), pre-fault steady state OCs are sampled and the response to a given contingency is simulated. After the time-domain simulations, the pre-fault OCs are grouped in two classes according to their response to the fault and a different label is assigned for secure and insecure OCs. For instance, δ_0 is secure and δ_0' is insecure. The training database is finally built considering the sampled OCs and their respective labels. As second step in Fig. 2.5 (b), the classifier δ_c that separates the secure and insecure OCs is trained using the generated training database. This classifier can be then used in real-time for instant security predictions.

The assessment and control of the system's security can be described as statistical classification and regression problems [13], respectively from ML. Let $\mathbf{x}(t) = [\mathbf{x}^L(t), \mathbf{x}^G(t), \mathbf{x}^V(t)]$ the state variables describing the loads (active/reactive loads), generation (active/reactive injected powers) and voltages with $\mathbf{x}^L \in \mathbb{R}^L \subseteq \mathbb{R}^n, \mathbf{x}^G \in \mathbb{R}^G \subseteq \mathbb{R}^n$ and $\mathbf{x}^V \in \mathbb{R}^V \subseteq \mathbb{R}^n$ at time step t and n the number of buses. However, the time dependence is omitted to make the notation

clearer. ε_{NET} is the network's physical interconnectivity that is defined as follows:

$$\varepsilon_{NET,ij} = \begin{cases} 1 & \text{if } i,j \text{ adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where $i, j \in [1, .., n]$ are two different buses, The system's security can be expressed as a function $f_a(\mathbf{x}, \varepsilon_{NET})$ where

$$f_a : (\mathbf{x}, \varepsilon_{NET}) \longrightarrow y_a = \begin{cases} 1 & \text{insecure} \\ 0 & \text{secure.} \end{cases} \tag{2.2}$$

Similarly, the optimal controller for system's security is a function $f_c(\mathbf{x}^L, \varepsilon_{NET})$ that

$$f_c : (\mathbf{x}^L, \varepsilon_{NET}) \longrightarrow (\mathbf{x}^G_{opt}, \mathbf{x}^V_{opt}), \tag{2.3}$$

where $(\mathbf{x}^G_{opt}, \mathbf{x}^V_{opt})$ are the cost-optimal generator settings fulfilling all power system constraints and security criteria. For large systems, the assessment and control functions f_a, f_c are often highly non-linear and non-convex, and hence finding (and evaluating) these functions can be challenging.

The approach of ML is to learn approximating functions

$$\begin{aligned} \tilde{f}_a &: (\mathbf{x}, \varepsilon_{NET}) \longrightarrow \tilde{y}_a \\ \tilde{f}_c &: (\mathbf{x}^L, \varepsilon_{NET}) \longrightarrow \tilde{\mathbf{y}}_c = [\tilde{\mathbf{x}}^G_{opt}, \tilde{\mathbf{x}}^V_{opt}] \end{aligned} \tag{2.4}$$

such that

$$||y_a - \tilde{y_a}||_p, \ ||\mathbf{y}_c - \tilde{\mathbf{y}}_c||_p \quad \text{are minimised.} \tag{2.5}$$

To learn these functions with supervised learning requires creating, as a first step, a database (X, Y) that includes OCs $\mathbf{x}$ (or $\mathbf{x}^L$ for security control) from historical observations and synthetically generated data, including their respective security labels y_a (or optimal generator settings y_c). The aim of using synthetically generated data is to enrich the variability of the

training database by considering extreme OCs that are unlikely to occur frequently but result in high-impact failures. The generation of a pre-fault OCs is very challenging as it requires sampling in a space fulfilling all the non-linear constraints of the AC model (feasible space). There are three different approaches for data generation: historical sampling, importance sampling and generic sampling. The first approach creates samples that follow the same probability distribution of a stochastic model fitted to the historical observations [75]. Importance sampling focuses on the sampling of OCs that have a higher relevance for the learning, and hence can add knowledge to the existing training database [38]. Finally, generic sampling aims to create samples that span the feasible space uniformly,

The supervised learning of the functions $\tilde{f}_a$ and $\tilde{f}_c$ involves then splitting the database (X, Y) into training (X_{train}, Y_{train}) and testing dataset (X_{test}, Y_{test}) with $X_{train}, X_{test} \subseteq X$, $Y_{train}, Y_{test} \subseteq Y$ such that $X_{train} \cap X_{test} = \emptyset$, $Y_{train} \cap Y_{test} = \emptyset$. A ML-algorithm learns the function $\tilde{f}_{train}$ from the training set

$$\tilde{f}_{train} : \mathbf{x}_{train} \longrightarrow \tilde{y}_{train} \tag{2.6}$$

with $\mathbf{x}_{train} \in X_{train}$ such that

$$|y_{test} - \tilde{f}_{train}(\mathbf{x}_{test})| \quad \text{is minimised} \tag{2.7}$$

with $\mathbf{x}_{test} \in X_{test}, \mathbf{y}_{test} \in Y_{test}$.

Some choices on the candidate function $\tilde{f}_{train}$ (e.g., parametrisation) need to be made before applying the ML training algorithm. One typical choice is to consider "rules" organised hierarchically and sequentially to predict the label/value [76]. For instance, many researchers have used DTs to learn these rules, a popular algorithm is CART (more example algorithms can be found in [9]), or neural networks (NNs). For a binary classification problem, let $\Omega^+ = \{\mathbf{x} \in \mathbb{R}^n : f_a(\mathbf{x}, \varepsilon) = 1\}$ with $\varepsilon = \varepsilon_{NET}$ and approximate it as the union of N convex polytopes $\tilde{\Omega}^+ = \cup_{i=1}^{N} P_i$ with $P_i = \cap_{j=1}^{M_i} H_{ij}$ of M_i half-spaces $H_{ij} = \{\mathbf{x} \in \mathbb{R}^n : h_{ij}(\mathbf{x}, \varepsilon) > 0\}$.

h_{ij} is the indicator function defined as follows:

$$h_{ij}(\mathbf{x}, \varepsilon) = \begin{cases} 1 & \sum_{k=1}^{n} \omega_{ij,k} x_k + b_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

where $\omega_{ij,k}$ and b_{ij} are the weights and the bias term, respectively. Hence, the approximation function $\tilde{f}_a$ in Eq. (2.4) can be written as a disjunction of conjunctions, also known as disjunctive form:

$$\tilde{f}_a(\mathbf{x}, \varepsilon) = \bigvee_{i=1}^{N} \left(\bigwedge_{j=1}^{N} h_{ij}(\mathbf{x}, \varepsilon) \right) \tag{2.9}$$

such that $\tilde{\Omega}^+ = \{\mathbf{x} \in \mathbb{R}^n : \tilde{f}_a(\mathbf{x}, \varepsilon) = 1\}$ and $\tilde{\Omega}^+$ is an approximation of Ω^+. In a similar way, $\tilde{f}_c$ can be also written as a disjunctive form as follows:

$$\tilde{f}_c(\mathbf{x}^L, \varepsilon) = \bigvee_{i=1}^{N} \left(\bigwedge_{j=1}^{N} h_{ij}(\mathbf{x}^L, \varepsilon) \right) \tag{2.10}$$

with $\tilde{\Omega}^+ = \{\mathbf{x} \in \mathbb{R}^L : \tilde{f}_c(\mathbf{x}, \varepsilon) = 1\}$ and $n_1 = |\tilde{\Omega}^+|$. In this case, the predicted values are calculated as follows:

$$\tilde{\mathbf{y}}_c = \frac{\sum_{k=1}^{n_1} \omega_{ij,k} x_k^L + b_{ij}}{n_1} \tag{2.11}$$

Different indicator functions $h_{ij}(\mathbf{x}, \varepsilon)$ correspond to different ML models, for example a NN can be defined using indicators in Eq. (2.8) but the same indicators, without the bias terms b_{ij}, can be used to represent a DT model as a set of split functions [77]. The two approximation functions $\tilde{f}_a$, $\tilde{f}_c$ can be also integrated in a single Multi-Task Learning (MTL) model, for example a NN in which the total loss $J(\alpha)$ is the weighted sum of the loss function of the two tasks [78]:

$$J(\alpha) = \alpha_1 |y_a - \tilde{y}_a| + \alpha_2 |\mathbf{y}_c - \tilde{\mathbf{y}}_c| \tag{2.12}$$

with $\alpha = [\alpha_1, \alpha_2]$ the set of the loss weights for the two tasks.

The predictive performance of the ML models for system's dynamic security and control is

generally measured using the accuracy or F_1-score for DSA, and distance metrics, such as the mean squared error, for control predictions. In the case of DSA, the standard test accuracy is defined by the ratio of accurately predicted cases divided by all predicted cases. The F_1-score is generally used in class imbalance problems and is defined as follows:

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{2.13}$$

where the true positives (TP) are the positive (secure) conditions predicted as positive, false positive (FP), or missed alarms, are the negative (insecure) conditions predicted as positive and finally the false negative (FN), or false alarms, are the positive conditions predicted as negative.

Not only maximising predicting performance is important, but also the interpretability and generalisation capability to other network reconfigurations (following topological changes). Although NNs generally provide high accurate predictions, DTs or ensemble of DTs have been mostly adopted for power system applications, above all for DSA, as they are more interpretable than NNs. Interpretability is important to build up the trust of operators in these approaches. High model interpretability supports operators in their understanding as to how a model predicts and maintains the security with little inspection, allowing operators to be still involved in the control loop [24, 39]. However, even the most accurate and interpretable model may no longer work properly when the system's configuration is different to the training configuration. As shown in Eq. (2.4), different ML-models are suitable for different ε_{NET}. The low generalisability of such models is a key barrier for their application to power systems and it is the focus of this thesis.

Chapter 3

Training Data Pre-processing

In machine learning based DSA the offline database is periodically updated to consider more recent security information [79]. However, the time distance between the offline and the online stage can compromise the performance of the classifier, irrespective on the type of classifier used. In fact, OCs can change very rapidly over time, resulting in current OCs which are different from those included in the initial knowledge base [11, 31]. The OCs from the online and offline stages can be described as originated from two different probability distributions. Importance sampling is generally a very useful approach to cope with these discrepancies between the training and testing distributions [80]. However, due to the high frequency of probability changes in the OCs, the importance factor cannot be estimated a priori in DSA applications. To track these changes, several efforts have been directed towards a periodic update (e.g. daily or hourly) of both the training database and the classification model. These updates represent a challenging task as they are undertaken very close to real-time operation, i.e. a few hours before operation. In this near real-time stage, more recent information regarding the residual uncertainty deriving from the increasing integration of renewables are included in the training database. The computational intensity of the updates in the near real-time stage is high as the system size requires a non-linear increase of dynamic simulations and this increase (more data) leads to a slower training process. Hence, one strategy for re-training

the model is proposed in [33, 41, 53], where only the data weights or small portions in an ensemble DT-based model are updated rather than to re-train the full DT. To additionally decrease the computational time in the training of the model, the dimension of the attributes can be reduced by applying feature selection techniques as a pre-processing step [50, 53]. The existing feature selection methods can be broadly classified into filter, wrapper and embedded methods [81, 82]. As wrapper and embedded methods require the highest computational times across these three classes, they may not be suitable to online DSA [41]. A few filter methods have been used in the past in DSA, resulting in a moderate trade-off between computations and accuracies [83]. Recently, in machine learning research, causality-based feature selection is investigated as causal features can improve the robustness and the interpretability of the predictive model across different settings [84, 85]. These approaches use independence tests to discover the causality between features from data. However, for power system DSA, the physical interconnectivity and the dependency between features are highly related [49]. Considering the system's physical knowledge and not only the data may improve the performance in terms of accuracy, computations and interpretability of causality-based FS. This is what this work investigates, causal feature selection for power system DSA.

3.0.1 Challenges of designing feature selection to DSA

The first challenge of feature selection for machine learning based DSA refers to their computation efficiency when highly relevant features should be selected [83]. If features are not selected effectively, the classifier results in inaccuracies. Simultaneously, if the feature selection requires a long time, it needs to be done early in the workflow, resulting in discrepancies between offline and real-time so large that the classifier becomes inaccurate. The challenge is to make efficient selection decisions and at the same time keeping the computation time of the selection process as short as possible. Thus, highly relevant features can be selected close to real-time, minimising the discrepancies between offline and real-time.

The second challenge has to do with the robustness of the selected features and the trained

classifier against the residual uncertainty between generated and real-time OCs [82]. The real-time OCs will always be (slightly) different from those included in the training database, even if the database generation and classifier training are moved very close to real-time operation. This difference is mainly caused by the high operational uncertainty as the underlying probability distributions of the OCs can change over a few hours [86]. Hence, it is important to identify highly relevant features for improving the robustness against uncertain OCs. The variability of the OCs represents a key challenge in the wider machine learning approaches to reliability assessment [9].

3.0.2 Proposed approach

A fast approach for FS to address the two aforementioned challenges is proposed. This approach is used in combination with a three-stages learning workflow (Fig. 3.1) that differs from the traditional two-stages scheme for DSA by the presence of a near real-time stage in between. The proposed approach uses the system's physics and data to discover the causal structure between features. Then, the approach identifies highly relevant features by learning the Markov Blanket (MB) on the derived structure. The key novelty is to inform the Markov Blanket search with the physical interconnectivity of the power system. The proposed approach is beyond using the concept of causality for feature selection [87] and brings a key advantage over other FS approaches that are exclusively data-driven. The correlation structure between the features and dynamic security is highly related to the network topology [49]. Hence, performing the FS based on the system's physics allows identifying features most relevant to security for the given topology. The proposed FS approach results in the highest accuracies and requires less computational times. The proposed workflow has two key benefits:

1. The accuracy of the final classifier is higher as the training is shifted closer to real-time operation, lowering the impact of discrepancies between offline training and real-time operation.

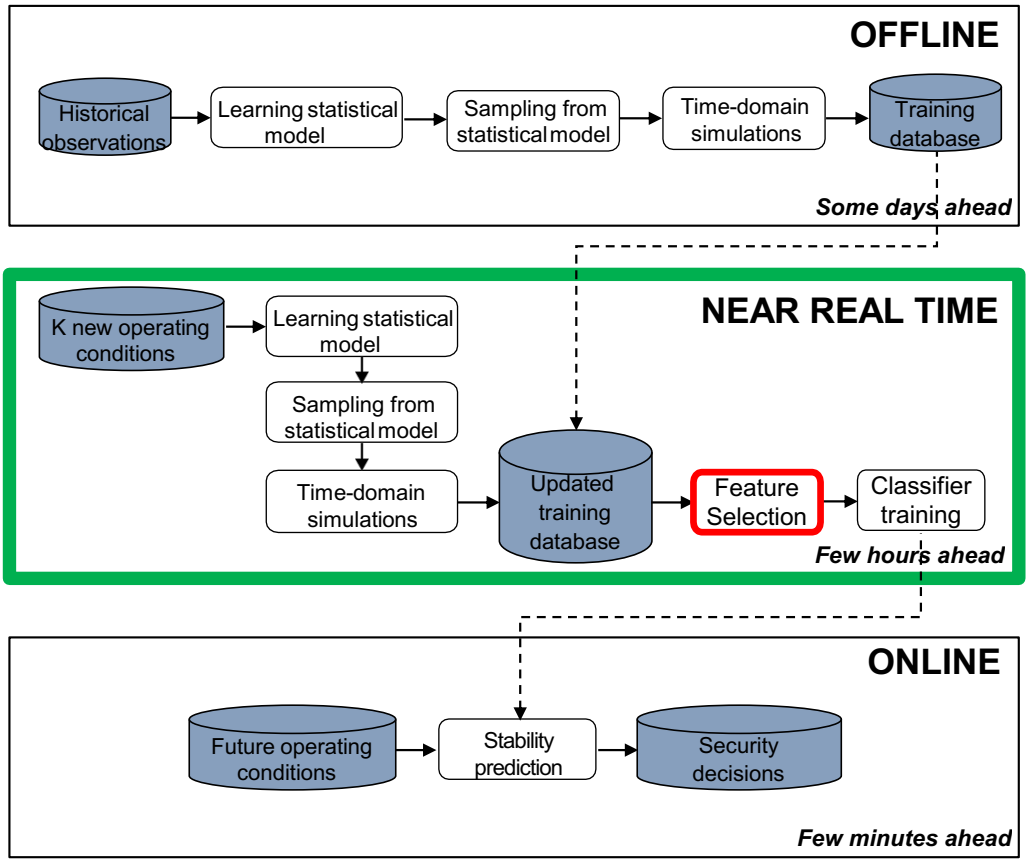

Figure 3.1: Data-driven three-stages workflow for classification. In green the proposed workflow and in red the proposed FS approach.

2. The robustness against uncertainties is higher as causal features are selected based on physical knowledge and not purely based on data.

A case study on the IEEE 68-bus system considering transient stability is used to illustrate the performance of the proposed workflow. First, the proposed approach is compared to a two-stage workflow for DSA. Then, the robustness of the classifier is tested on several datasets with varying discrepancies between estimated and actual parameters of the probability distributions of the OCs. Subsequently, the trade-off between prediction performance and computational complexity is investigated. The scalability to a large-scale system is projected with the French Transmission system confirming the potential of the approach.

3.1 Probabilistic Graphical Model

To combine the system's physics with a causality-based FS approach, a probabilistic graphical model that captures the statistical correlations according to the physical connections of the network topology is constructed [88]. The power network is defined as a physical graph $G(V, \epsilon)$, where V and ϵ represent the buses and lines, respectively. According to this approach, the voltage measurements obtained by smart meters over time at bus i are associated to a random variable v_i. The following basis justify the focus on voltage measurements: i) These measurements have become more accessible in recent years with usage of high fidelity PMUs. ii) Nodal voltage measurements are characterized by some conditional correlation properties that make them more suitable to an efficient representation of the network topology through graphical models [89].

To describe the probabilistic relationships among the voltage variables v_i, the measurements of each bus voltage associated to v_i are quantized and the joint probability distribution between such variables is computed as follows:

$$p(\mathbf{v}) = p(v_1, v_2, ..., v_n) = p(v_1)p(v_2|v_1)...p(v_n|v_1, ..., v_{n-1}) \tag{3.1}$$

where v_i represents the voltage measurements at bus i and n is the number of buses. Bus 0 is the reference bus with a unit magnitude and a zero phase angle voltage. If the quantization of a bus voltage has m levels, calculating this joint probability distribution is computationally expensive as the computational cost would be $\mathcal{O}(m^{n-1})$. To reduce this cost, a simplified distribution $p_a(\mathbf{v})$ can be used to approximate the true distribution $p(\mathbf{v})$ if the minimum information loss is guaranteed. A tree-dependent probabilistic graphical model can be chosen as approximation of $p(\mathbf{v})$:

$$p_a(\mathbf{v}) = \prod_{i=1}^{n} p(v_i|v_{pa(i)}) \tag{3.2}$$

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. This condition holds in

a transmission network if the current injections are independent [88]. As the voltages generally remain within the nominal range and the loads can be assumed as being independent, also the current injections can be approximated as 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 $p(\mathbf{v})$:

$$\begin{aligned} D(p||p_a) =& E_{p(\mathbf{v})}\log\frac{p(\mathbf{v})}{p_a(\mathbf{v})} = \sum p(\mathbf{v})\log\frac{p(\mathbf{v})}{p_a(\mathbf{v})} = \\ &= \sum p(\mathbf{v})\log p(\mathbf{v}) - \sum p(\mathbf{v})\sum_{i=1}^{n}\log p(v_i) \\ &\quad - \sum p(\mathbf{v})\sum_{i=1}^{n}\log\frac{p(v_i, v_{pa(i)})}{p(v_i)p(v_{pa(i)})} = \\ &= \sum_{i=1}^{n} H(V_i) - H(V_1, .., V_n) - \sum_{i=1}^{n} I(V_i; V_{pa(i)}) \end{aligned} \tag{3.3}$$

where I and H indicate the mutual information and entropy, respectively.

Minimizing the KL divergence is equivalent to minimize the information loss when $p(\mathbf{v})$ is approximated with $p_a(\mathbf{v})$. By following the Chow-Liu algorithm [90], the maximum spanning tree algorithm, which is based on mutual information, provides the optimal approximation of $p(\mathbf{v})$ in terms of this minimization. More precisely, the maximum spanning tree is constructed by selecting branches of successively higher values of mutual information and rejecting all branches that involve loops. Once the branches with highest mutual information are identified, $p_a(\mathbf{v})$ can be readily formed. Then, the undirected graph is transformed to a directed graph by choosing a root variable and setting the direction of all edges to be outgoing from it (the choice of root variable does not change the log-likelihood of the network). The resulting model is a Directed Acyclic Graph (DAG), known as Bayesian Network (BN) [91], where the correlation structure indirectly describes the grid topology. The variables considered are the voltage magnitudes of the buses. Hence, it is assumed that a state-estimation was performed beforehand [62]. It is also assumed that the network line parameters are fixed.

3.1.1 Tree Augmented Naïve Bayes structure

The Chow-Liu algorithm can be extended to learn the maximum likelihood Tree-Augmented Naïve Bayes (TAN) structure instead of the BN. In this model, each feature has as parents the classification target C and at most one other feature. In this study, C corresponds to the post-fault security state and the features represent the bus voltage measurements. The maximum spanning tree is constructed by comparing the conditional mutual information between each v_i and its parent given C [92].

3.1.2 Causal dependence in loopy structures

The described Chow-Liu algorithm for learning the TAN model shows low performance when applied to power networks with highly mesh topologies, e.g. transmission network, as it neglects the conditional dependencies of loops. However, taking into account possible loops implies losing the causal dependence between features. The causal dependence between features is crucial to the proposed MB-based approach as the MB can be identified only in a causal model. Two conditions are necessary to learn causal models from data [91]:

- *Markov Condition*: Each attribute $v_i \in \mathbf{v}$ is conditionally independent of its non-effect attributes given its direct cause attributes.
- *Faithfulness*: $\exists$ DAG G that is a perfect map for the probability distribution P of $\mathbf{v}$.

The Markov condition is guaranteed by the independence of current injections. On the other hand, considering loops implies losing the faithfulness assumption as the model would contain cycles. This issue can be solved by introducing an auxiliary variable with fixed value for each loop [91]. An example is provided in Fig. 3.2, where an auxiliary variable E is introduced between variables A and C in order to describe the conditional dependence between them respecting the causality of the previously learned model where A is cause of C. Since E is

clamped to a fixed value, then it results:

$$p(E) = 1 \implies p(A, C, E) = p(A, C) \tag{3.4}$$

Thus, the introduction of E does not change the probabilistic relationship between A and C, and hence the faithfulness assumption is still satisfied.

By using the auxiliary variable method, the learned TAN model with probability distribution p_a can now include a general loop l that has been neglected during the learning without losing the causality condition. Therefore, the probability distribution p_a of such an enriched TAN model is:

$$p_a(\mathbf{v}) = p(e_l|v_{pa(e_l),1}, v_{pa(e_l),2}) \prod_{i=1}^{n} p(v_i|v_{pa(i)}) \tag{3.5}$$

where e_l is the auxiliary variable, which is associated to the considered loop, and $\{pa(e_l), 1\}$, $\{pa(e_l), 2\}$ represent the two parent nodes which are linked through l. Following Eq. (3.4), it results:

$$p(e_l) = 1 \implies p(v_{pa(e_l),1}, v_{pa(e_l),2}, e_l) = p(v_{pa(e_l),1}, v_{pa(e_l),2}) \tag{3.6}$$

Then, it results:

$$\begin{aligned} D(p||p_a) = &\sum_{i=1}^{n} H(V_i) - H(V_1, .., V_n) - \sum_{i=1}^{n} I(V_i; V_{pa(i)}) \\ &- \sum p(\mathbf{v}) \log \frac{p(e_l, v_{pa(e_l),1}, v_{pa(e_l),2})}{p(v_{pa(e_l),1}) p(v_{pa(e_l),2})} \end{aligned} \tag{3.7}$$

By adding $p(e_l)$ inside the denominator:

$$\begin{aligned} D(p||p_a) = &\sum_{i=1}^{n} H(V_i) - H(V_1, .., V_n) - \sum_{i=1}^{n} I(V_i; V_{pa(i)}) \\ &- \sum p(\mathbf{v}) \log \frac{p(e_l, v_{pa(e_l),1}, v_{pa(e_l),2})}{p(e_l) p(v_{pa(e_l),1}, v_{pa(e_l),2})} - \sum p(\mathbf{v}) \log p(e_l) \end{aligned} \tag{3.8}$$

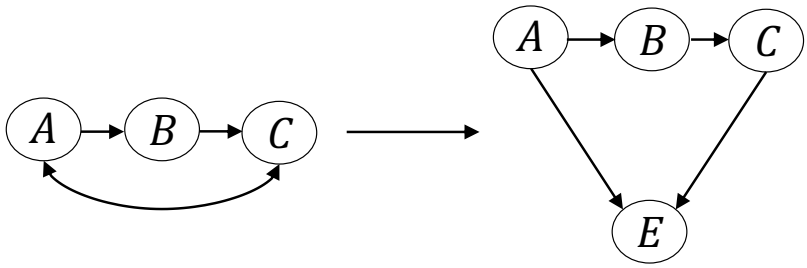

Figure 3.2: The auxiliary variable E for describing the loop between A and C respecting the causal dependence.

Since $p(e_l) = 1$, then the following equality holds:

$$D(p||p_a) = \sum_{i=1}^{n} H(V_i) - H(V_1, .., V_n) - \sum_{i=1}^{n} I(V_i; V_{pa(i)}) - I(e_l; v_{pa(e_l),1}, v_{pa(e_l),2}) \tag{3.9}$$

The first two terms are both independent of the dependence tree, whereas the last two terms represent the branch weights. The same proof can be done for all loops. As these last terms are both non-negative, minimizing the divergence measure is equivalent to maximize the total branch weight for both directed edges and loops.

3.1.3 Discretization

The TAN model typically works well on discrete data. In order to deal with continuous variables, the variables are generally discretized and the network is learned on the discretized domain. However, inappropriate discretization intervals may cause strong performance degradation because it may happen that the selected discretization step is able to capture only rough characteristics of the real distribution. The Clustering of $\sqrt{N}$-Interval Discretization (CloNI) is adopted as discretization method because it attempts to minimize the number of intervals while maintaining high accuracy [93]. According to this discretization method, the discretization step is different in the offline and near real-time stage.

3.2 Approximate Markov Blanket

The TAN model, which includes directed edges and loops, plays a fundamental role in the feature selection process because of its potential application as a MB technique [87]. The MB of a feature v_i provides a complete picture of the local causal structure around v_i. A number of MB-based feature selection approaches have been proposed [94]. However, they are slow and inefficient in terms of information found as they are based only on independence tests [95]. A different approach focuses on identifying the MB by first performing a BN learning step [96]. Under the faithfulness assumption, the MB of a variable v_i in a BN is unique and consists of parents, children and spouses of v_i [91].

By following a similar approach, the constructed TAN model is used to identify the MB of the classification class C. Since all of the features are target's children in the TAN model, the MB of C includes all features. In order to select the most relevant features for classification by taking advantage of the causal dependence structure of the TAN model, an approximation of the MB is derived [97].

Definition 1: *For two features v_i and v_j $(i \neq j)$, v_j is an Approximate Markov Blanket (AMB) of v_i if $SU_{j,C} \geq SU_{i,C}$ and $SU_{i,j} \geq SU_{i,C}$, where the symmetrical uncertainty SU measures the correlation between features and between feature and class C.*

Generally, the AMB-based feature selection algorithms discard features which are included in the MB of another feature as redundant to it, hence irrelevant to classification. Since the causal dependencies between features are already known in the derived TAN model, the MB(C) can be directly approximated by performing pairwise comparisons between each parent and children nodes. The mutual information is used as correlation measure. Algorithm 1 shows in detail how the identification of the AMB of C is performed. More specifically, ε is the set of directed edges, for which v_i is the parent of v_j. In constrast, $\tilde{\varepsilon}$ is the set of loops. AMB(C) and SP(C) are the approximate MB and spouse set of C, respectively. The pairwise comparisons over the directed edges are performed and all features with an higher relevance to C are included in the

Algorithm 1 AMB TAN based feature selection

Initialization: $\mathrm{AMB}(C) = \{\emptyset\}$, $\mathrm{SP}(C) = \{\emptyset\}$
Output: $\mathrm{AMB}(C)$

for $(i, j) \in \varepsilon$ **do** ▷ *Pairwise comparisons over directed edges*
  if $I(v_j; C) \geq I(v_i; C)$ & $I(v_j; v_i|C) \geq I(v_i; C)$ **then**
    $\mathrm{AMB}(C) = \mathrm{AMB}(C) \cup \{v_j\}$
    $\mathrm{SP}(C) = \mathrm{SP}(C) \cup \{v_i\}$
  else if $I(v_i; C) \geq I(v_j; C)$ & $I(v_j; v_i|C) \geq I(v_j; C)$ **then**
    $\mathrm{AMB}(C) = \mathrm{AMB}(C) \cup \{v_i\}$
  else
    $\mathrm{AMB}(C) = \mathrm{AMB}(C) \cup \{v_i, v_j\}$
  end if
end for
for $(i, j) \in \tilde{\varepsilon}$ **do** ▷ *Pairwise comparisons over loops*
  if $\{v_i, v_j\} \notin \mathrm{SP}(C)$ **then**
    if $I(v_j; C) \geq I(v_i; C)$ **then**
      $\mathrm{AMB}(C) = \mathrm{AMB}(C) - \{v_i\}$
    else if $I(v_i; C) \geq I(v_j; C)$ **then**
      $\mathrm{AMB}(C) = \mathrm{AMB}(C) - \{v_j\}$
    end if
  end if
end for
for $v_{pa(i)} \in \mathrm{SP}(C)$ **do** ▷ *Deletion from spouses*
  if $v_i \notin \mathrm{AMB}(C)$ **then**
    $\mathrm{SP}(C) = \mathrm{SP}(C) - \{v_{pa(i)}\}$
  end if
end for
$\mathrm{AMB}(C) = \mathrm{AMB}(C) \cup \mathrm{SP}(C)$ ▷ *Final* AMB *evaluation*

$\mathrm{AMB}(C)$. Their parents are added to the spouse set. From this initial AMB set, some features are deleted by comparing the correlation to C over the loops. Finally, all features $v_{pa(i)}$ whose children v_i have been removed from $\mathrm{AMB}(C)$ in the previous step are deleted from $\mathrm{SP}(C)$. The final $\mathrm{AMB}(C)$ is given by the union of $\mathrm{AMB}(C)$ and $\mathrm{SP}(C)$.

3.2.1 Sampling weights

The AMB TAN approach is applied to the updated database in the near real-time stage, as shown in Fig. 3.1. Due to time restrictions, only few OCs from the near real-time probability distribution are included in the knowledge base. To increase their impact on the feature selection approach, an importance estimation method is adopted [80]. The sampling weights are

calculated through a logistic regression classifier. A selector variable $\delta = 0$ and $\delta = 1$ is assigned to samples from the offline and the near real-time probability distribution, respectively. By applying the Bayes theorem, the weights w can be expressed in terms of δ:

$$w(\mathbf{v}) = \frac{p(\mathbf{v}|\delta = 1)}{p(\mathbf{v}|\delta = 0)} = \frac{p(\delta = 0)}{p(\delta = 1)} \frac{p(\delta = 1|\mathbf{v})}{p(\delta = 0|\mathbf{v})} \tag{3.10}$$

where the first term is the ratio between the number of samples from the two distributions and the second term is calculated by training a logistic regression classifier with δ as class variable. Then, a weighted form of the mutual information is used to measure the correlation between the features and the class variable [98]. In Algorithm 1, $I(v_i, C)$ is replaced with $wI(v_i, C),\ i = 1...n$, where n is the number of features and $dim(w) = N$, which is the total number of OCs in the updated database.

3.2.2 Computational complexity

The computational complexity of the proposed FS method is composed of: i) computational time of the CloNI algorithm; ii) computational time of the TAN learning algorithm. Focusing on the discretization method, given n and N the number of variables and samples, the computational time is:

$$\mathcal{O}\left(n \cdot \left(N \log N + \sqrt{N}\right)\right) \approx \mathcal{O}\left(n \cdot N \log N\right) \tag{3.11}$$

By parallelizing the process over n variables, the computational cost is reduced to $\mathcal{O}(N \log N)$. On the other hand, the naïve bayesian model learning generally requires $\mathcal{O}(n^2)$ mutual information tests. It is necessary to check if the number of tests increases when the mutual information tests are replaced by conditional mutual information tests for the TAN learning algorithm. In order to evaluate the conditional mutual information, the training data is partitioned set by class values [99]. Then, the mutual information conditioned to each class value is evaluated.

Given x, y discrete random variables and z binary variable, it results:

$$\begin{aligned} I(x;y|z) =& \sum_{z\in Z} p(z) \sum_{y\in Y}\sum_{x\in X} \cdot \log\left(\frac{p(x,y|z)}{p(x|z)p(y|z)}\right) \\ =& \frac{k_{z=0}}{N}\cdot I(x;y|z=0) + \frac{k_{z=1}}{N}\cdot I(x;y|z=1) \end{aligned} \tag{3.12}$$

where $I(x;y|z)$ represents the mutual information between x and y conditioning on z, k_z is the number of samples for which z assumes each class value and N is the total number of samples. By defining D_i the partition for which $Z = z_i$, noticing that the mutual information is symmetrical in D_i and $\sum_i |D_i| = N$, the computational complexity for (3.12) is:

$$\mathcal{O}\left(\sum_i |D_i| \left(\frac{n}{2}\cdot(n-1)\right)\right) = \mathcal{O}\left(N\left(\frac{n}{2}\cdot(n-1)\right)\right) \approx \mathcal{O}\left(\sqrt{N}\cdot\frac{n}{2}\right) \tag{3.13}$$

where the last equivalence is obtained by parallelizing over n variables and by applying the CloNi algorithm over N samples.

The proposed algorithm has a time complexity which scales linearly in the number of variables. Moreover, the way by which it scales in the number of instances depends only on the chosen discretization algorithm.

3.3 Case Study

Several studies were undertaken to demonstrate the benefits of the AMB TAN FS approach for power system DSA. First the proposed three-stage approach using the AMB TAN FS is compared to the traditional two-stages approach for DSA. Then, the trade-off of accuracy and computations was explored in detail for the proposed FS method and compared to existing techniques. Finally, the computational savings of using the proposed workflow were investigated.

3.3.1 Test system and assumptions

The IEEE 68-bus system (Fig. 3.3) was first used for testing the performance of the proposed workflow [3], and later the computations were scaled to a large-scale system (French Transmission System corresponding to 1,955 transmission lines, 1,886 buses and 411 generators). A set of 12000 OCs was generated and each of them represented a pre-fault condition of the system considering the full AC model. These OCs were generated by sampling the active loads from a multivariate Gaussian distribution via Monte Carlo sampling and assuming a Pearson's correlation coefficient $c = 0.75$ between all load pairs. The multivariate Gaussian distribution can either fit the available historical data or it can be randomly drawn as in this case. The sampled active loads from the Gaussian distribution were then mapped to a marginal Kumaraswamy distribution using the inverse transformation method. The cumulative distribution function of the Kumaraswamy distribution is:

$$f(x) = abx^{a-1}(1 - x^a)^{b-1} \tag{3.14}$$

where $a = 1.6$, $b = 2.8$ are shape parameters and $x \in [0, 1]$ are observations of the variable X representing the active loads. The benefit of using the Kumaraswamy distribution for modelling the stochastic nature of the loads is that it is highly flexible to adapt to the skewness of the load distributions by appropriately modifying the shape parameters [100]. Finally, the active loads were scaled to be within $\pm 50\%$ of the nominal values provided in [3]. The reactive powers follow the active powers proportionally as constant impedances were assumed. Then, i.i.d power factors were sampled in the range of $[0.95, 1]$ for the generators. As the resulting OCs may be infeasible, the full AC model was considered in a mathematical optimization problem to minimize the absolute difference between these power factors. In that way, active and reactive powers corresponding to feasible OCs were obtained for the generators. The optimization problem was implemented in Python 3.5.2 and Pyomo package and solved with IPOPT 3.12.4 [101]. The transients of three-phase faults over 22 different lines were simulated ($k = 1, \ldots 22$). If during 10s simulation time all the differences between each two phase angles

Table 3.1: Accuracy performance using different classification models

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

of the generators were less than 180°, then the OC i was considered stable $Y_{i,k} = 1$, otherwise unstable $Y_{i,k} = 0$. A fault clearance time of 0.1s was used. In the resulting 22 datasets, the percentage of unstable observations was between 1% and 91% with 46% as mean. The simulation was performed in Matlab R2016b Simulink.

In the machine learning part, voltage magnitudes were considered as the relevant features according to Section 3.1. Consequently, each OC X_i was composed of all 68 buses voltage magnitudes. The data were pre-processed by applying existing feature selection techniques and the AMB TAN causality-based approach. The Minimum Redundancy Maximum Relevance (MRMR), Correlation-based Feature Selection (CFS) and Joint Mutual Information (JMI) as filter techniques, SVM Recursive Feature Elimination (RFE) as embedded method, and Sequential Forward Selection (SFS) as wrapper method, were used. Then, the CART learning from the *scikit-learn* algorithm was used to train DTs. In Table 3.1, 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 on the final accuracies. DT with depth equal to 3, SVM with linear kernel [28], AdaBoost and XGBoost with 50 estimators [102, 103], and single layer feed-forward ANN with 10 neurons [25], were used. It resulted that all testing accuracies of these approaches were very similar, therefore DTs with maximum depth $D = 3$ were preferred as they are more interpretable. Across the studies, 10 different combinations of training/testing set were computed for each classifier. The training and testing split was 70%/30% in all studies and these two sets were drawn from the same probability distribution, unless indicated otherwise. One DT was learned for each of the 22 datasets. The DT learned for the contingency on line $31-38$ using the AMB TAN causality-based approach as pre-processing step is shown in Fig. 3.4. As part of this study, the size of the training database was varied in the offline and near real-time stage. The F_1 score was used as criterion for the accuracy.

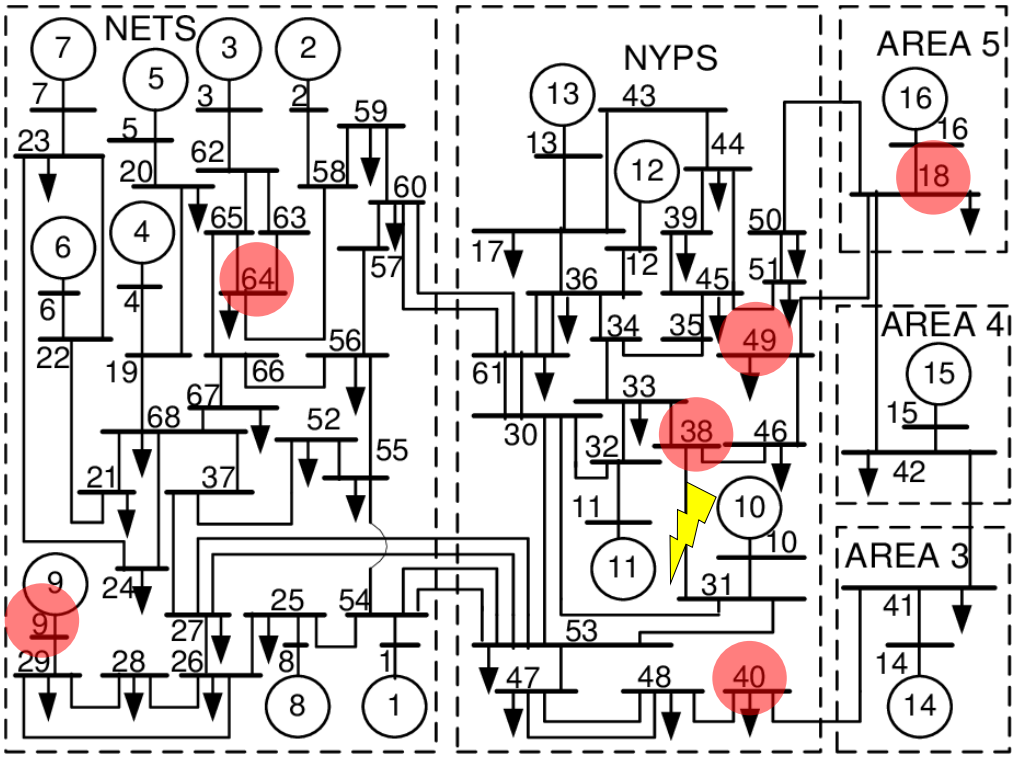

Figure 3.3: The IEEE 68-bus system [3]. In red the features selected through the AMB TAN FS approach for the contingency on line 31 – 38.

3.3.2 Inaccuracy of the security rules

In this study, the advantages of the proposed three-stages workflow in terms of the accuracy performance on new OCs were investigated. The offline stage was considered to be several weeks before real-time operation. In the offline stage, it was assumed the operators knew the load distribution parameters to be $a_0 = 1.6, b_0 = 2.8$ and $c_0 = 0.75$. To simulate the changing distribution of OCs, the parameter b was assumed to decrease over time, as shown in Fig. 3.5. It was assumed operators were aware of this trajectory at the online time. Two training approaches were compared. In the first one the classifier was trained only in the offline stage, whereas in the second one it was re-trained in the near real-time stage. At this stage, a few OCs can be generated and simulated from an updated load distribution that corresponds to the trajectory in Fig. 3.5. In the two approaches, 12000 OCs were used to perform FS and classifier training. However, in the traditional approach, these 12000 OCs were generated offline from a probability distribution different from the actual one. Then, SFS was used before

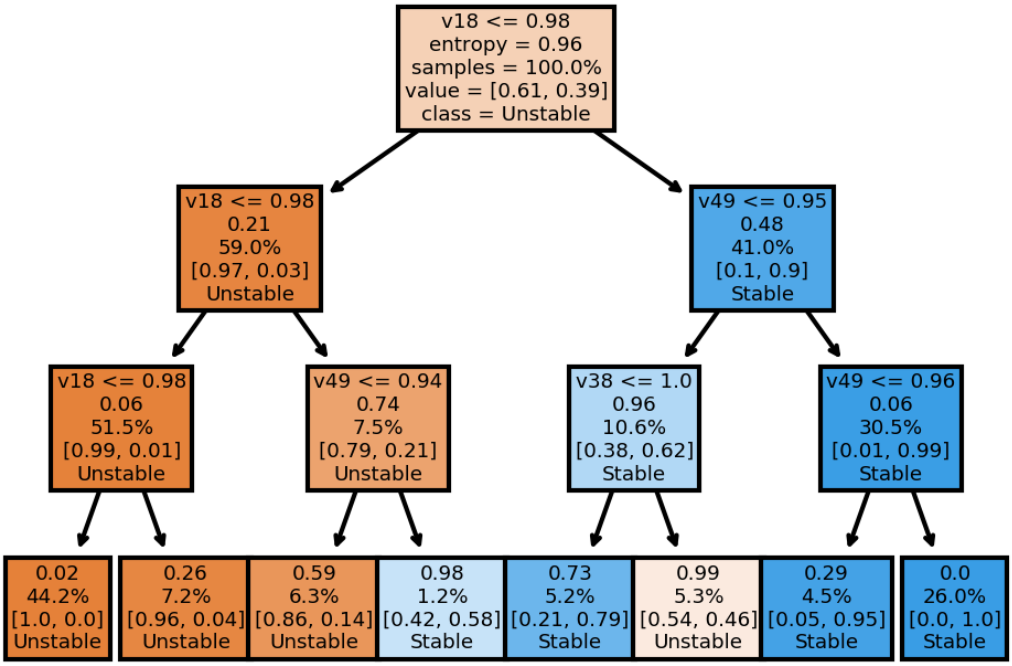

Figure 3.4: The DT learned for the contingency on line $31 - 38$. The entropy was used as uncertainty measure. The voltages v_{18}, v_{38}, v_{49} were selected as features.

training the classifier. In the proposed approach, the bulk of data came from the offline stage and only 150 OCs from the near real-time distribution were included in the training database due to limited computational resources. Subsequently, the weight of each OC in the updated database was calculated, AMB TAN approach was used as feature selection and then the classifier was trained. Subscripts $0, n, t$ were used to indicate the offline, the near real-time and the online stage, respectively. Then, the two approaches were tested on 50 new observations for which $a_t = 1.6, b_t = 2.4, c_t = 0.75$ according to Fig. 3.5. In the proposed approach, the average accuracy across all contingencies improved by roughly 1%. The same study was then repeated for five different linear trajectories in the parameter changes. The results are summarized in Table 3.2 and demonstrate the improvements in the accuracy across all tested trajectories. The individual contributions of the AMB TAN FS and the classifier training to this improvement were also investigated. By leaving the AMB TAN FS out, the improvement dropped by 0.25%. Finally, the features selected through the proposed approach across all contingencies were analysed to demonstrate the increased interpretability of considering the causal relationships and the physical knowledge in the learning process. The most features

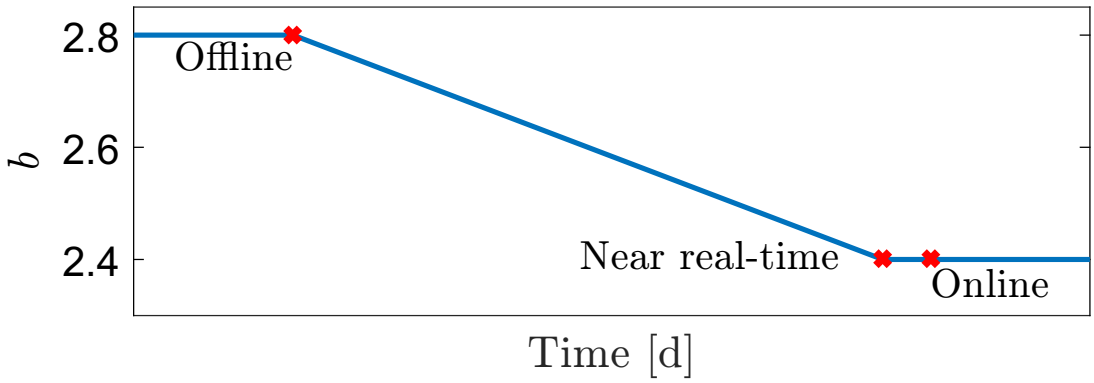

Figure 3.5: Trajectory of the parameter b by assuming a constant value between near real-time and online stage.

Table 3.2: Accuracy improvement by re-training

Training stage		**Testing stage**	**Accuracy**
Offline	*Near real-time*	*Online*	**improvement**
$b_0 = 2.8$	$b_n = 2.2$	$b_t = 2.2$	+0.14%
$a_0 = 1.6$	$a_n = 1.2$	$a_t = 1.2$	+0.22%
$a_0 = 1.6$	$a_n = 1.4$	$a_t = 1.4$	+1.32%
$c_0 = 0.75$	$c_n = 0.25$	$c_t = 0.25$	+1.52%
$c_0 = 0.75$	$c_n = 0.50$	$c_t = 0.50$	+0.22%

included in the MB of each contingency were located very close to the faulted bus, as shown in Fig. 3.3. At the same time, similar MBs were identified for contingencies with close faulted buses.

3.3.3 Residual uncertainty

Here, the residual uncertainty in the load distributions that leads to discrepancies between the estimate and the actual probability distribution was investigated. It was assumed the parameter b decreased until one day before the operation. In the near real-time stage, three possible trajectories of b were considered, as shown in Fig. 3.6. As in the previous study, 150 OCs were generated from the probability distribution of the near real-time stage and used to enrich the offline database. Then, all steps from weighting, applying AMB TAN FS and training the classifier, were performed. Subsequently, the proposed approach and the offline training based approach were tested on three different test sets of 50 OCs corresponding to trajectories

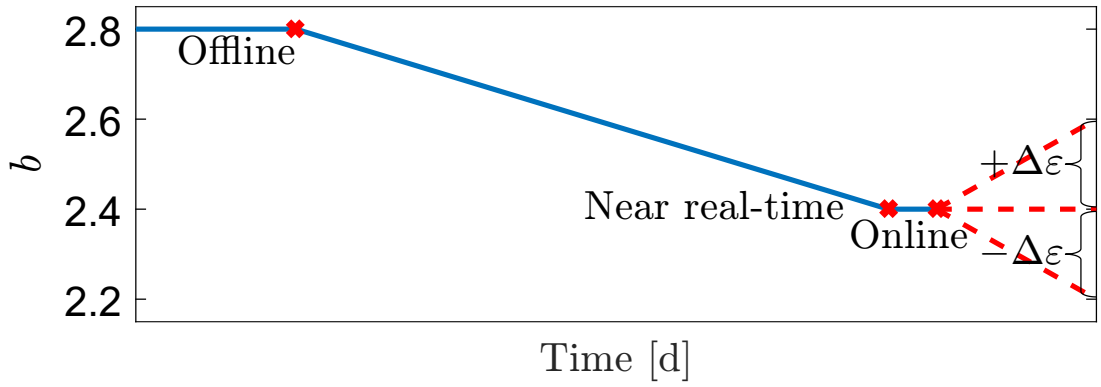

Figure 3.6: Trajectory of the parameter b accounting for the residual uncertainty between near real-time and online stage.

shown in Fig. 3.6 (red dotted lines). The parameters of the three test sets were calculated as follows:

$$b_{t1} = b_n - \Delta\varepsilon,\ b_{t2} = b_n,\ b_{t3} = b_n + \Delta\varepsilon \tag{3.15}$$

with $\Delta\varepsilon = 0.2$. The result of this test is that the use of the AMB TAN FS in the near real-time stage increased the accuracy mean value over all contingencies by 0.7% for b_{t1}, by 0.9% for b_{t2} and decreased by 0.8% for b_{t3}. Subsequently, the same study was undertaken for five different trajectories as per Table 3.2. When the change in the load probability distribution in the online stage progresses with the same trend of offline and near real-time stage as for b_{t1} and b_{t2}, the accuracy mean value improved by 0.7% and 0.72%, respectively. If the change in the load probability distribution progresses with a different trend in near real-time and online stage as for b_{t3}, the OCs included in the training database were not informative, and hence the accuracy mean value decreased by 0.5%. However, the decrease in the predictive performance for b_{t3} was lower in mean than the improvements for b_{t1} and b_{t2}.

3.3.4 Trade-off between accuracy and computational time

The following study focused on the trade-off between accuracy and computational time of the feature selection approaches for machine learning based DSA. Various FS approaches were used to select the features on which train the DT afterwards by using $|\underline{\Omega}| = 12000$ OCs. Subsequently,

Table 3.3: Statistical analysis of distributions shown in Fig. 3.7

	FS Method					
	AMB	MRMR	CFS	JMI	RFE	SFS
Accuracy Mean	0.89	0.71	0.75	0.77	0.85	0.89
Accuracy Std	0.08	0.24	0.26	0.23	0.13	0.08
Accuracy CI	0.01	0.02	0.02	0.02	0.01	0.01
Time Mean	$87s$	$10s$	$353s$	$153s$	$75s$	$352s$
Time Std	$11s$	$3s$	$213s$	$81s$	$47s$	$94s$

the training was repeated 10 times with varying splits of training/testing data to compute an average value for each of the 22 contingencies. In total 220 DTs were trained for each of the 6 FS approaches. The mean values of each contingency for computational time and accuracy are presented in Fig. 3.7 and summarized in Table 3.3. The confidence interval (CI) for the accuracies was evaluated considering a confidence level of 95%. Hence, the resulting accuracies do not vary much from the values indicated in Table 3.3, making then the comparison between the feature selection approaches in terms of the accuracies more faithful. Three main advantages of the AMB TAN approach can be observed: (1) AMB TAN resulted in higher accuracies than SVM-RFE, (2) AMB TAN requires 75% less computational time than SFS, resulting in the best trade-off between computational time and accuracy, and (3) the distribution in computational times is narrow. This narrow distribution enables precise estimations of the required computational times, allowing to reliably schedule the feature selection and classifier training later in the workflow.

The French transmission system was used to illustrate these advantages for larger systems. The computational times were estimated by scaling up from the IEEE 68 bus system and using $\mathcal{O}(n/2)$ and $\mathcal{O}(n^2)$ with n number of buses for the AMB TAN and SFS approach, respectively. The resulting computational times are reported in Table 3.4. There, the SFS approach would require a computational time of $75h$ to identify features, where the AMB TAN approach requires around $30min$.

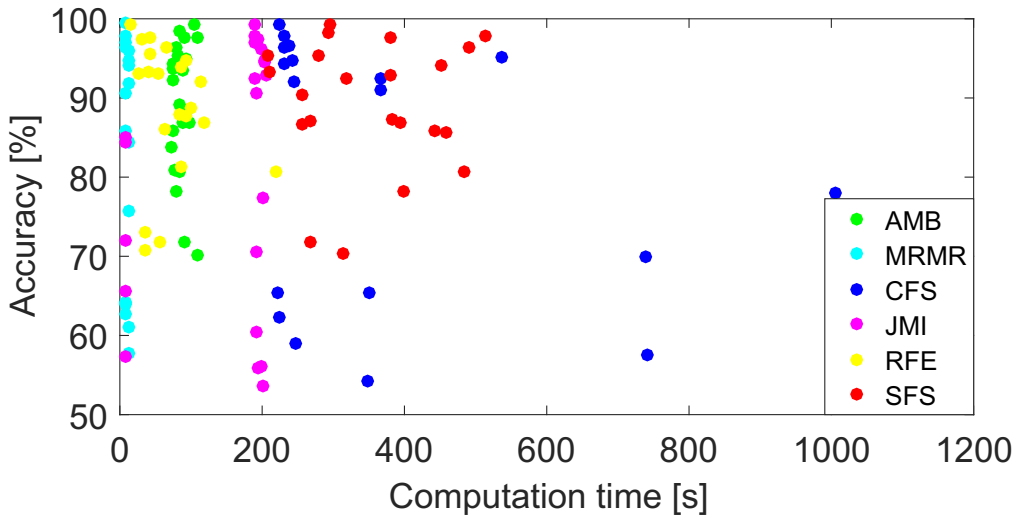

Figure 3.7: Accuracy and computation time for 22 contingencies with different FS techniques.

3.3.5 Size of the training database

In this study, the required size of the training database when using various feature selection approaches was investigated. This was an important analysis to conduct as operators need to decide how to optimally allocate the computational budget on data generation and training (FS and classifier). The size of the training database $|\underline{\Omega}|$ was varied, as the number of operating conditions used in feature selection and classifier training. The four best FS approaches (AMB TAN, MRMR, RFE and SFS) in terms of the accuracy and computations trade-off were first studied with larger database size, i.e. $|\underline{\Omega}| = 40000, 80000, 120000$, to guarantee the scalability of the workflow. For each step, the four different FS were applied and the DT trained. This was repeated 10 times with different training/testing data combinations. The results were averaged and shown in Table 3.5 for the single contingency with the lowest accuracy across all contingencies. The AMB TAN, RFE and SFS approaches resulted in highest accuracies, however AMB TAN outperformed in terms of computations. Since SFS resulted in similar accuracies to RFE in reduced computational times, only the SFS approach was considered in the following studies. Then, the same analysis was conducted with reducing database size, i.e. $|\underline{\Omega}| =$ 1000, 2000, ...12000 and applying the best three FS approaches (AMB TAN, MRMR and SFS). The results were averaged and illustrated in Fig. 3.8 for the same single contingency. The AMB

Table 3.4: Estimate of computational time for FS for a large system.

Approach	**Case Study**	
	IEEE 68-bus system	French transmission system
AMB TAN	$87s$	$\sim 30min$
SFS	$6min$	$\sim 75h$

Table 3.5: Accuracy and computation time according to different FS techniques by increasing the size of the training database.

Training data	$\lvert\underline{\Omega}\rvert = 40000$		$\lvert\underline{\Omega}\rvert = 80000$		$\lvert\underline{\Omega}\rvert = 120000$	
	Accuracy	*Time*	*Accuracy*	*Time*	*Accuracy*	*Time*
AMB TAN	0.81	$271s$	0.81	$628s$	0.81	$1083s$
MRMR	0.47	$17s$	0.56	$54s$	0.61	$52s$
RFE	0.81	$1116s$	0.82	$4772s$	0.82	$10555s$
SFS	0.83	$1031s$	0.83	$2466s$	0.83	$4296s$

TAN and SFS approach resulted in highest accuracies. However, AMB TAN outperformed SFS in terms of computations. The best database size for this contingency was around $|\underline{\Omega}| = 5000$ as the accuracy did not improve anymore for larger database. At $|\underline{\Omega}| = 5000$, AMB TAN required 70% less computational time than SFS. If in the real-time stage no FS approach was used and the training database was reduced to $|\underline{\Omega}| = 5000$, the accuracy decreased by 5%.

For the same contingency, the required training OCs in the offline stage when introducing the near real-time stage with the AMB TAN approach was investigated. The offline database size was reduced to $|\underline{\Omega}| = 500$, which is the minimum number of OCs to guarantee the feasibility of the SFS approach. It was assumed 4500 OCs were included in the database during the near real-time stage as $|\underline{\Omega}| = 5000$ was shown to be the best size in terms of accuracy performance. The AMB TAN approach was used, the classifier was trained and then tested on 1000 new OCs from the same load distribution. It turned out that the offline database size could be reduced up to 95%, resulting in the same accuracy performance.

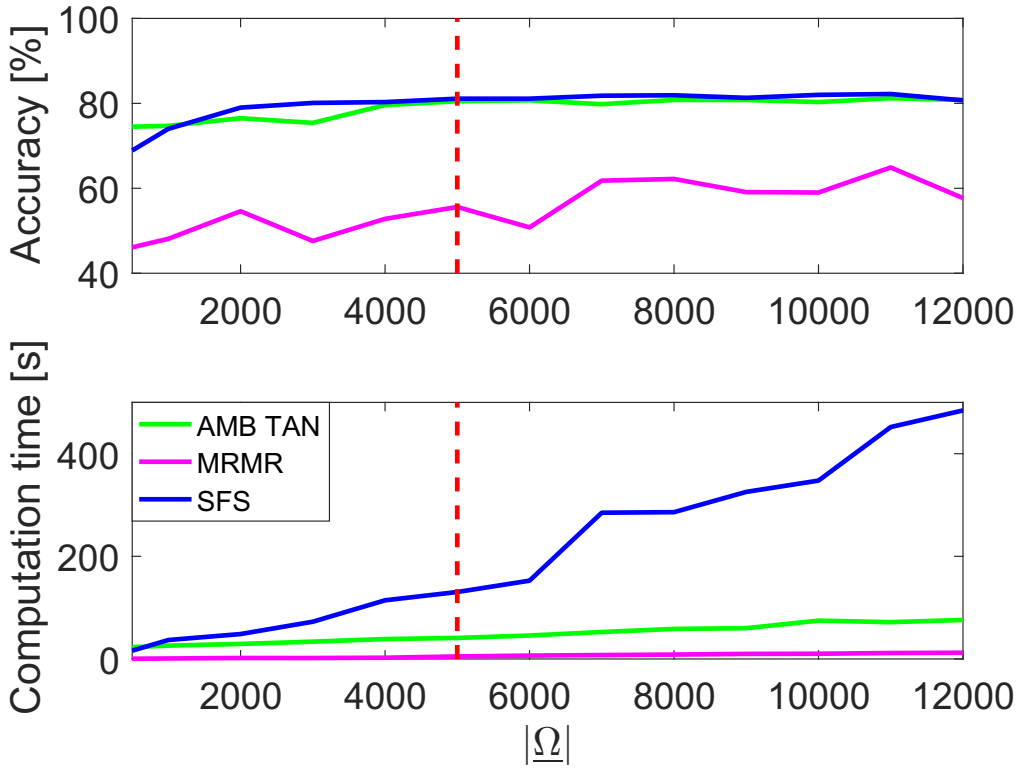

Figure 3.8: Accuracy and computation time according to different FS techniques by varying the size of the training database.

3.3.6 Computational efficiency

In this study, the computational savings obtained by using the proposed workflow for DSA were investigated. For the offline and near real-time stages, the impact of using or not the proposed AMB TAN approach on the computational cost of data generation and training of the machine was analysed. For this comparison, 12000 OCs were available in the offline and near real-time stage. However, the use of the AMB TAN approach allowed to reduce the offline and near real-time training databases to $|\underline{\Omega}| = 500$ and $|\underline{\Omega}| = 5000$, respectively, resulting in same accuracy performance (Section 3.3.5). Table 3.6 and Fig. 3.9 summarize the computational costs for data generation, feature selection and training of the machine when using the proposed AMB TAN approach and when not using the approach. The simulation time for each observation was 0.2s. The proposed AMB TAN approach enabled a total computational saving in DSA workflows (offline and near real-time stages) up to 75%. This result is a major finding of this work.

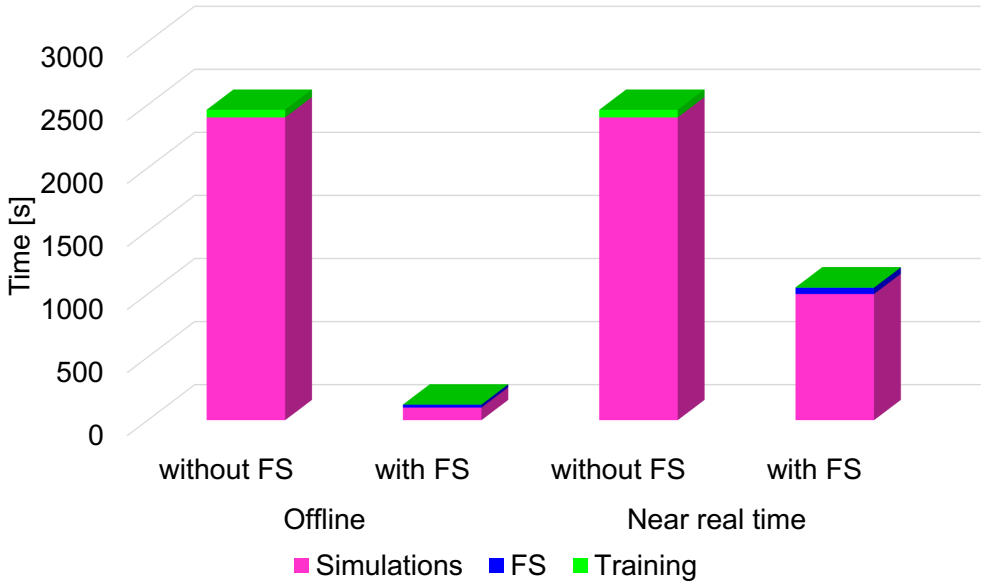

Figure 3.9: Computation times for offline and near real-time stages using and not the proposed AMB TAN approach.

Table 3.6: Computation times for offline and near real-time stages using and not the proposed AMB TAN approach.

Stage	Offline		Near real-time	
Approach	without FS	with FS	without FS	with FS
Nº observations	12000	500	12000	5000
Simulation time	2400s	100s	2400s	1000s
FS time	0s	23s	0s	50s
DT training time	60s	1s	60s	3s
Total	2460s	124s	2460s	1053s

3.3.7 A combination-based approach

To optimize the balance between high accuracy and low computations, a further analysis on a combination-based feature selection approach was conducted. This is worth investigating because different feature selection approaches showed the best performance for different contingencies, as shown in Fig. 3.7. In fact, the performance of MRMR was higher than AMB TAN for some contingencies. For this reason, the performance when combining these two FS approaches was investigated. For all contingencies, MRMR was first applied as its computational times were negligible. Then, only for contingencies where the accuracy was lower than a threshold, AMB TAN was used to select more relevant features. The threshold was selected to be 0.8 and

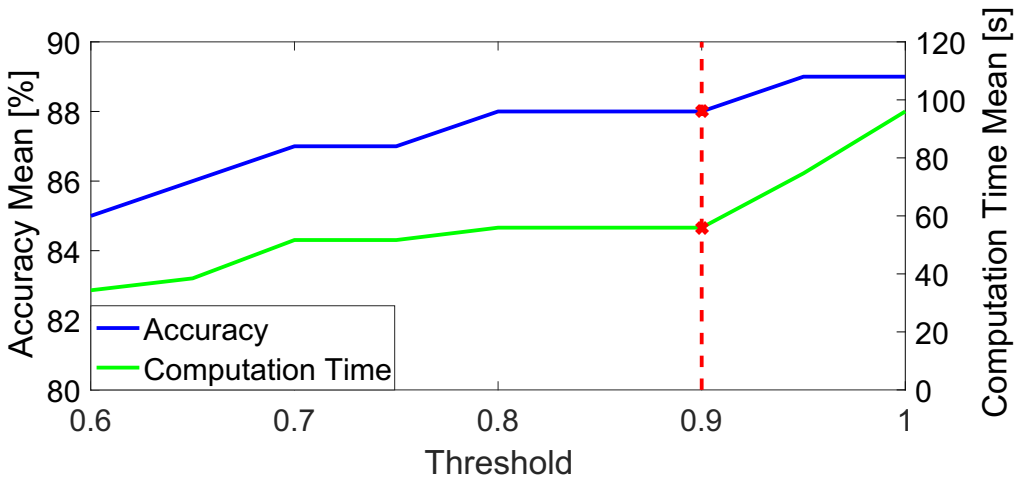

Figure 3.10: Mean values of accuracy and computation time by using the combination-based approach and varying the accuracy threshold.

the mean of accuracy and computational time were compared across all 22 contingencies for the AMB TAN versus this combined approach. The result was that the computational time reduced by 35% when using the combined approach, whereas the accuracy drops by only 0.5%. Then, the threshold value was studied in the range of $[0, 1]$ with step-size 0.05. The full results are shown in Fig. 3.10. It turned out that a threshold higher than 0.9 leads to under-performing the AMB TAN approach and 0.8 is the optimal threshold.

3.3.8 Discussion

The proposed causality-based FS approach in combination with a three-stages workflow showed promising results for online DSA applications, resulting in an accuracy increase of 1%. It resulted in the best trade-off in terms of accuracy and computational performance across other FS approaches, being four times faster than the SFS approach while obtaining the same high accuracy. The features selected through AMB TAN improved the interpretability of the classifier as they were all located close to the faulted bus and similar features were selected for similar contingencies. Moreover, AMB TAN has a very narrow distribution of the computational times. Consequently, operators can reliably schedule the learning task very close to the real-time operation as the required computational time can be estimated accurately. These

improvements in terms of computations are significant higher when moving to a large system, as illustrated on the French system (99% reduction). In terms of robustness against uncertain OCs, the proposed causality-based approach demonstrated to improve the accuracy for various trends in the change of the load probability distributions. The AMB TAN FS approach also reduced the amount of required training data by 60% and by 95% in comparison to the SFS approach for the offline and near real-time stage, respectively. This reduction is significant as a key bottleneck of online DSA approaches is the amount of data needed for large systems, and hence the computational cost for simulating the OCs from this large amount of data. Consequently, the proposed AMB TAN enabled a total computational saving for offline and near real-time stages, including data generation, feature selection and model training, up to 75%. A further reduction in the computational time was obtained by combining AMB TAN approach to another feature selection approach, resulting in almost the same accuracy (only 0.5% reduction). Overall, all these benefits represent a fundamental step forward to deploy machine learning approaches for DSA.

A few key limitations in designing feature selection to data-driven DSA approaches still exist. A lot of data are still required and if too little data are available, the feature selection approaches are not capable of identifying the underlying statistical dependencies, resulting in low prediction accuracy. The machine learning approaches that are used along AMB TAN FS 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 3.1). As the MB strongly varies from one contingency to another, single machine learning approaches for each contingency should be trained. In this work, the focus is on feature selection. However, when designing the entire machine learning workflow, every single step should be investigated and considered when allocating computational budgets. Relying on machine learning based DSA workflow rather than investing in new assets has a risk that should be considered in the decision making process. The proposed combination of the causality-based FS and near-real time stage should be also tested against other stability metrics. Finally, as topology changes may become more frequent, investigating how the proposed FS method performs under topology

changes becomes important [52]. The performance of the AMB TAN FS approach should be also tested when the occurrences of severe weather events or topological changes make the assumption on independent currents not valid anymore. In this context, incorporating the physical knowledge into the learning approach of the causal structure between features may also improve the robustness of a single classifier across similar contingencies or network's topologies.

3.4 Conclusion

The challenge of designing computationally efficient and robust feature selection approaches to machine learning-based DSA was investigated, showing that DSA can suffer from discrepancies between real-time and offline. Not considering these discrepancies along with the time horizon of generating data and training the machine results in inaccuracies. In response, a novel causality-based FS approach in combination with a near real-time stage was proposed to train the classifier closer to real-time operation. By using the system's physics to learn the causal structure and to identify the Markov Blanket, this approach outperformed other feature selection approaches in robustness, interpretability and significantly improved the computational time while the predictive accuracy is as high as state-of-the-art feature selection approaches. The IEEE 68 bus system and transient stability were used to illustrate these benefits, showing the required computational time for feature selection was reduced by 75%. This reduction in the computational time is becoming more significant for large systems as demonstrated for the French transmission system. Moreover, the required training database was reduced by 60%. This reduction is important as often the number of time-domain simulations for a training database is a barrier to using data-driven DSA. The proposed feature selection approach in combination with a three-stages workflow is a significant step forward to include dynamics in future's security assessment by the support of machine learning, enabling an effective operation of the grid assets closer to their limitations. In the future, the entire workflow should be investigated when allocating computational budgets for other objectives of DSA approaches.

Chapter 4

Machine Learning Model Re-training

A set of approaches to deal with topology changes in DSA frameworks is to retrain a new prediction model every time the network topology changes [52]. Since the frequency of topology changes is increasing these approaches are computationally impractical as they would require generating data and retraining very often. A new approach in [4] is building upon these approaches (Fig. 4.1) and makes them computationally more practical. Here, a data-driven based index is proposed 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 operating conditions 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

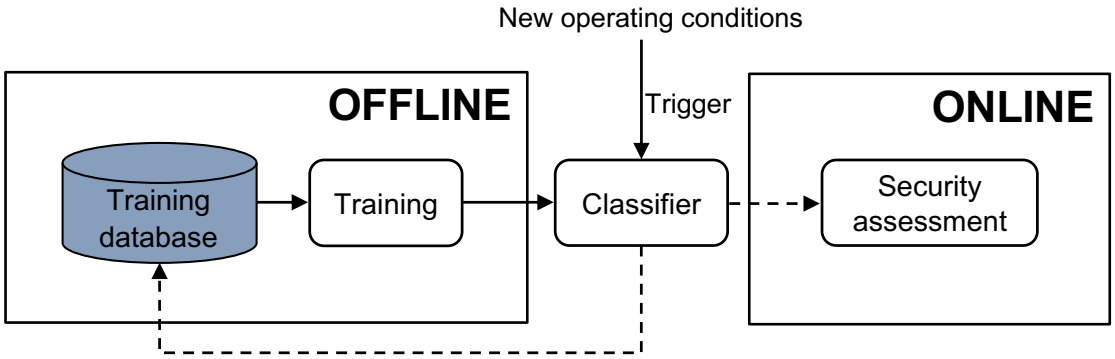

Figure 4.1: Data-driven workflow for DSA based on triggered re-training.

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.

In this chapter, a novel approach is proposed to address frequent changes in the network topology (Fig. 4.2). 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 using the proposed MB based feature selection approach (Section 3.2) 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 [104], 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 [105]. The proposed feature selection approach uses the

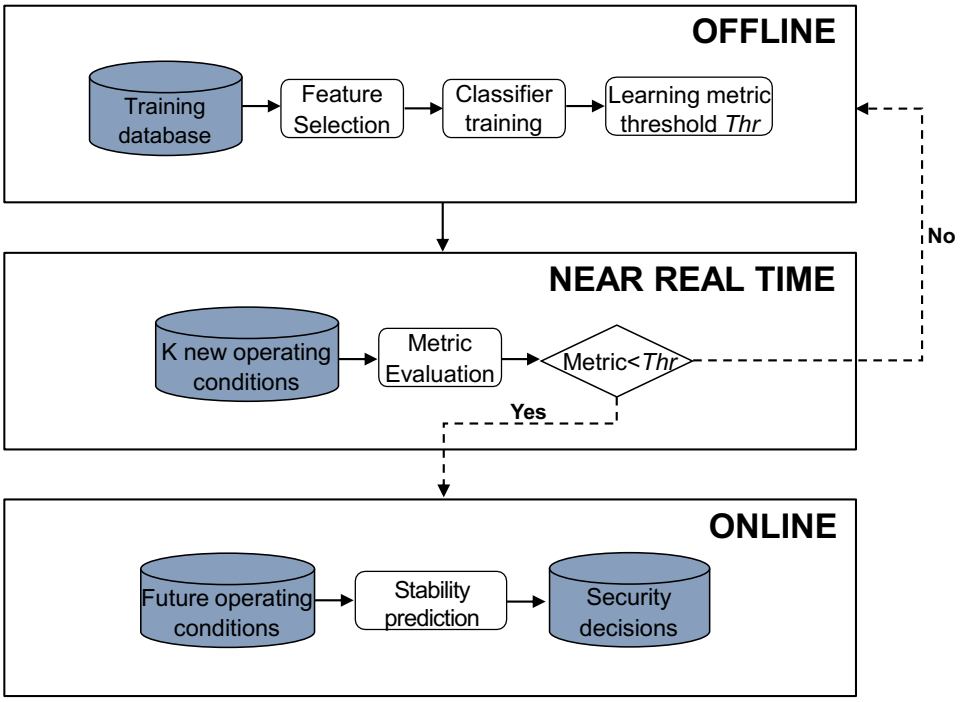

Figure 4.2: Data-driven workflow for DSA to deal with topology changes.

system's physics and data to discover the causal structure between features and then identifies highly relevant features by learning the approximate 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 [87]. 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 operating conditions based on their similiarity. 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.

4.1 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.

4.1.1 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 operating conditions, 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 [4, 106]. The ED between two OCs x_i and x_j originated from two distributions is calculated as follows:

$$ED(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + ... + (x_{in} - x_{jn})^2} \tag{4.1}$$

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. 4.3, 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 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 [107]. 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 $d_H(X, Y)$:

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\} \tag{4.2}$$

The two terms to calculate the Hausdorff distance are shown in Fig. 4.4. 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.

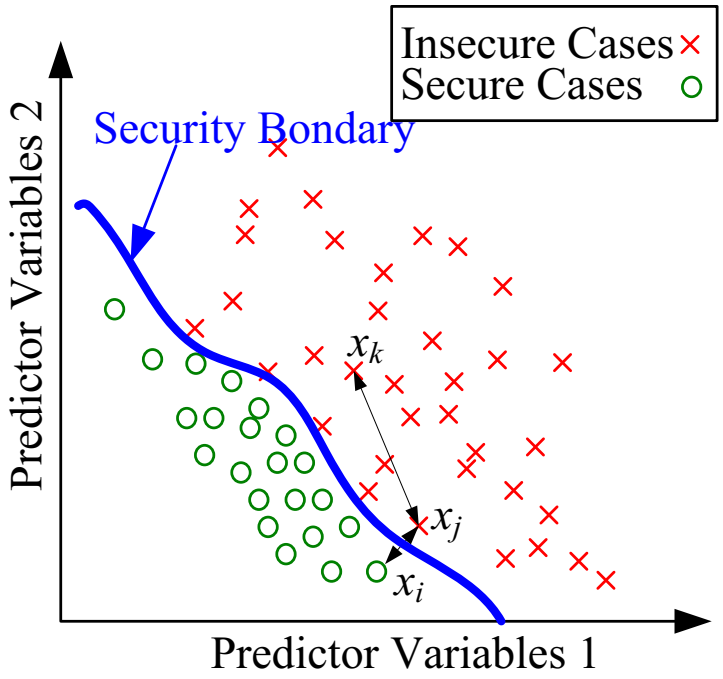

Figure 4.3: The drawback of the Euclidean Distance [4]. The ED between x_i and x_j is lower than the ED between x_j and x_k although x_i and x_j are on different sides of the security boundary.

4.1.2 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 [108, 109]. 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

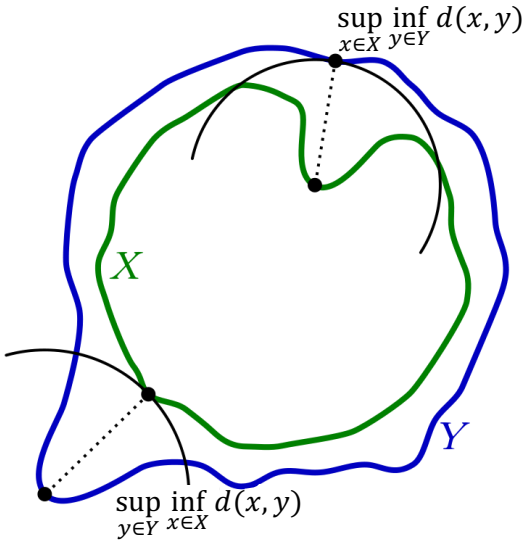

Figure 4.4: The Hausdorff distance between the two convex hulls in green and blue.

distance D:

$$D = |Y_{bus}^{-1}| \tag{4.3}$$

with

$$D_{ij} = \begin{cases} d_{ij} & \text{if } j \text{ adjacent to } i \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

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.

4.1.3 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

Algorithm 2 Training database construction

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

```
for i ∈ H_0 do
    if i ∈ H_1 then
        Tr = Tr ∪ (i, y_{i,0})
    end if
end for
for i ∈ H_1 do
    if i ∉ H_0 then
        Compute y_{i,1}
        Tr = Tr ∪ (i, y_{i,1})
    end if
end for
```

convex hulls, H_0 and H_1, 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. i represents a general operating condition, $y_{i,0}$ are the security labels of $i \in H_0$ that are available from the offline stage, and $y_{i,1}$ are the security labels of $i \in H_1$ that are not available and needed to be computed. Firstly, the OCs contained in H_0 which are also contained in H_1, and hence are still relevant, are included in Tr together with their corresponding security labels $y_{i,0}$. Secondly, time-domain simulations are performed for the OCs contained in H_1 which are not contained in H_0, and hence no similar to those already available. These OCs together with the newly computed security labels $y_{i,1}$ are then included in Tr. Only for these last OCs, time-domain simulations are performed to compute the security labels, resulting in significant reductions of the computational time needed to generate the new training database Tr.

4.2 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.

4.2.1 Test system and assumptions

The IEEE 68-bus system (Fig. 4.5) was used as one of the three test systems [3]. 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. 4.5 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. These OCs were generated by sampling the active loads from a multivariate Gaussian distribution via Monte Carlo sampling and assuming a Pearson's correlation coefficient $c = 0.75$ between all load pairs. The multivariate Gaussian distribution can either fit the available historical data or it can be randomly drawn as in this case. The sampled active loads from the Gaussian distribution were then mapped to a marginal Kumaraswamy distribution using the inverse transformation method. Finally, the active loads were scaled to be within $\pm 50\%$ of the nominal values provided in [3]. 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 ($k = 1, \ldots 9$) for all 42 topologies. A fault clearance time of 0.1s was used. If within 10s simulation time all the differences between each two phase angles of the generators were less than 180°, than 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

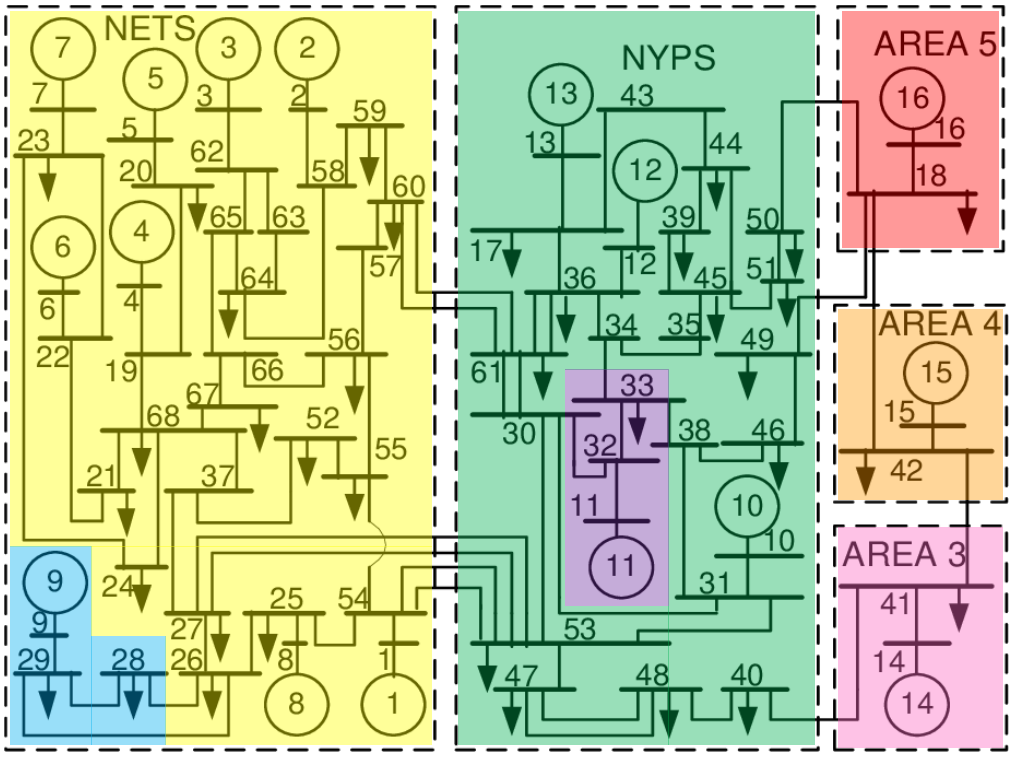

Figure 4.5: The IEEE 68-bus system [3]. In colour the seven areas with similar characteristics identified through spectral clustering analysis.

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 [110]. 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 3.1. 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 4.1, 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 training. DT with depth equal to 3, SVM with linear kernel [28], AdaBoost and XGBoost with 50 estimators [102, 103], and single layer feed-forward ANN with 10 neurons [25] 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%.

4.2.2 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 operating conditions 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-

Table 4.1: 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

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 chapter were not tested as they were computationally not feasible. All classification models were trained offline on the generated $20,000$ conditions for the reference topology and then the three approaches were tested against 42 topology changes on $1,500$ new operating conditions 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. 4.6 (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 and accuracy for the line contingency between bus 31 and 38 is shown in the two Figs. 4.6 (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. 4.7). 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.

4.2.3 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 (Sec. 4.1.1) and electrical regions are computed using the spectral clustering approach (Sec. 4.1.2). These 7 regions are highlighted in Fig. 4.5. Subsequently, if another contingency has common features to the MB of the trained contingency, then the classifier can still be used.

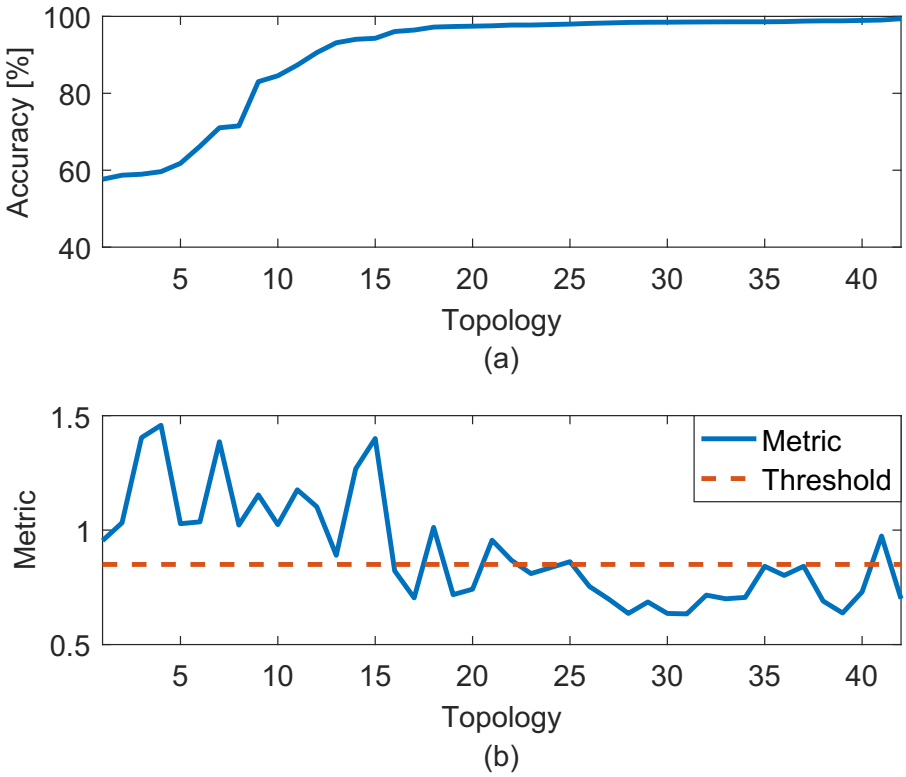

Figure 4.6: The accuracy and metric of classifier corresponding to contingency on line $31-38$ for 42 topologies. The topologies (x-axis) were ordered according to increasing accuracies.

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. 4.8). 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.

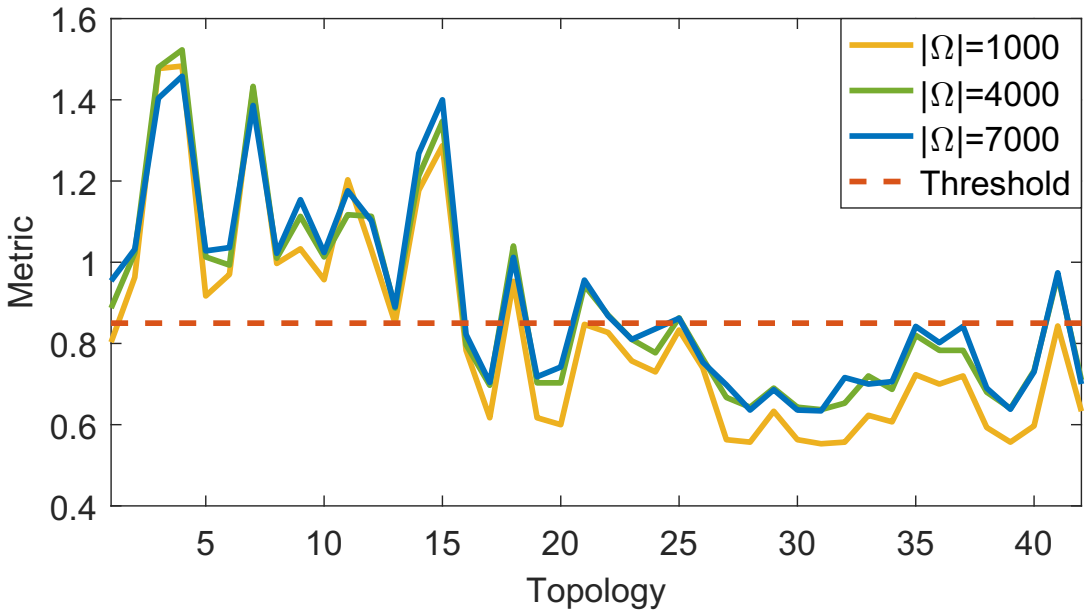

Figure 4.7: The metric for classifier corresponding to contingency on line 31−38 for 42 topologies with varying training database sizes.

4.2.4 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).

4.2.4.1 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

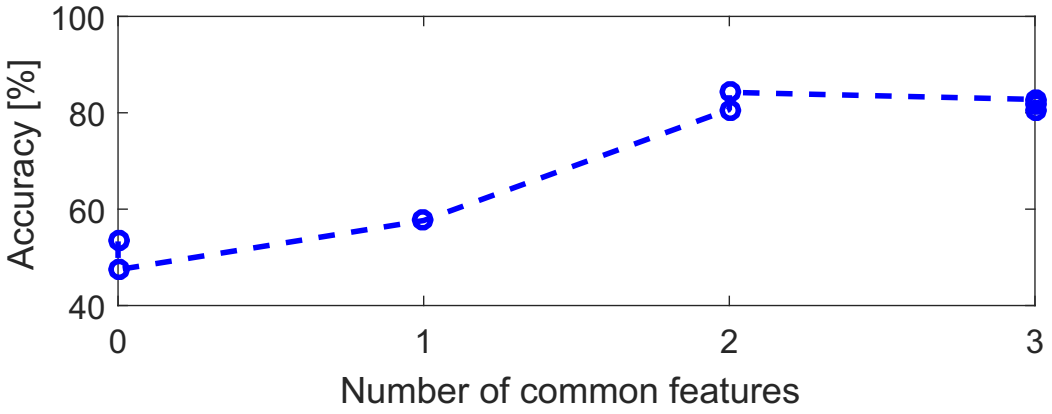

Figure 4.8: The correlation between accuracy of the reference classifier c_0 for different contingencies c_i and similarities in their MBs.

in Table 4.2. 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 operating conditions from the training database that were relevant for the new topology.

4.2.4.2 Generation of new data

An effective approach to generating new operating conditions is to considering only operating conditions that add knowledge to the database. In other words, operating conditions 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 chapter. The results in Table 4.2 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.

4.2.5 Training strategy

The training strategy of the proposed approach for DSA was compared to existing approaches (i-iii) introduced in Section 4.2.2, and additionally to a three-stages workflow using the proposed metric where the training is exclusively based on the new operating conditions (iv). Although this approach is computationally inefficient, it provides a theoretical upper limit for the accuracy achievable from new operating conditions. All approaches were tested on $1,500$ new operating conditions generated from the power flow distribution following the topology changes. The results of the test accuracy are shown in Table 4.3 for the 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 operating conditions 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 4.2.2), of 60% for new machines to be trained (Section 4.2.2) and of 55% for new data to be generated following high-impact topological changes (Section 4.2.4), 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 operating conditions. This approach is computationally inefficient as simulating the new operating conditions at each

Table 4.2: 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
Nº offline OCs	20,000	4,307	4,307
Nº new OCs	1,500	1,500	684
Accuracy	83.26%	85.54%	85.39%

topological change would significantly increase the time span in which no accurate security assessment is in place. The results were summarised in Table 4.4. The estimation showed that the proposed approach has the potential to reduce the time for data generation by up-to 85% from 472h to 71h.

4.2.6 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 predic-

Table 4.3: 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
Nº offline OCs	$20,000$	$20,000$	$4,307$	0
Nº new OCs	0	$1,500$	684	$1,500$
Accuracy	76.62%	76.61%(w.c)	85.39%	83.85%

Table 4.4: Estimation of computational times for data generation for a large system using different DSA workflows.

	Approach	
	(iii) Proposed workflow	(v) Frequent regular updates
Nº offline OCs	$1,000,000$	$7,000,000$
Nº new near real-time machines	400	$1,000$
Nº simulations per new trained machine	680	$1,500$
Simulation time per OC	$0.2s$	$0.2s$
Total	$71h$	$472h$

tion (Section 4.2.2) 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 4.1). 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.

4.3 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.

Chapter 5

Corrective Control for System Security

In the past the grid assets were treated as static infrastructure, which were fixed within the network, except during times of forced outages. Changes in the network's topology were mainly used for maintenance purposes. The introduction of transmission switching in the 1980s has changed this view and nowadays operators can control the grid assets to overcome various situations, such as voltage violations, line overloads [55], line losses and cost reduction [56], or system security [57]. Corrective switching, including transmission line switching, bus-bar switching, transformer tap changing, is one example of topology control which is implemented today [111]. The line disconnection following a short-circuit is another example of topology change for corrective control. In particular, ad-hoc connections of plug-and-play DERs play a key role in supporting a reliable grid service when the topology changes [88, 112]. The use of topology changes for corrective control is still very limited and, more generally, the use of corrective control is limited to emergency failures. Generally, the preventive control, e.g. generation rescheduling, aims at changing the system's operating condition once it is found to be insecure for a particular disturbance when the system is still in a secure state. Conversely, corrective control, e.g. load shedding or corrective switching in this case, is applied when the disturbance has already occurred and the system is already in an insecure state. However, in such a challenging scenario of modern operations where generation and demand are barely

predictable and the power electronic devices by which DERs are integrated into the system do not provide inertia, very conservative preventive measures or investments in new system infrastructure are required to keep the system stable if no corrective control is used. These two strategies are costly and inefficient [113].

Corrective control is an alternative strategy for efficiently and cheaply ensuring stable normal operations [114]. Corrective control adjusts system 'post-fault' operation in response to a disturbance or a fault [73,74]. Currently, corrective control is only used for emergency responses, and an example of control action is load shedding. A promising opportunity is to use corrective control also in normal operations leveraging the new technological advantages of very fast-acting (corrective) converter-interfaced generator (CIG) controllers. However, a research gap is a control method that can compute optimal corrective control measures for real-time considering the system stability in normal operations. It is challenging to consider corrective control in real-time system operation. The system operation can be modelled as a nonlinear ACOPF optimisation which can consider preventive control. However, the ACOPF cannot consider corrective control easily. Considering corrective control requires modeling post-fault system states and corresponding post-fault decisions in the optimisation. The post-fault dynamics are highly nonlinear differential-algebraic equations (DAEs) which require to model transient constraints in the optimisation. Modelling such constraints makes the ACOPF optimisation very slow. This is challenging as ideally the ACOPF would be solved every five minutes to account for rapid fluctuations of demand and generation, but this is infeasible in real-time.

Approaches to consider corrective control in real-time operation can either aim at predicting operation decisions with ML instead of solving the optimisations or aim at finding approximations for transient constraints for the ACOPF. As solving the ACOPF in real-time is not possible, recently, ML models are designed to predict the ACOPF solution. For example, NNs can either directly predict the optimal solution [115] or classification and regression models can predict the economic OPF costs [116], however, these approaches do not predict transients and are not developed yet for corrective control. Numerical approaches (Forward Euler or Runge-Kutta)

can compute the exact transients from the DAEs which are computationally hard above all in large systems. Approximations of the transients for ACOPFs can be easier computed through heuristics, discretisations, simplifications, or energy functions [117, 118]. Heuristics connect the ACOPF optimiser with a simulator. The ACOPF generation dispatch is simulated for dynamic stability [73], and when found unstable, the dispatch is modified until the system stabilises. NNs can be also used to substitute dynamic simulations in some iterations as in [119]. Another approach discretises the DAEs in time [120, 121] and considers the discretised transient constraints in the ACOPF which can result in many constraints and significantly slows down ACOPFs in larger systems. The single-machine equivalent (SIME) approach [122] simplifies the dynamical model. However, the simplified stability limits for each generator cannot consider coupling effects. Another approach uses energy functions and derives the certification of safety [123, 124]. The energy functions are a specific form of Lyapunov functions that guarantee the system convergence to stable equilibrium points. The sublevel sets of these functions can be then used as region of attraction (ROA in Fig. 5.1) that represents the subspace of all operating states converging to a steady-state equilibrium [5]. The ROA can be a security constraint for the ACOPF. This approach does not require computing the full system transients, results in low computational cost, however, is only applicable for very simple systems. Unfortunately, all of the aforementioned approaches are unsuitable for real-time corrective control as they either cannot consider corrective actions or are too slow for real-time.

Recently, a novel approach from ML seems promising to address a key issue in identifying Lyapunov functions. The key issue of the Lyapunov function based approach is to identify the function for large and complex systems [125]. Particularly, for complex dynamical systems such as power systems the functions are very hard to find [126]. For instance, a semi-definite program is efficient only when the dynamics are polynomial and the Lyapunov function is restricted to be a sum-of-squares polynomial [5]. However, assuming linear or polynomial approximations pose much restriction on the system and the Lyapunov function. Recently, in ML research, NNs seems suitable to model the Lyapunov function and avoid linear or polynomial approximations [127–129]. In [130], a NN learns the control law and the Lyapunov function

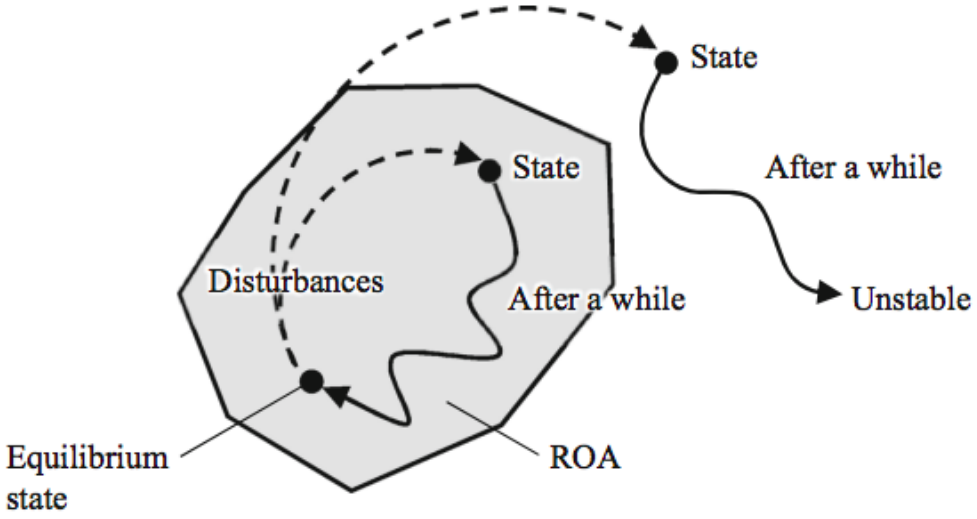

Figure 5.1: Definition of region of attraction [5].

that maximise the ROA of a general nonlinear dynamical system. The learner uses stochastic gradient descent to find the optimal parameters of the control law and the neural Lyapunov function that minimises the Lyapunov risk. This risk measures the violations of the Lyapunov conditions. This NN-based approach can assess the transient stability of power systems in [131–133], however, this approach was never used to consider the system's transient response within the ACOPF.

5.0.1 Proposed approach

This chapter proposes the NN-LYAPUNOV CONTROL approach, an OPF-based optimisation that considers NN-trained ROAs as stability constraints and preventive and corrective control as decisions. The flowchart of the proposed control approach is shown in Fig. 5.2. The NN training approach transforms offline the differential and nonlinear constraints for transient post-fault stability into algebraic constraints which are manageable by the OPF optimisation. The OPF is then solved online for each next OC to obtain the optimal pre- and post-fault setpoints. The proposed NN-LYAPUNOV CONTROL estimates and maximises the ROA of the post-fault system [130] resulting in highly efficient corrective and preventive control actions. The contribution of this work is threefold:

1. Deriving algebraic stability constraints for the post-fault system using NN-based Lyapunov functions. The NNs learn the optimal controller and the Lyapunov function to obtain larger ROAs compared against standard control methods. This optimises the balance of transient stability and economic cost.

2. Considering the system's transient response within the OPF using the derived algebraic constraints. The DAEs for transient stability are transformed into algebraic operational constraints using the learned ROA for the post-fault system. This results in transient stable, cost-optimal post-fault OCs.

3. Using corrective control not only as a backup strategy but as an active role in maintaining the system stability. This reduces operational costs and carbon emissions while maintaining adequate levels of stability.

A case study tested the approach on the IEEE 9 bus system with integrated DERs for two scenarios, for 25% of renewable sources and for 40% of renewable sources. The scenarios considered wind farms and fleets of electric vehicles with V2G technology as corrective control. The study compares the approach against baselines such no control, only preventive control and ML-based corrective control. Case studies on the IEEE 39-bus and 118-bus systems finally tested the stability performance of the proposed approach on large systems.

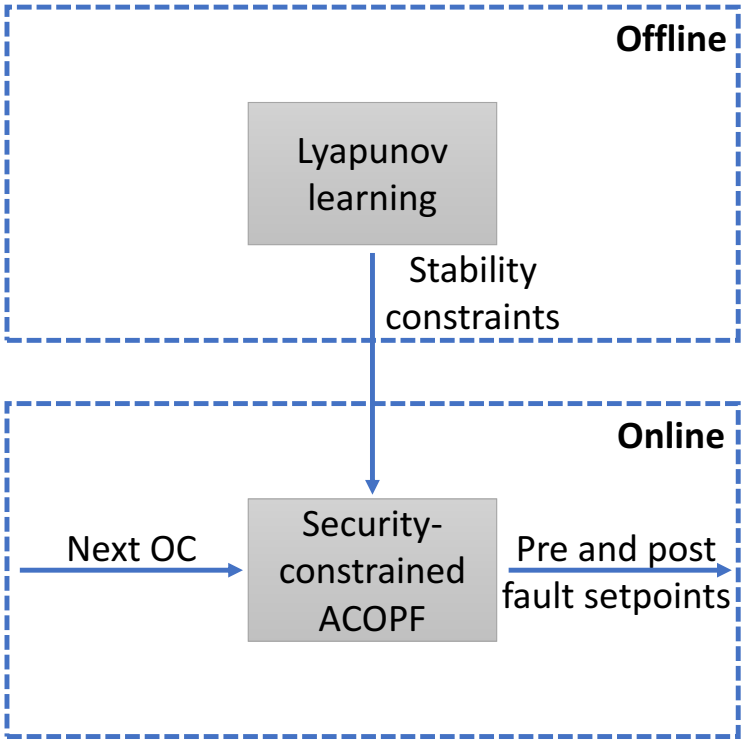

Figure 5.2: The proposed control approach for security-constrained ACOPF.

5.1 Power System's Dynamics

The classical model of a multimachine system is used [134]. Let N be the number of buses, ε the set of transmission lines, B_{ij} the susceptance matrix of lines $(i,j) \in \varepsilon$, G the set of synchronous generator buses and B the set of load buses. In this model, the loads are represented by passive impedances and the mechanical power $P_{m,i}$ for each generator bus $i \in G$ is assumed constant over the timescale of transients. The Kron reduced model is used to aggregate the load buses into generator buses. Each generator bus has the conventional momentum of inertia $M_i = \dfrac{\omega_R}{2H_i}$ where H_i is the inertia and ω_R the frequency reference, and the damping factor D_i. The dynamical model of a multimachine system is a second order nonlinear and differential equation, also known as swing equations:

$$\begin{aligned} \dot{\delta}_i &= \omega_i, \ \forall i = 1...|G| \\ M_i\dot{\omega}_i &= P_{m,i} - \sum_{j=1,j\neq i}^{N} B_{i,j}\sin(\delta_i - \delta_j), \ \forall i = 1...|G| \end{aligned} \tag{5.1}$$

with δ_i and ω_i the phase angle and frequency at each generator bus i. The transient stability is then assessed using the Integral Square Generator Angle (ISGA) index that is defined as follows [135]:

$$\begin{aligned} \delta_{coak} &= \frac{1}{M_{tot}} \sum_{i \in G} M_i \delta_{i,k}, \ \forall k = 0, ...T \\ J_k &= \frac{1}{M_{tot}} \sum_{i \in G} M_i (\delta_{i,k} - \delta_{coak})^2, \ \forall k = 0, ...T \\ \bar{J} &= \frac{1}{|T|} \sum_{k=0,...T} J_k \end{aligned} \tag{5.2}$$

where δ_{coak} is the centre of angle at the time step k and $M_{tot} = \sum_{i \in G} M_i$ is the total inertia. The ISGA index $\bar{J}$ is the time-average of the element-wise J_k. Higher values of $\bar{J}$ correspond to unstable operating conditions following a disturbance.

5.2 Learning the Lyapunov Function for Stability

The key role of the ROA in power system stability is described in the following transient stability problem (adapted from [136]). Suppose that at time t_0 the power system is subjected to a severe transient disturbance (fault), e.g. a short circuit. During the fault, the system responds by large excursions of the system variables. At t_1 the fault is cleared, the system reaches a new state $x_1(t_1)$ and switches to post-fault system. The transient stability problem considers whether the trajectory $x(t \geq t_1)$ with initial conditions $x(0) = x_1(t_1)$ will converge to an asymptotically stable equilibrium point. The system will return to steady-state operation only if $x_1(t_1)$ belongs to the ROA of the post-fault steady-state. Hence, the larger the ROA, the more operating conditions can reach post-fault stability, and a system with large ROA can be considered robust against large disturbances.

Several approaches can estimate the ROA of a generic nonlinear system such as the power system. A straightforward approach is to use time-domain simulations to check every point in

the neighborhood of the stable equilibrium point. This approach provides the exact ROA but is impractical for large-scale systems due to high computational costs and does not provide any closed form for control design purposes meaning that it does not allow to enlarge the ROA. More promising is the approach based on Lyapunov functions where the sublevel sets of these functions estimate the ROA. The corresponding training of the NN to identify the optimal controller and Lyapunov function is described [130].

Given the system's dynamics $(\dot{\delta}, \dot{\omega}) = f_u(\delta, \omega)$ described in Eq. (5.1) with controller $u = P_m$ and state $x = (\delta, \omega)$, a Lyapunov function V can be used to establish the stability of the post-fault state $x^* = (\delta^*, \omega^*)$ as follows:

Definition 1: If in a ball $D_R = \{x | ||x||_2^2 \leq R^2\}$ with radius R, there exist a continuous differentiable scalar function V such that:

1. V is positive definite in D_R
2. $\dot{V} = \dfrac{dV(x(t))}{dt} = \mathcal{L}_{f_u} V(x(t))$ is negative definite in D_R

then $x^* = (\delta^*, \omega^*)$ is asympotically stable and V is a Lyapunov function.

The sublevel set of the Lyapunov function S_c $\forall c \geq 0$ is defined as

$$S_c = \{x \in D_R | V(x) \leq c\} \tag{5.3}$$

and can be used to approximate the ROA. The ROA is an invariant subset such that all system trajectories starting inside this subset asymptotically converge back to the post-fault state. For a state deviation from x^* falling within the ROA, the system can be assessed to be stable as the system trajectories resulting from the deviation will converge back to x^*. Therefore, the ROA can be used to certify the stability of the post-fault state as long as a Lyapunov function for the post-fault system can be identified.

A multilayered feedforward NN with *tanh* activation function is assumed as structure of the Lyapunov function. The learning framework, shown in Fig. 5.3, is an unsupervised learning

task that is composed of a learner and a falsifier. Using u to denote both the NN parameters and the controller, the learning module in Fig. 5.3 updates u to improve the likelihood to satisfy the Lyapunov conditions that are formulated as a cost function also named the Lyapunov risk. The Lyapunov risk measures the degree of violation of the following conditions: i) $V_u(x)$ is positive, ii) $\mathcal{L}_{f_u}V_u(x)$ is negative, iii) $V_u(x^*) = 0$. Therefore, the design objective is to minimise the following Lyapunov risk $L(u)$ by updating the controller and parameters u:

$$\begin{aligned} L(u) =& \frac{1}{N}\sum_{k=1}^{N}\Big(\max(0, -V_u(x_k)) + \max(0, \mathcal{L}_{f_u}V_u(x_k))\Big) \\ &+ V_u^2(x^*) + \beta \end{aligned} \tag{5.4}$$

where N is the number of training samples and β is the cost term that regulates how quickly the Lyapunov function value increases with respect to the radius of the level sets:

$$\beta := \frac{1}{N}\sum_{k=1}^{N} ||x_k||_2 - \alpha V_u(x_k) \tag{5.5}$$

where α is a tunable parameter. The cost term β allows to adjust the controller and the NN parameters u to maximise the corresponding ROA of the learned Lyapunov function. Because of the local nature of stochastic gradient descent, it is hard to learn good control functions through random initialization of control parameters. Therefore, the parameters u are initialized to the linear quadratic regulator (LQR) solution that is obtained by linearising the dynamics in a small neighborhood of the origin [137]. On the other hand, the initialization of the neural network Lyapunov functions can be completely random.

As a second step of the diagram in Fig. 5.3, for each learned controller and Lyapunov function pair (V_u, u), there is a falsifier that finds the states violating the Lyapunov conditions, also called counterexamples, and these counterexamples are then added to the training set for the

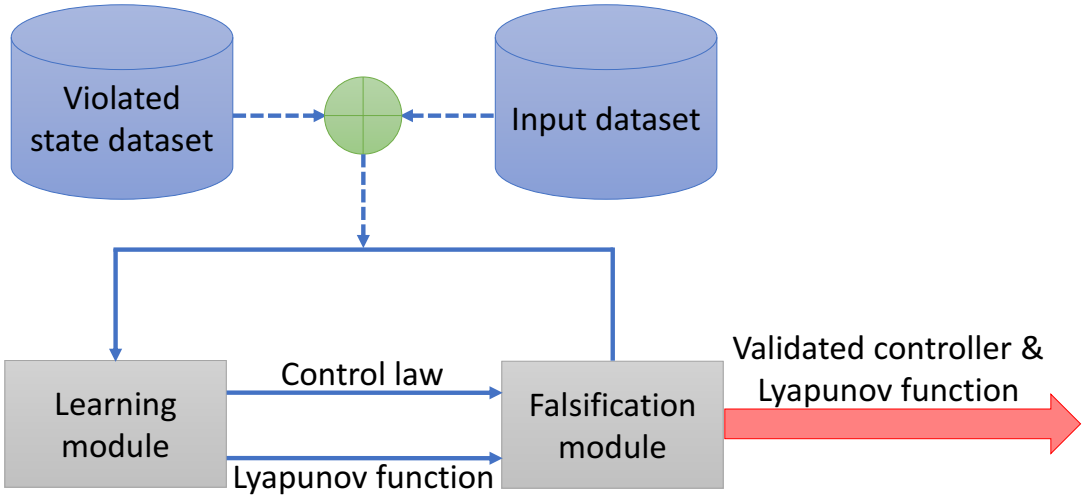

Figure 5.3: The neural learning of the optimal controller and Lyapunov function.

next learning step. The falsification constraints are defined as follows:

$$\Phi_\varepsilon(x) := \left(\sum_{k=1}^{N} x_k^2 \geq \varepsilon\right) \wedge \left(V_u(x) \leq 0 \vee \mathcal{L}_{f_u} V_u(x) \geq 0\right) \tag{5.6}$$

where ε is a positive constant parameter that bounds the tolerable numerical error. This parameter allows to avoid numerical issues as values within the ball with radius ε are physically insignificant and does not affect the learned Lyapunov function. To solve the falsification constraints that would require the minimisation of highly nonconvex functions, the SMT solver dReal based on the δ-completeness property is used [138]. When the δ-complete algorithm concludes that Eq. (5.6) is unsatisfiable, then the Lyapunov conditions hold for all states, otherwise counterexamples are obtained and added to the training set. The learning stops when no counterexamples are found. The outputs of the neural learning are the parameters u of the validated Lyapunov function V_u and the weight matrix W of the corresponding optimal controller $u = g(x) = W \cdot x$.

Once the NN Lyapunov function for the post-fault system is learned, the corresponding ROA provides bounds for the maximum state deviation for which the system trajectories are bounded within the ROA and asymptotically converge to the origin. In the steady-state post-fault

system $\omega = \dot{\delta} = 0$, then the lower bound in the state space is $\delta_{\min} = (\delta, \omega = 0) \in S_{c_0}$ with $S_{c_0} = \{S_c | c = 0\}$ such that

$$\sum_{i=1}^{|G|} |\delta_i - \delta_{i,\min}|^2 \text{ is minimised} \tag{5.7}$$

where $\delta_{i,\min} = \min\{\delta_i : (\delta, \omega = 0) \in S_{c_0} \ \wedge \ \delta_{j \neq i} = 0\}$. Similarly, the upper bound in the state space is $\delta_{\max} = (\delta, \omega = 0) \in S_{c_0}$ such that

$$\sum_{i=1}^{|G|} |\delta_i - \delta_{i,\max}|^2 \text{ is minimised} \tag{5.8}$$

where $\delta_{i,\max} = \max\{\delta_i : (\delta, \omega = 0) \in S_{c_0} \wedge \delta_{j \neq i} = 0\}$. Finally, the bounds in the input space can be easily derived as $u_{\min} = g(\delta_{\min}, \omega = 0)$ and $u_{\max} = g(\delta_{\max}, \omega = 0)$ and used as operational constraints in the following optimisation problem.

5.3 Stability Constrained ACOPF

This section formulates the optimisation problem to identify a feasible operating condition that fulfills all power network constraints in the pre-fault state and the power networks and stability constraints in the post-fault state. The subscripts p and P are used to differentiate the pre-fault and the post-fault variables, respectively. The objective function is the minimisation of the operational costs and hence of the pre-dispatched generation costs. No costs are assumed for corrective control following the fault. The proposed optimisation problem is a modification of a relaxed ACOPF formulation that uses the Semi Definite Programming (SDP) relaxation and efficiently finds a global optimal solution [139]. This modification minimises the mean squared distance from a predicted generator dispatch $[\hat{P}_{G_i}, \hat{Q}_{G_i}]$ to a feasible generator dispatch $[P_{G_i,p}, Q_{G_i,p}]$. These mean squared differences in the pre-fault active and reactive powers are α_i and γ_i for generator $i \in G$. Then, the full optimisation problem is:

$$\min \sum \alpha_i + \lambda \gamma_i \tag{5.9a}$$

subject to

$$P_{G_i,P} = P_{G_i,p} + \Delta P_{G_i} \qquad \forall i \in G \tag{5.9b}$$

$$P_{G_i}^{\min} \leq P_{G_i,p} \leq P_{G_i,p}^{\max} \qquad \forall i \in G \tag{5.9c}$$

$$Q_{G_i}^{\min} \leq Q_{G_i,p} \leq Q_{G_i,p}^{\max} \qquad \forall i \in G \tag{5.9d}$$

$$u_{\min} \leq \Delta P_{G_i} \leq u_{\max} \qquad \forall i \in G \tag{5.9e}$$

$$-P_{Db} + \sum_{i \in G} P_{G_i,p} = \mathrm{tr}(\mathbf{Y}_{b,p}\mathbf{W}_p) \qquad \forall b \in B \tag{5.9f}$$

$$-P_{Db} + \sum_{i \in G} P_{G_i,P} = \mathrm{tr}(\mathbf{Y}_{b,P}\mathbf{W}_P) \qquad \forall b \in B \tag{5.9g}$$

$$-Q_{Db} + \sum_{i \in G} Q_{G_i,p} = \mathrm{tr}(\bar{\mathbf{Y}}_{b,p}\mathbf{W}_p) \qquad \forall b \in B \tag{5.9h}$$

$$-Q_{Db} + \sum_{i \in G} Q_{G_i,P} = \mathrm{tr}(\bar{\mathbf{Y}}_{b,P}\mathbf{W}_P) \qquad \forall b \in B \tag{5.9i}$$

$$(V_b^{\min})^2 \leq \mathrm{tr}(\mathbf{M}_{b,p}\mathbf{W}_p) \leq (V_b^{\max})^2 \qquad \forall b \in B \tag{5.9j}$$

$$(V_b^{\min})^2 \leq \mathrm{tr}(\mathbf{M}_{b,P}\mathbf{W}_P) \leq (V_b^{\max})^2 \qquad \forall b \in B \tag{5.9k}$$

$$\begin{bmatrix} -(S_k^{\max})^2 & \mathrm{tr}(\mathbf{Z}_{kl}\mathbf{W}_p) & \mathrm{tr}(\bar{\mathbf{Z}}_{kl}\mathbf{W}_p) \\ \mathrm{tr}(\mathbf{Z}_{kl}\mathbf{W}_p) & -1 & 0 \\ \mathrm{tr}(\bar{\mathbf{Z}}_{kl}\mathbf{W}_p) & 0 & -1 \end{bmatrix} \leq 0 \, \forall k \in \varepsilon \tag{5.9l}$$

$$\begin{bmatrix} -(S_k^{\max})^2 & \mathrm{tr}(\mathbf{Z}_{kl}\mathbf{W}_P) & \mathrm{tr}(\bar{\mathbf{Z}}_{kl}\mathbf{W}_P) \\ \mathrm{tr}(\mathbf{Z}_{kl}\mathbf{W}_P) & -1 & 0 \\ \mathrm{tr}(\bar{\mathbf{Z}}_{kl}\mathbf{W}_P) & 0 & -1 \end{bmatrix} \leq 0 \, \forall k \in \varepsilon \tag{5.9m}$$

$$\begin{bmatrix} -(S_k^{\max})^2 & \mathrm{tr}(\mathbf{Z}_{km}\mathbf{W}_p) & \mathrm{tr}(\bar{\mathbf{Z}}_{km}\mathbf{W}_p) \\ \mathrm{tr}(\mathbf{Z}_{km}\mathbf{W}_p) & -1 & 0 \\ \mathrm{tr}(\bar{\mathbf{Z}}_{km}\mathbf{W}_p) & 0 & -1 \end{bmatrix} \leq 0 \, \forall k \in \varepsilon \tag{5.9n}$$

$$\begin{bmatrix} -(S_k^{\max})^2 & \mathrm{tr}(\mathbf{Z}_{km}\mathbf{W}_P) & \mathrm{tr}(\bar{\mathbf{Z}}_{km}\mathbf{W}_P) \\ \mathrm{tr}(\mathbf{Z}_{km}\mathbf{W}_P) & -1 & 0 \\ \mathrm{tr}(\bar{\mathbf{Z}}_{km}\mathbf{W}_P) & 0 & -1 \end{bmatrix} \leq 0 \, \forall k \in \varepsilon \tag{5.9o}$$

$$\begin{bmatrix} -\alpha_i & P_{G_i,p} - \hat{P}_{G_i} \\ P_{G_i,p} - \hat{P}_{G_i} & -1 \end{bmatrix} \leq 0 \qquad \forall i \in G \tag{5.9p}$$

$$\mathbf{W}_p \geq 0 \tag{5.9q}$$

$$\mathbf{W}_P \geq 0 \tag{5.9r}$$

where λ is a user-defined scaling parameter to cope with the different orders of magnitudes of α and γ. Pre- and post-fault active and reactive power injections P_{G_i}, Q_{G_i}, the post-fault ΔP_{G_i} and pre- and post fault voltage matrices $\mathbf{W_p}$ and $\mathbf{W_P}$ are the decision variables. Equation (5.9e) guarantees the post-fault state is transient stable with $u_{\min}$ and $u_{\max}$ the bounds introduced in Section 5.2 and ΔP_{Gi} the deviation from the pre-fault generator dispatch due to corrective control. All other parameters are introduced and better explained in [139]. The pre-fault voltage coordinates $w_p = [V_{d1,p}, ...V_{dN,p}, V_{q1,p}, ..V_{qN,p}]$ with $V_{i,p} = V_{di,p} + V_{qi,p}$ the voltage phasor in rectangular coordinates, can be recovered through

$$\text{rank}(\mathbf{W}_p) = \text{rank}(w_p w_p^T) = 1 \tag{5.10}$$

This optimisation is convex and can be solved with a second-order cone solver, e.g. Mosek, SCS [139]. The outputs are the optimised pre-fault generator dispatches (P_{Gi}, Q_{Gi}), the generator bus voltages (V_{Gdi}, V_{Gqi}) and the corrective power injections ΔP_{Gi} for each generator bus $i \in G$.

5.4 Case Study

This section provides a brief tutorial on the proposed approach to learn the Lyapunov based stability constraints, subsequently analyses the stability of the proposed approach and the reductions of economic costs and carbon emissions.

5.4.1 Test system and assumptions

A modified version of the IEEE 9-bus system from [140] was used as test system in the following studies, unless indicated otherwise. The modification included integrated DERs, all lines had a minimum resistance of 10^{-4} p.u., a storage capacity of 20MWh was available at each generator bus using EV fleets, and two scenarios with renewable sources were considered: a) 25% of fossil fuel generation was replaced by wind power, b) 40% of fossil fuel generation was replaced by wind power. Generation redispatch and energy storage using the EVs were considered for preventive and corrective control [141], respectively. 1000 load scenarios were sampled from a Latin hypercube with uniform distribution around 50% of the nominal value for the active power and 20% for the reactive power. A short circuit at bus 8 at time 0.1s was considered as fault. The fault was then cleared at 0.25s by opening the line between buses 8 and 9. Corrective control was applied only once the fault was cleared as there is always a latency of up to 0.25s between the fault occurrence and the real activation of the corrective control due to communication delays. During the time interval $[0.1, 0.25]$s, transient assistive measures (TAMs) were applied [73]. These measures usually last for few milliseconds, so they are not optimised. The transient stability was analysed over a simulation time $T = 10$s and the optimal post-fault operating condition was then assessed as stable if the index ISGA ≤ 0.47, otherwise unstable [135]. 93% and 99% of OCs were unstable when no preventive and corrective control was applied at scenarios (a) and (b), respectively. These high rates of unstable OCs showed that high shares of renewable sources make the system strongly unstable.

The NN structure to learn the Lyapunov function had 3 linear layers, one input, one hidden with 6 neurons and one output layer, all using *tanh* as activation function. 1000 training data $x_i = [\delta_1, ..., \delta_{|G|}, \dot{\delta}_1, ..., \dot{\delta}_{|G|}]$, with $i = 1...1000$ and $|G| = 3$, were fed into the input layer. Each training data x_i was uniformly sampled between $[-1, 1]$ p.u. corresponding to the maximum phase angle deviation 57° from the initial pre-fault condition. The NN structure was implemented using the package PyTorch 1.7 with Python 3.8.5 [142]. The optimiser ADAMW was used with a learning rate of 0.01. Finally, the optimisation problem (5.9a)-(5.9r) was implemented in CVXPY 1.1.5

with $\lambda = 100$ and solved using the SCS solver with default settings. The slack bus phase angles were set at 0° to get unique solutions to the optimisation. The time-domain simulations were then carried out in Matlab R2016a using a sixth stage-fifth order Runge-Kutta method (ode45 function).

The performance of the proposed approach was finally tested on larger systems using: (c) the IEEE 39-bus system with 10 machines and system parameters taken from [133], and (d) the IEEE 118-bus system with 19 machines and system parameters taken from [143]. In both systems, a storage capacity of 30MWh was available at each generator bus using EV fleets. There, the NN to learn the Lyapunov function had 5 linear layers, with 4 nodes at each hidden layer. 500 load scenarios were sampled with uniform distribution around 5% of the nominal value for the active and reactive power. The same transient assumptions of the IEEE 9-bus system were considered for these two larger systems. Short circuit faults were considered at buses 4 and 12 in (c) and (d), respectively.

5.4.2 Why NNs to learn the Lyapunov function

This study investigated the learning of the NN based Lyapunov function and compared the proposed approach with regards to stability and cost for preventive control against a more analytical learning approach for Lyapunov functions. The Lyapunov function for the post-fault system was learned by the

(i) 3 layers feed-forward NN-LYAPUNOV approach (Section 5.2)

(ii) initial LQR LYAPUNOV controller approach.

The neural Lyapunov learning converged in 3110 iterations by setting the numerical error parameter $\varepsilon = 0.5$ and $\delta = 0.01$ for the falsification step. This learning procedures found a Lyapunov function that is proved to be valid within the region $||x||_2 \leq 1$ and the following

optimal controller:

$$u = W \cdot x \tag{5.11}$$

with

$$\begin{aligned} u =& [P_{m,1}, P_{m,2}, P_{m,3}] \\ x =& [\delta_1, \delta_2, \delta_3, \dot{\delta_1}, \dot{\delta_2}, \dot{\delta_3}] \\ W =& \begin{bmatrix} -0.102 & -0.469 & -0.256 & 0.021 & -0.217 & 0.231 \\ 0.108 & -0.342 & -0.247 & -0.736 & -0.152 & 0.056 \\ -0.287 & -0.504 & -0.195 & 0.109 & -0.216 & -0.615 \end{bmatrix} \end{aligned} \tag{5.12}$$

The learned NN-based Lyapunov function is shown in Fig. 5.4 for the pair $(\delta_1, \dot{\delta}_1)$ with the dashed red circle defining the valid region. For both scenarios a) with 25% and b) with 40% wind power described in Section 5.4.1, the Lyapunov function for the post-fault system was learned only once as the Lyapunov function is an invariant property of the nonlinear system itself, i.e. the conclusion of the system stability is independent from the initial conditions [144]. Therefore, the same Lyapunov function can be used to determine the stability region in both scenarios. The corresponding ROAs of the Lyapunov functions learned using (i)-(ii) are compared in Fig. 5.5. It resulted the NN learning allowed to significantly enlarge the ROA compared to the LQR solution. The larger ROA enhanced the balance between costs and system stability as shown below.

The stability bounds were obtained by setting $\dot{\delta}_i = 0$ as the post-fault operating condition is a steady-state condition with time derivatives of the generator phase angles equal to zero. These state-space bounds $(\delta_{\min,i}, \delta_{\max,i})$ were finally replaced in Eq. (5.11) to derive the corresponding bounds in the input space, $u_{\min} = [-18, -5, -29]$ MW and $u_{\max} = [18, 6, 27]$ MW, to be included in the ACOPF (Eqs. (5.9a)-(5.9r)) as stability constraints.

The need of a NN based learning for the optimal controller and Lyapunov function was investigated by comparing the approaches (i)-(ii) with regards to stability and cost of preventive control. The ACOPF was solved using the two stability constraints derived from the NN and

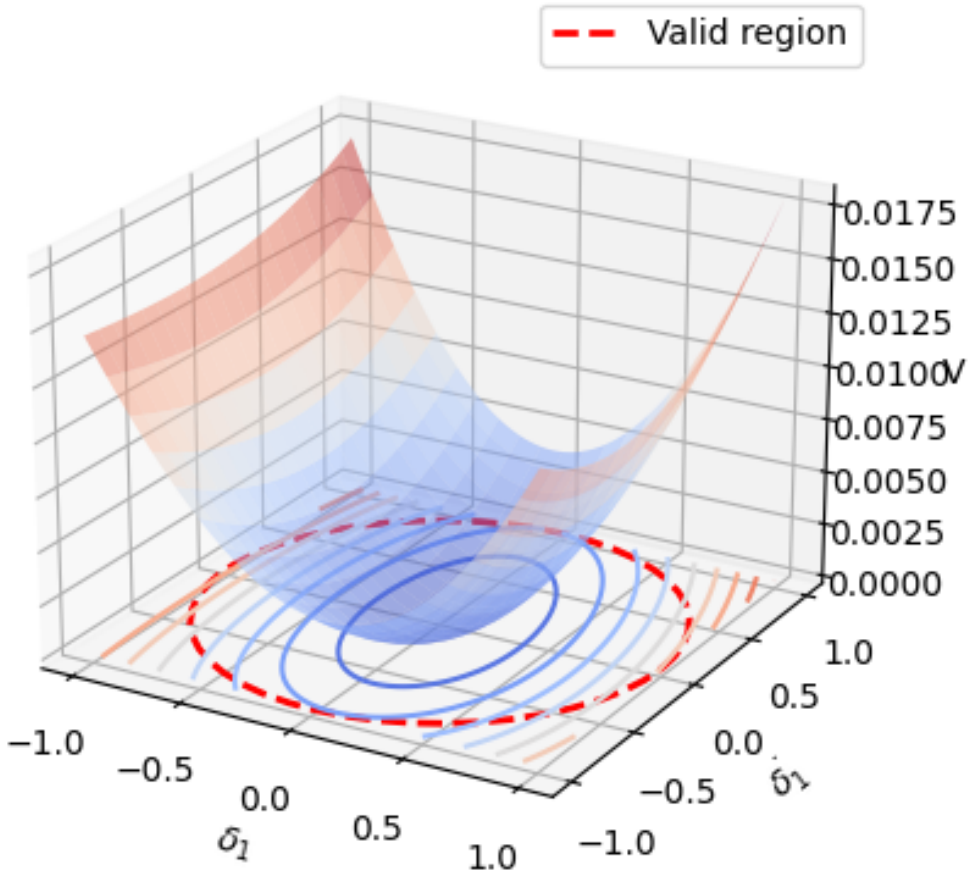

Figure 5.4: Lyapunov function learned using the NN learning procedure.

the LQR based Lyapunov functions. Then, the fault was simulated using the optimal pre-fault generator dispatch as initial conditions and applying the optimised corrective power injections when the fault was cleared.

The stability and cost results for approaches (i)-(ii) are summarised in Table 5.1 for scenario (a)-(b) and compared against the baseline approach WIND CURTAIL. In the baseline approach, the wind power was curtailed and replaced by fossil fuel generation as preventive control without carrying out any optimisation. In this way, the system's inertia was increased to improve the system's response to the fault. The preventive curtailment of the wind power almost halved the unstable OCs, however both the approaches (i)-(ii) outperformed the WIND CURTAIL approach by making the system transient stable for all the studied 1000 load scenarios.

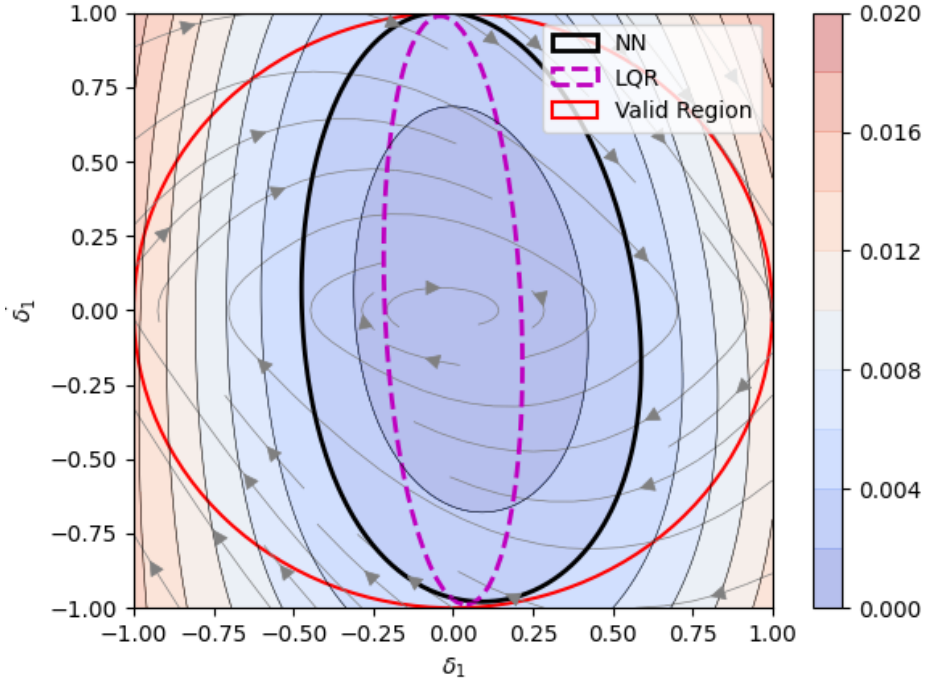

Figure 5.5: ROAs estimated using different Lyapunov functions.

Table 5.1: Stability and costs of NN and LQR based Lyapunov functions with 25% (a) and 40% (b) renewables.

	Approaches		
	WIND CURTAIL	(i) NN LYAPUNOV	(ii) LQR LYAPUNOV
$(a)-Unstable\ OCs$	418/1000	0/1000	0/1000
$(a)-Cost\ reduction$	–	−2.7m (−48%)	−2.6m (−46%)
$(b)-Unstable\ OCs$	510/1000	0/1000	0/1000
$(b)-Cost\ reduction$	–	−3.7m (−62%)	−3.6m (−60%)

In terms of cost savings, approach (ii) decreased the cost of preventive control by 46% and 60% for scenario (a)-(b) compared to the WIND CURTAIL approach. Approach (i) based on the neural learning of the Lypaunov function further reduced these costs by 2%. The higher cost reduction when using approach (i) instead of approach (ii) confirmed that using the LQR controller to find the Lyapunov function did not allow to fully leverage the available storage capacity from EVs as the corresponding ROA and the consequent transient bounds significantly restricted the deviations in generation from the initial pre-fault OC. Therefore, a larger generation redispatch was considered to maintain the system transient stable resulting in an increase of the operating costs for preventive control.

5.4.3 Stability analysis

In this study, three control approaches for cost-optimal transient stable operations were compared in terms of stability:

(i) optimised PREVENTIVE CONTROL

(ii) DT CONTROL, a decision tree based corrective control

(iii) NN-LYAPUNOV CONTROL.

The PREVENTIVE CONTROL approach optimised the generation redispatch to guarantee transient stability using the Lyapunov based stability constraints, however no corrective control was applied. The DT CONTROL approach optimised only the generation redispatch as preventive control. However, no constraints for transient stability were considered in the optimisation problem. The stability of the optimised pre-fault OC was then assessed using a DT. When the OC was unstable, corrective control was used to reduce the difference in generation between this optimised OC and the closest stable one in terms of euclidean distance [145]. The proposed NN-LYAPUNOV CONTROL approach optimised preventive and corrective control to balance operating costs and system stability (Eqs. (5.9a)-(5.9r)). Following the fault, TAMs were used for 0.15s before the activation of corrective control in all approaches.

The stability results for these three approaches are summarised in Table 5.2 for scenarios (a)-(b). It resulted that the proposed NN-LYAPUNOV CONTROL outperformed the baseline approach in Table 5.1 and the DT CONTROL approach in Table 5.2 by reducing the instability rate by 42% and 10% for scenario (a) and by 51% and 24% for scenario (b). Although similar stability performance of the NN-LYAPUNOV CONTROL were expected when using the PREVENTIVE CONTROL approach as the same stability constraints were considered in the optimisation, the number of unstable OCs slightly increased for scenario (a) with PREVENTIVE CONTROL. This is because the use of only preventive control to transient stability increased the pre-fault gen-

Table 5.2: Stability and costs of three control approaches with 25% (a) and 40% (b) renewable integration.

	Approaches		
	PREVENTIVE CONTROL	DT CONTROL	LYAPUNOV CONTROL
$(a) - Unstable\ OCs$	23/1000	95/1000	0/1000
$(a) - Cost\ reduction$	−2m (−35%)	−3.1m (−54%)	−2.8m (−48%)
$(b) - Unstable\ OCs$	0/1000	241/1000	0/1000
$(b) - Cost\ reduction$	−3.2m (−53%)	−4.1m (−67%)	−3.7m (−62%)

eration so much to make the TAMs less effective, and hence increased the number of unstable OCs.

5.4.4 Quantifying the value of corrective control

This study investigated the value of corrective control in terms of cost savings and CO2 emissions. The proposed control approach to transient stability combines preventive and corrective control to reduce the operating costs of preventive control and prevent wind power curtailment.

To quantify the cost benefits of actively using corrective control for system stability, the three approaches described in Section 5.4.3 were compared against the WIND CURTAIL approach with regards to the cost of preventive control. The cost results are summarised in Table 5.2. As expected, the WIND CURTAIL approach resulted in the highest operational costs as fossil fuel generation replaced the curtailed wind power to meet the power balance and the cost of fossil fuel generation is much higher than wind power. However, the proposed NN-LYAPUNOV CONTROL reduced the costs by 48% and 62% for scenario (a) and (b), respectively, compared to the WIND CURTAIL approach as the optimised use of corrective control allowed preventing wind curtailment. Similarly, the optimised PREVENTIVE CONTROL resulted in a lower cost reduction than the corrective based approaches as no resources were available in addition to the preventive strategies to maintain the system transient stable. Although the DT CONTROL approach resulted in slightly higher reductions of the operating costs compared to the proposed NN-LYAPUNOV CONTROL, this approach cannot guarantee the same high stability performance

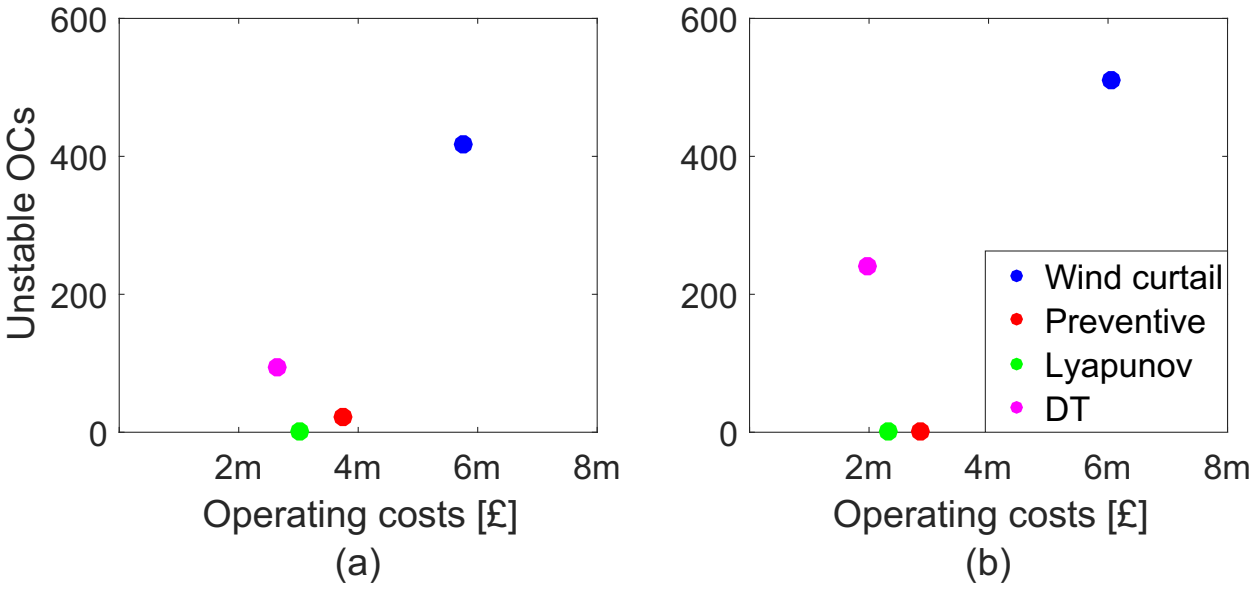

Figure 5.6: The balance between stability and operating costs for four control approaches with 25% (a) and 40% (b) renewable integration.

of the proposed one. The key finding is that the NN-LYAPUNOV CONTROL proposed approach resulted in the best balance between system stability and operating costs for preventive control as shown in Fig 5.6 (the proposed approach is in green).

To quantify carbon emissions, the proposed control approach prevented the curtailment of 128MWh and 192 MWh of wind power for each unstable OC for scenario (a)-(b), respectively, corresponding to a reduction of 420 kg of CO2 emissions for each uncurtailed MWh. For scenario (b) with 40% of wind power generation and with the 99% of unstable OCs, the proposed approach reduced the CO2 emissions by up-to 80m kg corresponding to the 60% of the total CO2 emissions when using WIND CURTAIL approach.

5.4.5 Performance on a larger systems

In this study, the proposed approach was tested on the (c) IEEE 39-bus and (d) 118-bus systems. For the case study settings, the numerical error parameter was set at $\varepsilon = 0.5$ and $\delta = 0.01$ for the falsification step, so that the neural Lyapunov learning converged in 4668

Table 5.3: Stability of three control approaches in large systems.

	Approaches		
	NO CONTROL	LQR CONTROL	LYAPUNOV CONTROL
$(c) - Unstable\ OCs$	470/500	425/500	0/500
$(d) - Unstable\ OCs$	347/500	178/500	0/500

and 9275 iterations for (c) and (d), respectively. The proposed NN-LYAPUNOV CONTROL was compared against i) the NO CONTROL approach when no preventive and corrective control was applied, and ii) the LQR-LYAPUNOV CONTROL where the stability constraints for the security-constrained OPF were derived from the initial LQR controller.

The stability results are summarised in Table 5.3. There, the NN-LYAPUNOV CONTROL outperformed approaches (i)-(ii) by making the system transient stable for all the studied 500 load scenarios in both systems. This result demonstrates that enlarging the ROA can leverage the available EV storage capacity to improve the system stability in these larger systems.

The computational times of the proposed NN-LYAPUNOV CONTROL approach for different system sizes are also investigated. Two components need to be considered: i) computational time to learn the Lyapunov function; ii) computational time to solve the ACOPF. The learning times of the Lyapunov function for different systems are in [133], showing that training a NN for the IEEE 118-bus system can take around 45min. Considerations with this long training time is part of the discussion section. The solver time for the SDP based ACOPF is only 2.1s for each OC for the 118-bus system, but scales to 24min for a 3375-bus system [139], that may be still an acceptable time when the fault is foreseen. Switching to use commercial solvers such as Mosek and developing advancements in operational research may likely reduce this computational time further.

5.4.6 Discussion

The proposed control approach optimally balances the cost of normal operation considering preventive and corrective control and the transient stability. This approach has a promising performance from several viewpoints. From the stability viewpoint, the NN-LYAPUNOV CONTROL reduced the unstable OCs by up-to 99% and 51% when not using control and when using preventive WIND CURTAIL approach, respectively. Although the stability performance of the proposed approach was comparable to the optimised PREVENTIVE CONTROL, the NN-LYAPUNOV CONTROL outperformed the other approaches with regards to normal operating costs by a reduction of up-to 62% when compared against the WIND CURTAIL approach. This is a key step forward to secure power system operations as the costs for renewable curtailment are generally very high, for example renewable curtailment costed €372.7m in Germany in 2016 corresponding to the 43% of the total cost for congestion management [146]. Only the DT CONTROL resulted in slightly lower operating costs. However, the cost of experiencing unstable OCs is much higher than the difference in the operating costs between the proposed approach and the DT CONTROL one. Therefore, the proposed approach resulted in the best balance between stability and operating costs. Importantly, the proposed NN-LYAPUNOV CONTROL approach reduced carbon emissions by 60% through avoiding the curtailment of wind.

The proposed control approach for transient stability still has a few limitations that need to be considered. The learning times for the stability constraints are quite high for larger systems. However, this learning step is carried out offline (Fig. 5.2) well ahead of real-time operation with no limitations on computational resources, and may only be done once (or in regular time intervals). The unforeseen fault scenario was not investigated in this work as no preventive control would be available for unexpected faults, and this work focuses on a cost-optimal combination of preventive and corrective tools. Training offline a NN to instantly predict the cost-optimal solution of the SDP based optimisation in real-time could be a potential research direction to investigate [115] to improve the applicability of the proposed approach to real-time applications. Only a NN model was tested to learn the controller and the Lyapunov

function. However, the performance of such models were better discussed in [130] showing that it is worth using NNs for this task. Intuitively, larger NN models with larger numbers of layers and neurons would be needed for larger systems and this would increase the learning time [133]. Therefore, the trade-off between larger NN model's sizes and higher learning time should be also investigated. The performance of the proposed approach on future systems with higher shares of renewables than 40% should be also tested. Finally, relying on machine learning based control approached rather than investing in new grid infrastructure or curtailing wind power in advance has a risk that should be considered in the decision making process.

5.5 Conclusion

The need for novel operating methods to deal with the new dynamical phenomena was investigated showing that future grids can suffer from highly unstable operations. In a scenario where renewable sources and DERs make the generation more uncertain and the demand more flexible and CIGs scale the timescale of interest down to a few milliseconds, fast corrective control methods available in real-time are needed. In response, a new real-time operating approach that utilises the high flexibility that DERs (EVs in this work) have to offer was proposed. The key advancement of this approach is that corrective control can be used in normal operation, not only as backup strategies when preventive control fails. The vision of this work is that corrective control actively participates in maintaining the system stability. The proposed approach optimises the combination of preventive and corrective applications to reduce the operating costs and carbon emissions, and enhance the system stability. Concluding from the studies on the IEEE 9-bus system with high shares of renewable generation and integrated DERs and on the IEEE 39-bus and 118-bus systems: the approach outperformed existing control approaches in balancing operating costs and stability, resulting in reductions of up to 62% and 51% for the costs and the number of unstable OCs. Also in a larger system, the approach resulted in high stability performance. This paper recommends considering the proposed method in a real operating tool, then, as this paper showed, a significant step forward could be made toward

reducing wind curtailment and carbon emissions by up-to 60%.

Chapter 6

Security Assessment in Future Systems

Power system operation with a high share of renewables becomes more uncertain and has lower inertia to respond securely to disrupting dynamics [11,147]. Not only new renewable generation but also new flexible loads and new transmission devices have completely different dynamics than conventional synchronous generations with which the power system was designed. These new technologies are interfaced with converters which result in lower inertia in the system [148] and faster dynamics as the technology can interact and generate unexpected time couplings with existing system components as synchronous generators [6, 149]. The timescale of such relevant dynamics is similar to electromagnetic transients (and gets reduced from a few microseconds to several milliseconds as shown in Fig. 6.1). In response, operators can harden the power infrastructure or develop new software tools for the uncertain operation (much cheaper) to keep up the security of supply [147]. Unfortunately, most current real-time software tools are limited to only the assessment of the static security of a shortlist of system faults. With current software, the assessment of the dynamic security is infeasible in real-time as they require time-domain simulations which have long computational times as they rely on numerical integration [65, 72] (e.g., 56 s for the French system [24]).

Methods from ML show high potential for real-time probabilistic DSA [46, 150]. The idea is to carry out simulations offline, then train a ML model offline that can be used in real-

time for predicting the security instead of simulating it. This approach is a promising idea as ML predictions require nearly zero computational time, and hence very large numbers of possible operating conditions can be assessed in real-time for a very large number of possible fault scenarios. In the last decades, many researchers investigated this idea with different settings [9, 151]. Most works, however, propose the same sequence of steps from generating data over data pre-processing, selecting features toward training and validating the ML model. These works analysed the individual steps in detail, e.g. for data generation [34, 152], for data processing and feature selection [29, 79, 153], for model training [50, 154] and selection. Over decades this line of research has built upon each other improving successively previous conclusions in each of the steps. However, most of the works following this ML-based idea have carried out their studies on standard IEEE test systems, which typically assume high inertia in their dynamical models representing power systems from the past. Unfortunately, most works do not consider large amounts of renewable energy sources or adequate demand models and their flexibility or converter-interfaced generations or storage.

This chapter aims to investigate whether the conclusions in ML-based real-time DSA made in the past high-inertia (HI) systems are still valid for low-inertia (LI) systems. A few research has considered more LI-based assumptions in their proposed ML-based DSA concepts. For instance, [155] considered 30 % renewable energy sources, HVDC links, and forecasts for 2030 for generating operating conditions, but the static and small-signal analysis does not involve simulating the full dynamic model. [156] considers dynamic models for CIGs and assumes uncertainty in photovoltaic and wind power.

This work has three contributions that offer insights, for the first time, on whether the conclusions from ML-based DSA on HI systems can be transferred to LI systems. These insights are important to verify which findings from past research in ML-based DSA need to be revisited and how researchers may carry out ML-based research for LI DSA in the future. The first contribution summarises the changing assumptions of dynamics LI power systems that influence ML-based DSA, including the dynamical modelling of CIGs, transmission lines and loads.

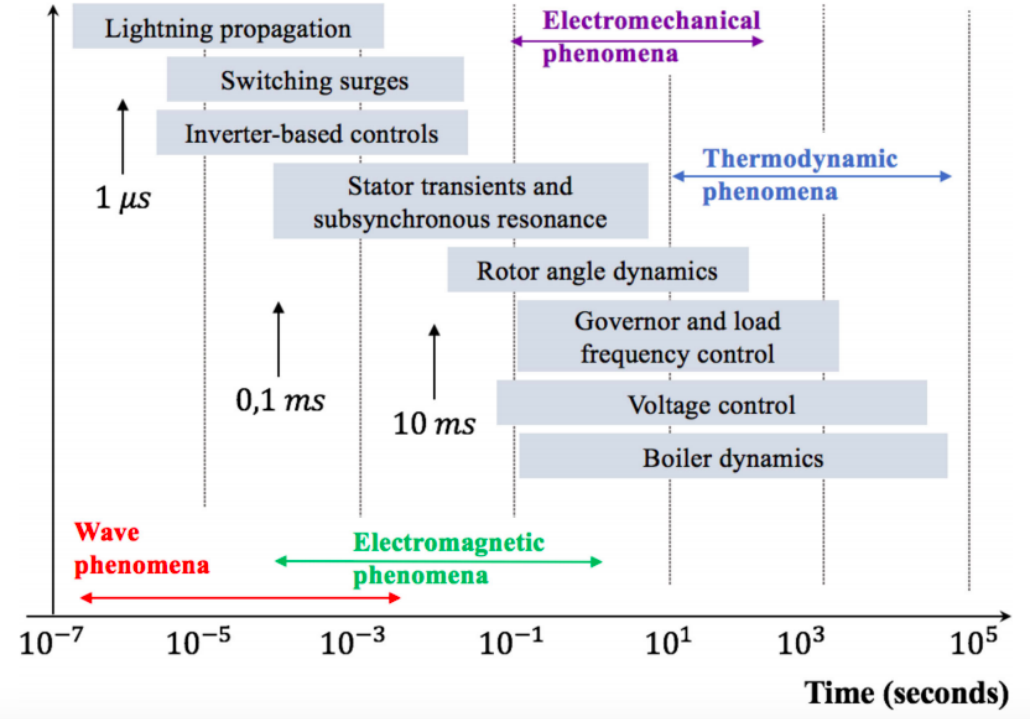

Figure 6.1: Timescales in Low Inertia systems [6].

The second contribution is analysing the impact of these assumptions on ML-based DSA. The impact of each assumption is analysed for each step individually in the ML-based DSA. The third contribution proposes a modified test system in the case study that corresponds to a LI system and can be used for ML-based research for DSA. Case studies on the original and the proposed modified versions of the IEEE test systems with 14 and 68 buses compare the performance of ML-based DSA in HI and LI systems considering transient stability.

6.1 Dynamic Security of Low Inertia Systems

This section describes the general ordinary differential equation (ODE) formulation of the dynamical model of LI systems [8]. In addition, this section introduces the new security definition and the new classes of security for such systems.

The nodal and generation sets are N and G, respectively, where G includes synchronous gen-

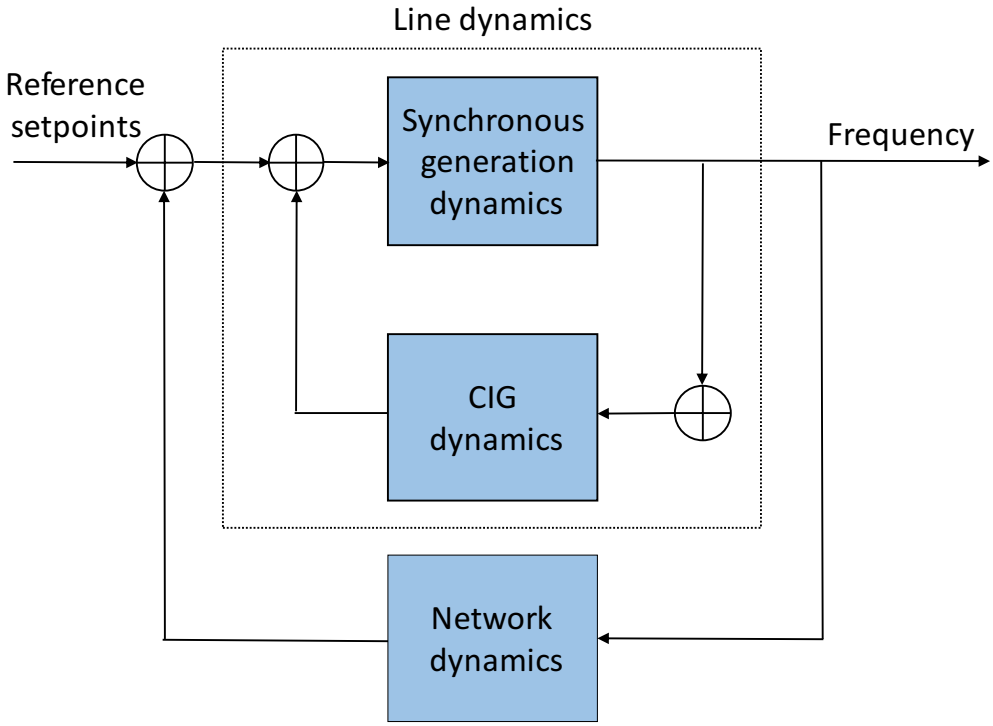

Figure 6.2: Block diagram of LI network (modified version of [7]).

erators S and converter-interfaced generators C, i.e. $G = S \cup C$. The set C is further classified into grid-forming converters $C_F \subset C$ and grid-following $C_f \subset C$ converters. The dynamical model of LI system (Fig. 6.2) can be then described as follows:

$$\dot{x} = f(x, u) \tag{6.1}$$

where

$$\begin{aligned} x =& (x^F_{c_1}, ..., x^F_{c_{c_F}}, x^f_{c_1}, ..., x^f_{c_{c_f}}, x_{g_1}, ..., x_{g_S}, x_n) \\ u =& (u^F_{c_1}, ..., u^F_{c_{c_F}}, u^f_{c_1}, ..., u^f_{c_{c_f}}, u_{g_1}, ..., u_{g_S}) \end{aligned} \tag{6.2}$$

are the system states and inputs, respectively. The models of the dynamical components to consider in LI systems, i.e. CIGs, synchronous generators, loads and transmission lines, are described in the following sections.

6.1.1 Dynamics of converter-interfaced generators

A unified power converter model that compromises both grid-forming and grid-following mode of operation within a single control structure is considered [8]. The two operation modes are shown in Fig. 6.3. More precisely, the main control architecture is a two-level voltage source converter: an outer system-level control provides a reference for the converter's terminal voltage that is tracked by a cascading device-level controller. In the implemented unified model, the single control design (both in terms of system- and device-level control) is modified in such a way that it fits the input-output characteristic of a grid-forming voltage source converter while still preserving its original dynamical properties. The general configuration of the such control scheme is illustrated in Fig. 6.4, with three main segments, i.e., system-level control, device-level control and device model. However, the Phase-Locked Loop (PLL) has no function in the grid-forming operation, as it is bypassed and the frequency setpoint ω_c is directly fed to the Active Power Controller. The device model consists of a DC-link capacitor, a lossless switching block, which transforms the the DC voltage v_{dc} into a three-phase AC voltage v_{sw}, and an output RLC filter (r_f, l_f, c_f).

The filter dynamics are described as follows:

$$\begin{aligned} \dot{i}_f =& \frac{w_b}{l_f}(v_{sw} - v_f) - \left(\frac{r_f}{l_f}\omega_b + j\omega_b\omega_r\right) i_f \\ \dot{v}_f =& \frac{\omega_b}{c_f}(i_f - i_g) - j\omega_b\omega_r v_f \end{aligned} \tag{6.3}$$

with ω_b and ω_r the base and reference frequency, i_g the transformer current that is injected into the grid, v_f and i_f the filter voltage and current. The transformer dynamics are:

$$\dot{i}_g = \frac{\omega_b}{l_t}(v_f - v_t) - \left(\frac{r_t}{l_t}\omega_b + j\omega_b\omega_r\right) i_g \tag{6.4}$$

with r_t, l_t and v_t the transformer's resistance and inductance and the voltage at the connection

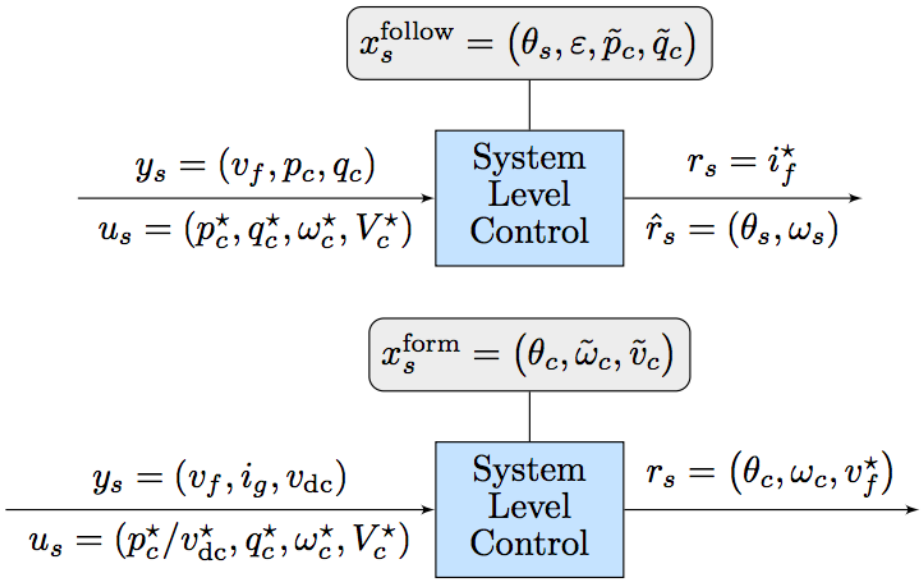

Figure 6.3: System-level control of grid-following operation mode (top) and grid-forming operation mode (bottom) [8].

terminal, respectively. The converter's AC voltage magnitude V_c and the instantaneous active and reactive power injections (p_c, q_c) can be directly obtained from the filter voltage v_f and the current i_g:

$$V_c = v_f, \quad p_c = v_f^T i_g, \quad q_c = v_f^T j^T i_g \tag{6.5}$$

The AC-side controller operates with the reference angle θ_c and velocity ω_c provided by the system level control which defines the reference voltage v_f^*. The reference current i_f^* is provided using a proportional-integral (PI) voltage controller:

$$\begin{aligned} \dot{\xi} =& v_f^* - v_f \\ i_f^* =& K_P^v(v_f^* - v_f) + K_I^v \xi + K_F^v i_g + j\omega_c c_f v_f \end{aligned} \tag{6.6}$$

where (K_P^v, K_P^i), (K_I^i, K_I^i) and (K_F^v, K_F^i) are the proportional, integral and feed-forward gains respectively, ξ is the integrator state and subscripts v and i denote the voltage and current controllers. By employing frequency and voltage control, R_c^ω and R_c^v, the outer control loop, described by the internal state variables $(\tilde{p}_c, \tilde{q}_c)$ for active and reactive power, regulates (p_c, q_c)

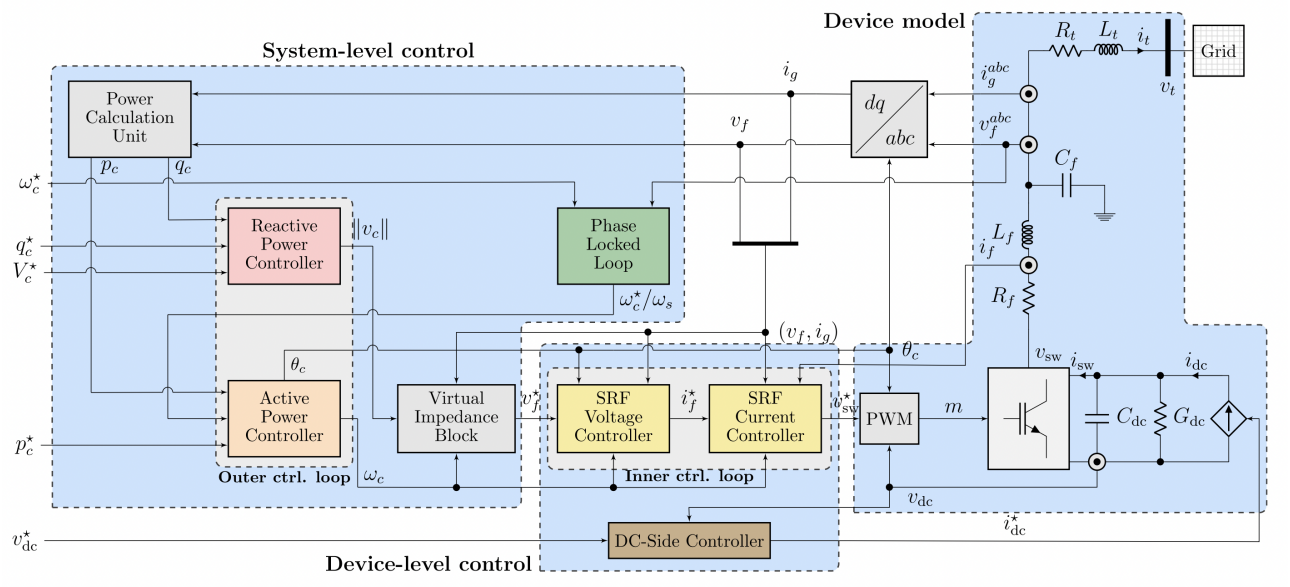

Figure 6.4: The control structure of the implemented unified converter model [8].

to the respective setpoints (p_c^*, q_c^*) as follows:

$$\begin{aligned} \dot{\tilde{p}}_c &= K_I(p_c^* - p_c - R_c^\omega(\omega_s - \omega_c^*)) \\ \dot{\tilde{q}}_c &= K_I(q_c^* - q_c - R_c^\omega(V_c - V_c^*)) \end{aligned} \tag{6.7}$$

with ω_c^* and V_c^* representing the frequency and voltage setpoints.

For the grid-following converters, the PLL dynamics should be also included to achieve the frequency synchronization. The PLL estimates the phase angle θ_s of the voltage v_f and the synchronous grid frequency ω_s. The PLL dynamics are described as follows:

$$\begin{aligned} \dot{\varepsilon} =& v_f \\ \omega_s =& \omega_0 + K_P^s v_f + K_I^s \epsilon \\ \dot{\theta}_s =& \omega_b \omega_s \end{aligned} \tag{6.8}$$

with $\omega_0 = 1$ p.u. being the nominal angular frequency, ε the integrator state, and (K_P^s, K_P^s) being the proportional and integral gains of the synchronization unit.

Given these dynamics, the following state and control inputs can be defined for CIGs:

$$\begin{aligned} x_c^F &= (i_f, v_f, i_g, \xi, \tilde{p_c}, \tilde{q_c}, \theta_c) \\ x_c^f &= (i_f, v_f, i_g, \xi, \varepsilon, \theta_s, \tilde{p_c}, \tilde{q_c}, \theta_c) \\ u_c^F &= (p_c^*, q_c^*, V_c^*, \omega_c^*, v_{dc}^*) \\ u_c^f &= (p_c^*, q_c^*, v_{dc}^*) \end{aligned} \tag{6.9}$$

6.1.2 Dynamics of synchronous generators

For a traditional two-pole synchronous generator, the internal dynamics are described by the swing equation:

$$M_g \dot{\omega}_r = \Delta p_e - D_g(\omega_r - \omega_0) \tag{6.10}$$

with M_g and D_g the inertia and damping constants and Δp_e the difference between mechanical and electrical power at the generator's output.

6.1.3 Load dynamics

Dynamic models of loads express the active and reactive powers as a function of voltage and time. Two main dynamical models for loads can be distinguished, the inductive model, and the exponential recovery load model [157]. These two models are mainly characterised by different time recovery ranges following disturbances. The inductive model is derived from the equivalent circuit of an induction motor with static and rotor resistances corresponding to a recovery time in the range of seconds (similar to electromagnetic transients). In this model, the active and reactive power, and hence the load current, are represented as functions of the past and present voltage magnitude and frequency of the load bus allowing to inject or withdraw current from the bus instantaneously when needed (current injection model [158]). Conversely, the exponential recovery load model is used to represent loads that slowly recover from a disturbance over large time periods (from several seconds to tens of minutes, similar to electromechanical transients).

6.1.4 Transmission line dynamics

Grid dynamics are typically neglected in synchronous generator-dominated power systems as these dynamics are much faster than those of the excitation and governor systems. With the inclusion of fast-acting CIGs, the line dynamics become relevant as they are of the same order of magnitude as the ones of converters (electromagnetic phenomena). The line dynamics can be represented using a conventional RL formulation:

$$\dot{i}_{jk} = \frac{\omega_b}{l_{jk}}(v_{nj} - v_{nk}) - \left(\frac{r_{jk}}{l_{jk}}\omega_b + j\omega_b\omega_c^*\right)i_{lj}, \forall j \in N, k \in K_j \tag{6.11}$$

with i_k and v_n being the nodal current and voltage, , r_{jk} and l_{jk} being the resistance and inductance of the line connecting $j \in N$ and $k \in K_j$ where $K_j \subset N$ is the subset of nodes adjacent to node j.

6.1.5 Security definition and classification

There is a need for new considerations on the classification and definition of the power system stability phenomena as a result of the increasing share of CIGs into bulk power systems [6]. The different dynamic behaviour of CIGs compared to synchronous generators behaviour leads to a mix of new types of transient phenomena that need to be constantly assessed. As long as the dynamic response to a fault affects only the CIG and does not cause the cascading instability of the main system, the conventional definition of power system security still applies. The conventional definition of security defines the ability of an electric power system to withstand sudden disturbances without major service interruptions in real-time. However, two new stability classes, i.e. converter-driven stability and resonance stability, need to be considered as the integration of power electronics devices scales the timescale of interest down to electromagnetic transients (Fig. 6.5). Resonance instability occurs when the magnitudes of voltage, current or torque exceed specific thresholds following the oscillations of periodic energy exchanges. These

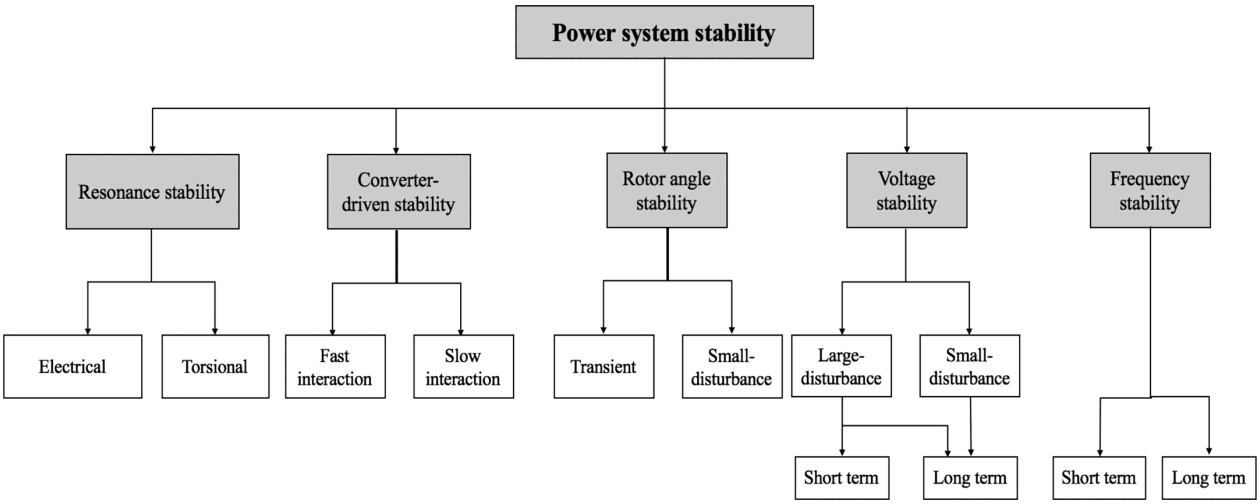

Figure 6.5: New classification of power system stability [6].

instabilities are caused by the resonance between series compensation and the mechanical torsional frequencies of the turbine-generator shaft (i.e. torsional resonance) or by the resonance between the series compensation and the electric characteristics of the generator (i.e. electrical resonance). Converter-driven instabilities involve a wide range of unstable frequency oscillations due to cross-couplings of CIG control loops with both the electromechanical dynamics of machines and the electromagnetic transients of the network.

The integration of CIGs does not affect the stability definitions for the existing stability classes, i.e. rotor angle, voltage and frequency stability. In terms of rotor angle stability, there is no consensus regarding the effects of CIGs on the small disturbance rotor angle stability. However, the fast control of such converters during and after the fault can significantly influence the transient rotor stability. Regarding voltage stability, the new dynamics of fast-acting load components, e.g. induction motors, controlled loads, HVDC links, may represent new potential causes of instabilities. Finally, although CIGs do not provide an inertial response as they are typically associated with renewable sources, they can provide fast primary frequency response, and this is crucial in LI systems in which fast controllers are needed to arrest frequency drops and keep the system's frequency within the acceptable range at all times.

6.2 ML-based Security Rules for LI Systems

This section investigates the impact of increasing the share of CIGs on conventional ML-based approaches for DSA. The four steps of these approaches are shown in Fig. 6.6, i.e. data generation, data pre-processing, model learning and model validation [9]. The following investigates these steps one by one with regards to past HI-based assumptions that may be outdated in LI systems.

6.2.1 Data generation

In conventional data-driven DSA workflows, database generation is the pivotal first step that lays a foundation for training a good quality model, and it is usually performed offline, starting from a list of possible OCs. The final database is a collection of these OCs that serve as pre-fault states, together with their respective post-fault security status after a contingency. In generating the pre-fault states, a combination of possible load levels and generator parameters are matched by solving the optimal power flow. In situations where historical data is insufficient, the loads are sampled from a multivariate normal distribution that assumes some correlation between the loads [34,152]. In the future, power systems with high renewable sources integration challenge this status quo with increased uncertainty. Firstly, there is a need to consider more "rare" cases that were previously ignored. Secondly, the search space that defines the OCs becomes much larger than that of current power systems. Current OCs of a power system are defined by a predefined set of variables, also known as features, i.e. currents, active and reactive power injections, load levels, bus voltages and phase angles. However, in future LI systems, the number of features to consider for each OC is expected to be significantly larger as each CIG will be, in turn, characterised by the terminal current/voltage. Therefore, generating possible operating conditions becomes more computationally intensive [155].

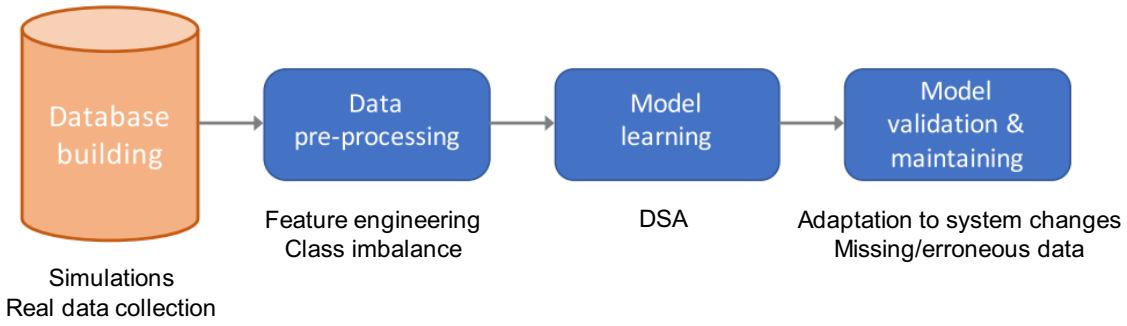

Figure 6.6: ML-based workflow for DSA [9].

6.2.2 Data pre-processing

Once the data is generated, the data is typically processed before being fed to a learning algorithm. This step becomes necessary in future LI systems to improve the predictive performance, reduce the training time and make the training data more interpretable. Many features are necessary to describe the OC of LI systems, and thus, the input features are expected to be highly redundant. Redundancy is one of the leading causes of low predictive performances in ML classifiers. Consequently, considering a large set of input features significantly increases the training time, making the training computationally infeasible in large systems even if done offline. Feature selection supports reducing the training time and improves the quality of predictions by selecting the features most relevant to the given classification target, and hence reducing the input space [155]. However, some of the most accurate feature selection approaches, i.e. wrapper methods and embedded methods, are very time-consuming. Therefore, their performance on the large input spaces of future operating conditions should be further investigated. Conversely, filter methods are another class of feature selection approaches based on statistical tests, which are fast even on large input spaces but generally result in low predictive performance [81].

6.2.3 Model learning

The high complexity of the dynamics of LI systems could be one of the leading causes of the low predictive performance of ML-based models for DSA. The dynamics of synchronous generators,

CIGs and transmission lines have significant differences in the time constants. Training a single classifier for such different timescales could result in very low accuracies. A classifier that works for short timescales, such as converters, may not work on much longer timescales such as turbines. Therefore, different models for each generator/converter insecurity pattern may be required to improve the predictive performance in a real-time setting. Clustering the different dynamics following the clearing of a fault and then training a different classifier for each cluster of dynamics could be a potential solution to overcome this challenge [156]. Since DTs have been proven to be suitable for real-time DSA purposes, multiple DTs may be used for these clusters of dynamics.

6.2.4 Model validation and real-time assessment

The increasing uncertainty in future power system operations resulting from the integration of renewable sources will significantly impact the quality of prediction of the ML-based models for DSA. In the traditional two-stages workflow of these approaches, the time distance between the offline and real-time stages compromises the performance of the offline trained classifier as the real-time operating conditions may be very different from those included in the training database. In such an uncertain scenario, training the classifier on immediate-future states or periodically updating the classifier may overcome the challenge of frequent changes in the distributions of the OCS [50,154,156]. However, these updates are computationally challenging as they are carried out in the real-time environment to minimise as much as possible the discrepancies between offline and real-time OCs. Therefore, the retraining or updating of the model with new data acquisition needs to be very fast [79].

6.3 Case Study

This section studies the impact of lowering the inertia on ML-based DSA. ML-based workflows for DSA were designed in HI systems and may not be valid anymore in LI systems. This case

study aims at demonstrating this impact on each step individually in the ML-workflow (i.e. data generation, data pre-processing, and model learning/validation). Subsequently, this study analyses the individual findings to provide insights on whether past research in ML-based DSA needs to be revisited or can be assumed as granted for LI systems and if the findings and assumptions also transfer to larger systems, such as to the IEEE 68-bus system.

The tested HI and LI systems had different assumptions. The tested HI systems were the original IEEE 14-bus [159] and IEEE 68-bus systems [160]. In these two HI systems, the dynamics were governed exclusively by synchronous generation and only static transmission lines. The tested LI systems were modifications of the IEEE 14 and 68-bus systems, where all the static transmission lines were replaced by dynamic lines, except the faulted line, and the 40 % (and 25 %) of thermal generation was replaced by CIGs in the IEEE-14 bus and IEEE-68 bus systems, respectively. The considered CIGs (the model in Section 6.1.1) were wind farms of similar power rating to the replaced synchronous generators. The wind farm model is a static generator connected to the network through the inverter model shown in Fig. 6.7 [10,161]. The control of power systems assumed an inner current control loop and an outer voltage control loop. The two systems were modelled in Julia 1.6.2 with the packages *PowerSystems.jl* [161], *PowerSimulationsDynamics.jl* [10]. Most of the defaults settings of the inverters for the LI system were used. The damping constant and the frequency droop gain of the outer controller were reduced by 99% to assume faster controls in the future and to study the impact of the faster controls with lower inertia on future system security.

The training data included $|\Omega| = 10,000$ pre-fault OCs and corresponding post-fault security labels for each of the HI and LI test systems. The pre-fault OCs considered different sampled setpoints for active and reactive power injections. The active load setpoints were sampled from a multivariate Gaussian distribution with a Pearsons' correlation coefficient $c = 0.75$. The method of inverse transformation was used to convert to a marginal Kumaraswamy distribution, where $a = 1.6$, $b = 2.8$ are distribution shape parameters. Then, the active load setpoints were scaled to be within $\pm 25\%$ and $\pm 10\%$ of the nominal values for the 14 and 68 bus system, respec-

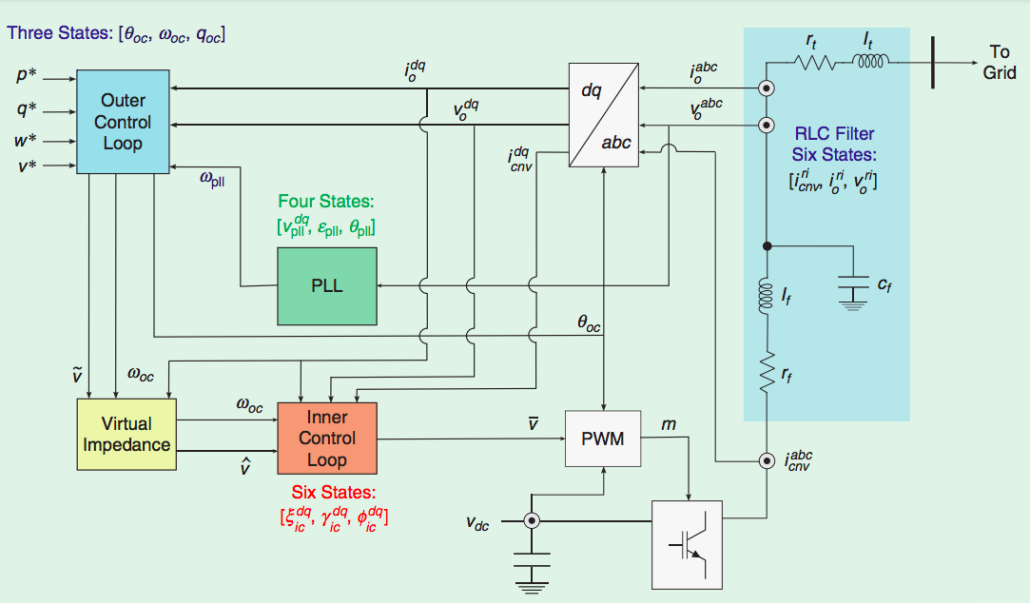

Figure 6.7: The block diagram of the grid-supporting inverter of wind farm models [10].

tively. The reactive powers followed the active powers proportionally as constant impedances were assumed. The AC power flow was then computed with the Newton-Raphson algorithm to obtain pre-fault OCs power generation. The solution to the AC power flow provides the full pre-fault variables X_i for each OC $i \in \Omega$, where the variables are active and reactive power generations, active and reactive power loads and voltages magnitudes and phase angles. The post-fault security label considered the transient security for 3 different three-phase faults k on line $2-3$ (Fault 1), $3-4$ (Fault 2), and $2-5$ (Fault 3) for the 14 bus system (Fig. 6.8) and for a fault on line $31-38$ for the 68 bus system (Fig. 4.5) with each a clearance time of 0.6 s. The fault on line $2-3$ for the 14 bus system was used for the following studies unless indicated otherwise. If within 10 s simulation time after the fault, all differences between each two phase angles of the generators were less than 180°, than the OC i was considered transient secure $Y_{i,k} = 1$, otherwise insecure $Y_{i,k} = 0$ and with that, the security label was computed. These transient simulations were modelled with the same packages *PowerSystems.jl* and *PowerSimulationsDynamics.jl*. The simulations were solved with the IDA package from Sundials solvers [162]. These simulations were performed twice, evaluating the security labels for the two systems, the HI and LI systems. The pre-fault OC and post-fault labels built the different databases for evaluating the impact on the ML workflow for DSA. All simulations were carried

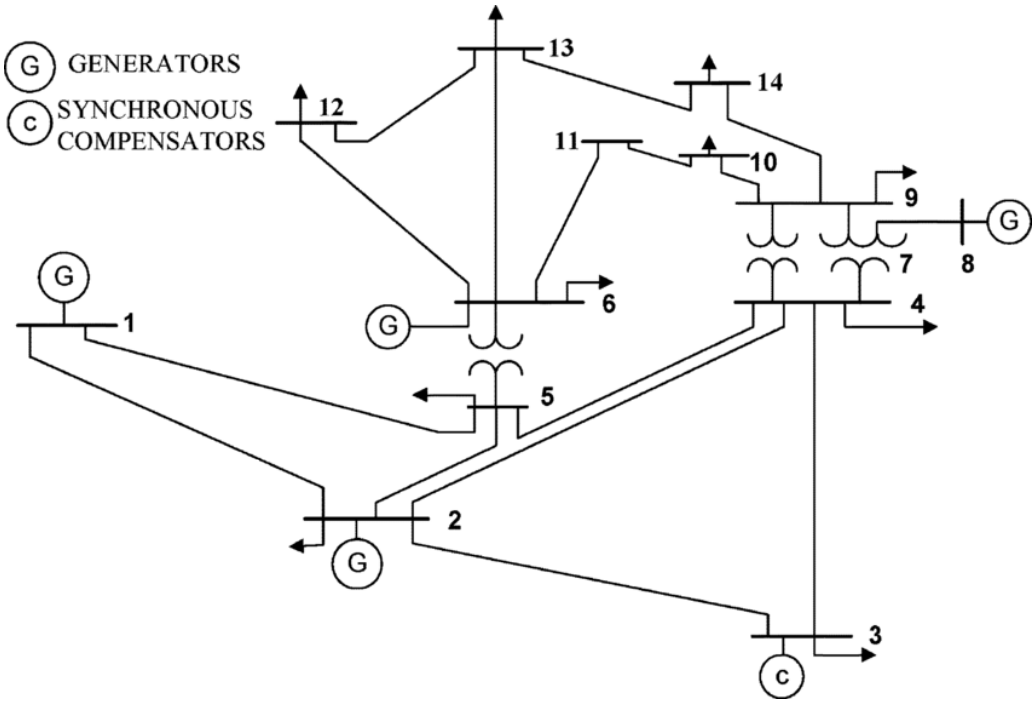

Figure 6.8: The IEEE 14-bus system with three-phase faults on lines $2-3$, $2-5$ and $3-4$.

on a standard machine with six cores and 16GB RAM.

The ML-methods studied included different FS methods as MRMR, CFS, JMI, SVM-RFE, SFS and different ML models, as DTs, SVMs, XGBOOST, feed-forward ANNs. A linear kernel for the SVM, 50 estimators for the XGBoost and 3 layers with 30, 15 and 5 neurons for the ANN, were used. Across the studies, 5 different combinations of training/testing sets were computed for each classification model. DTs were used to measure the classification performance unless indicated otherwise. The F1-score was most of the time used to assess classification performance as this score penalises the false negative and false positive. The false negatives have the highest cost in power system DSA (as false negatives can result in partial power blackouts in the worst case). For the fault on line $2-3$, $9,000$ OCs were used for training with a split of 70%/30% for training/testing and the remaining $1,000$ OCs for validation. For the other faults, 600 OCs were used for training with a split of 70%/30% for training/testing and the remaining 100 OCs for validation.

6.3.1 Differences in generating data for HI and for LI systems

This study focuses on the differences between an HI and an LI system when generating ML training data. The study focuses on 3 parts as a good training database for ML-based DSA (i) avoids redundancy in the data e.g., balanced shares of secure/insecure data, (ii) is easily separable (e.g., by a binary classifier), and (iii) is small while containing high levels of information.

The analysis (i) investigated the redundancy in the database. The share of secure/insecure OCs was a metric to investigate this redundancy, e.g., if all OCs in the database are secure, an ML model can not learn a security rule for classifications. Fig. 6.9 shows the share of secure OCs $\frac{N^+}{N^+ + N^-}$ in the HI and LI system for the same feature $\frac{P_1}{Q_1}$ active/reactive power generator level. Each point in the figure represents an average of secure OCs over 200 OCs with similar $\frac{P_i}{Q_i}$ values. The share of secure OCs decreased with increasing power generation in the two systems. However, the decrease in the LI system was much steeper and quickly converges to almost no secure OCs. This analysis showed that LI systems must be improved to keep the security as high as in HI systems (the tested LI system was not optimised for security), and the steep decrease showed the importance of optimising power generations to ensure the system's security.

The analysis (ii) investigated the separability of the data. Fig. 6.10 shows the distributions of secure and insecure OCs for the two systems, HI and LI, according to two generator power levels (active and reactive). A key finding was that the data is more separable in the LI system than in the HI system, e.g., even a linear function could separate secure from insecure OCs in the LI system, however, the security boundary (that is the function separating the secure from insecure OCs) seemed highly non-linear in the HI system (at least in these two features). The high non-linearity of the input data in the HI system suggested that in such a system i) more training features than in the LI system would be needed to obtain similar accuracy performance, ii) the dependence between system security and changes in power generations is not as strong as in LI systems. In LI systems, small changes in the generation make the system highly secure/insecure and this dependence results in a clearly defined boundary. Fig. 6.10 is

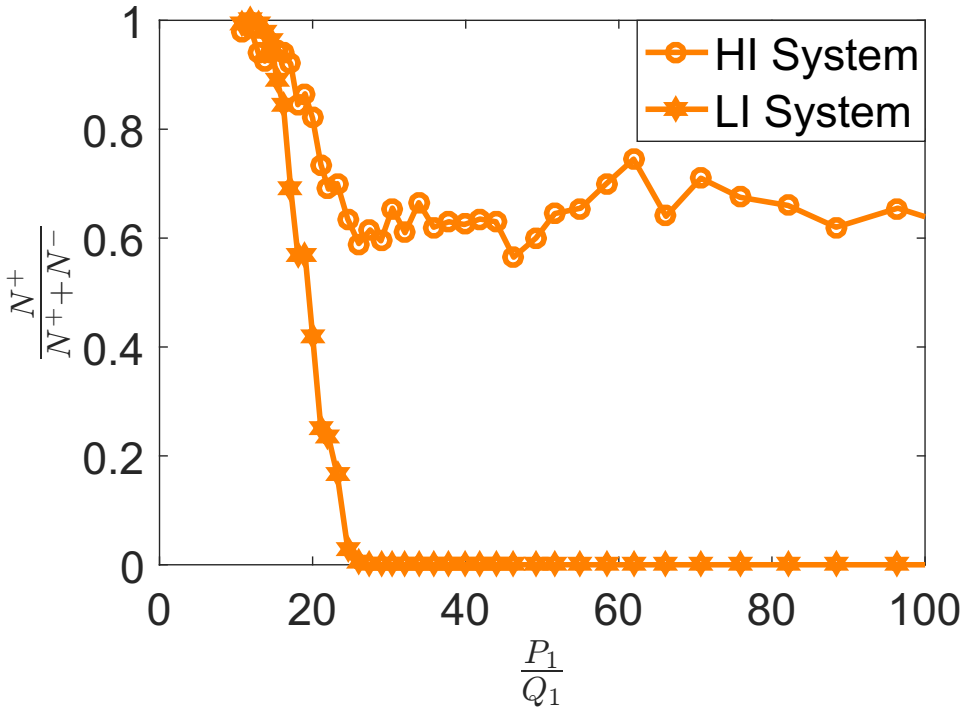

Figure 6.9: Share of secure OCs with increasing load levels for HI and LI systems.

zoomed in to better show the separating boundary between the secure/insecure OCs. When zooming out, some rare OCs (clearly separated from the others) can be noticed in both systems. Their number was reduced by 40% in the LI system compared to the HI system.

The analysis (iii) investigated the information per data (efficiency of database) in two methods: the first method was to train classifiers for the two systems with varying training database sizes figuring out the classifiers with the same predictive performance, and the second method was to analyse the information content of the two databases for LI and HI systems. In the first method, DTs were trained for the two systems for different sizes of training database Ω then their DT F1-score performances were analysed. Fig. 6.11 shows that the F1-score performance increases in the two systems. However, the F1-score was significantly higher in the LI system, which is related to the easier separability as pointed out in analysis (ii) and Fig. 6.10. The low DT performance in the HI system is related to the small test system size as in small systems, faults may result in very complex transients, which may be difficult to predict. These low numbers around $\sim$ 80% were also found for the IEEE-68 bus system

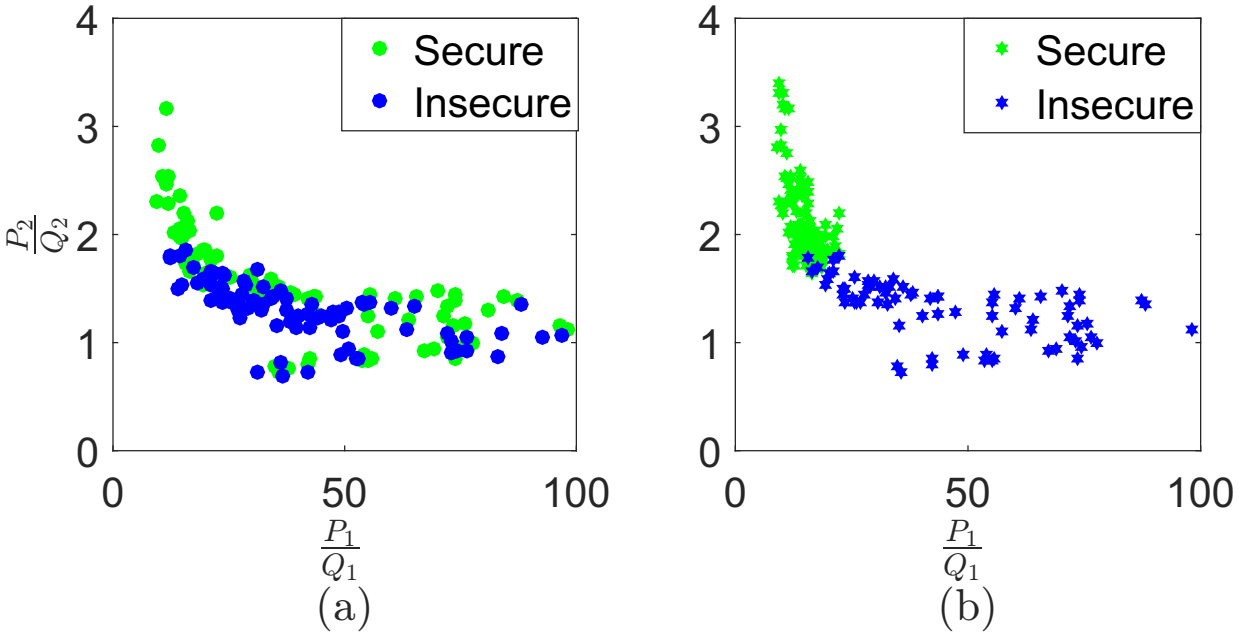

Figure 6.10: Secure and insecure OCs are less separable in (a) HI systems than in (b) LI systems.

Section 6.3.4, and in the literature, e.g., [79]. Additionally, the LI system classifier was very robust against reductions of the training database varying only around 2% compared to 20% in the HI system. In the second method, the mutual information I was computed to study the information content of the two training databases with the same size for HI and LI systems. The mutual information I between feature subsets and target Y and a summary of the two analyses, (i-ii) are in Table 6.1. The higher information in features X_i about the target Y resulted in higher classification performance in the LI system. The accuracy and F1-score were higher in LI than in HI systems, respectively by 23%, and 39%. The findings of high classification performance, the improved separability of security/insecurity in LI systems seem promising for a key role of ML for future real-time DSA.

6.3.2 Selecting features in LI systems

This study focuses on the differences between a HI and a LI system when pre-processing data to select features, the second step in ML-based workflows for DSA. Effective pre-processing of

Table 6.1: Performance metrics of training database.

	$\frac{N^+}{N^+ + N^-}$	$I(X;Y)$	ACCURACY	F1-SCORE
HI System	0.7	2.1	75.3%	59.2%
LI System	0.2	10.4	97.8%	98.5%

data reduces redundancy, enhances classification performance and reduces the training computational times. The past research on HI systems showed two findings: (i) the active and reactive power levels of generators are effective features for DSA, and (ii) typically, the wrapper and embedded feature selection methods have the highest performance for offline applications, and the filter feature selection methods are the fastest and hence best for real-time applications. The following two analyses investigate whether the same findings hold for LI systems.

The analysis (i) investigates the importance of the feature subsets power generations, power loads, voltage magnitudes and phase angles. This analysis used two metrics, the F1-score of a trained DT and the mutual information $I(X;Y)$. The analysis restricts the training data X to the corresponding feature subset. In other words, the subset providing the highest F1-score and mutual information are the best features. This analysis was undertaken for the LI and HI systems, and the results in Table 6.2 show that the subset of power generations was the best feature subset in both systems. E.g., in the HI system, the F1-score is 20 % higher when choosing power generation as features than when choosing other features. This finding is aligned with Fig. 6.10 which showed power generation as a strong indicator for system security in LI systems. However, in LI systems, voltage phase angles show to provide a significant higher mutual information, and this is also confirmed in the literature as voltage phase angles have conditional correlation properties allowing to capture the system's topology, and hence more relevant information for the security [89].

The analysis (ii) investigates whether the same FS methods are the best in LI systems as in HI systems, where the different FS methods select the best features out of all available features. For each different FS method, a DT was trained on the selected features. Then, the DT classification performance was compared to identify the best FS method for the LI system and

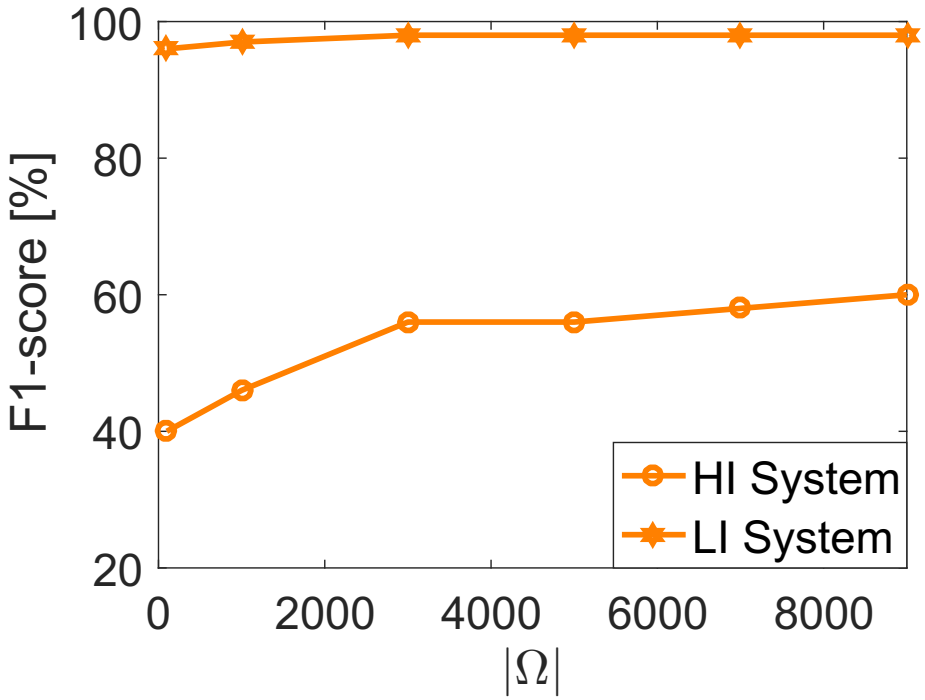

Figure 6.11: Prediction performance for different training database sizes.

for the HI system. The compared FS methods were MRMR, CFS JMI as filter method, SVM-RFE as embedded methods, and SFS as wrapper methods. The results in Table 6.3 show that the embedded method SVM-RFE and the wrapper method SFS had the highest F1-scores in the LI and HI systems, respectively. The analysis showed that SFS selected power generation and voltage phase angles feature as the mutual information values and F1-scores were high (Table 6.2). Similarly, SVM-RFE selected most features from voltage phase angles according to the high mutual information value as in Table 6.2. However, SVM-RFE also selected features from power generation and loads as they improved the accuracy.

These two analyses showed that the conclusions made in the past that corresponded to data pre-processing in ML-based workflows remain valid in the future LI systems: wrapper and embedded approaches typically result in the highest classification performance.

Table 6.2: F1-score (of trained DTs) and mutual information I for different feature subsets.

		Power Generation	Power Loads	Voltage Angles	Voltage Magnitudes
HI system	F1-score	64.9%	46.3%	43.2%	42.6%
	$I(X;Y)$	0.5	0.5	0.6	0.5
LI system	F1-score	98.5%	97.2%	93.4%	95.8%
	$I(X;Y)$	1.7	2.4	4.7	1.6

Table 6.3: F1-scores of different FS methods for training DTs.

	CFS	MRMR	JMI	SVM-RFE	SFS
HI system	52.1%	60.4%	60.5%	62.0%	64.5%
LI system	98.2%	98.2%	98.5%	98.8%	98.7%

6.3.3 Training ML models for LI and HI systems

This study aims at investigating whether the classification models recommended for HI systems will still be high-performing in future LI systems. The best ML model and the criteria to select the best model highly depend on the type of problem and application. In the following criteria finding a good trade-off between interpretability of the model and accuracy (F1-score) of the model was studied as an example [39]. The models studied were DTs, SVM, XGBoost and feed-forward ANN and the training considered all the 60 available input features. Cross-validation was used to tune the hyper-parameters of the models.

The results in Table 6.4 show that ANN outperforms the other models in LI and HI systems in terms of the F1-score. E.g., ANN outperforms SVM by around 8 % and 2 %, in HI and LI systems, respectively. The performance however of all models improved from HI to LI systems which were caused by the higher separability in LI systems as discussed earlier with Fig. 6.10. However, the values in the HI system are generally lower than what has been reported in the literature which is caused by the hard DSA classification problem at hand. The system security is represented by a highly non-linear security boundary and it seemed challenging to predicting for these selected faults with little training data and high load variability. Interestingly, the DTs showed high F1-score accuracy in LI systems of up to 98.5%. Additionally, DTs were promising regarding the selecting criteria of the trade-off to interpretability, as DTs offer higher levels of

Table 6.4: F1-score with different classification models.

	DT	SVM	XGBOOST	ANN
HI System	59.2%	55.4%	62.9%	63.1%
LI System	98.5%	97.2%	98.5%	99.1%

interpretability over ANN models. Other criteria can be selected in different applications, e.g. accuracy performance is prioritised in energy and load forecasting applications where ANN can be preferable [78, 163].

6.3.4 Larger LI systems and other faults

This study aims to generalise the previous studies to other faults (analysis i) and larger HI and LI systems (analysis ii). Analysis (i) investigated the 14-bus system with faults on two additional lines $3-4$ (Fault 2) and $2-5$ (Fault 3), and analysis (ii) investigated a fault on the IEEE 68-bus system.

The analysis (i) investigated whether the same conclusions for data generation, feature selection and model training (Secs. 6.3.1–6.3.3) were drawn for other faults in the 14-bus system. The results in Table 6.5 show that the same conclusions can indeed be found for these faults. The LI system experienced more insecure operating conditions than the HI system as a result of the reduced inertia. The training features X_i had an information content about the target Y 10 times higher in the LI system resulting in an enhanced predictive performance with improvements of $\sim 20\%$ in the accuracy and $\sim 45\%$ in the F1-score. Again, the higher accuracy performance in the LI System resulted from the linear separating boundaries making the classification task much easier in the LI system than in the HI system.

The analysis (ii) investigated whether the same conclusions for data generation, feature selection and model training count also for a larger system, e.g. the IEEE 68-bus system. The results in Table 6.6 show that, indeed, most of the conclusions are the same. The share of secure operating conditions decreased by around 80% from the HI to the LI system. One difference is that the information content of the database was similar in the LI 68-bus system and the HI

Table 6.5: Performance metrics for different faults.

		$\frac{N^+}{N^+ + N^-}$	$I(X;Y)$	Accuracy	F1-score
HI system	Fault 2	0.7	2.1	70.2%	47.5%
	Fault 3	0.7	2.1	72.5%	49.5%
LI system	Fault 2	0.6	11.9	93%	94.1%
	Fault 3	0.5	14.3	92%	94.2%

Table 6.6: Performance metrics for a larger system.

	$\frac{N^+}{N^+ + N^-}$	$I(X;Y)$	Accuracy	F1-score
HI System	0.9	8.2	91.5%	77.1%
LI System	0.1	8.5	89.6%	94.5%

68-bus system, and also the DT accuracy was similar. However, the F1-score performance was around $\sim 20\%$ higher in HI than in LI systems.

These two analyses showed that most conclusions drawn in Section 6.3 for the overall workflow (Secs. 6.3.1–6.3.3) generalise to different faults and on a large system.

6.3.5 Discussion

The key finding of this work is that using ML for real-time DSA becomes even more promising in LI systems as it was already in HI systems. It is known that LI systems result in lower system security (when not optimised), but this work's key insight is that the prediction of security (and learning of it by ML) becomes easier in LI systems. Additional findings of this work are the following differences (and similarities) when learning a ML model for a HI and a LI system:

1. When generating data, more operating conditions in LI systems may be insecure as the inertia is lower and dynamic couplings may occur.

2. The wrapper and embedded feature selection methods are perform well in both LI and HI systems. Power levels are valuable features to assess future system security as lowered inertia results in a stronger dependence of system security to changes in power generations.

3. ML can easier learn the classification boundary in LI systems than in HI systems which results in strong improvements of security predictions. The reason for this improvement is that LI systems may have clear (e.g. linear) separating boundaries between secure/insecure operating conditions.

4. Wrapper methods for feature selection (such as SFS) result in the highest predictive performance, above all in HI systems. Therefore, they can be preferred when there are no limitations on computational resources, for example in the offline stage. Otherwise, a filter method such as JMI may be preferred.

5. ANNs result in the highest predictive performance in both LI and HI systems. However, DTs have a higher performance in LI than in HI systems, and hence are promising as DTs are more interpretable than ANNs.

This chapter and its case study were designed for a typical ML-based workflow for DSA that corresponds to the majority of ML for DSA approaches from the literature. However, there may be other ML-based DSA approaches that use completely different workflows. For these workflows, the conclusions of this work are likely to correspond to as well as they relate to the underlying changes of the stability phenomena and how this changes the training data, however, this can not directly be concluded from this work.

A limitation of this work is that the model and design of the LI system were not optimised. However, all the studies were conducted under the same conditions for both the HI and LI system to make the comparative analysis fair. The low security of the LI system with increasing levels of power generation highlights the need to add virtual inertia to future systems to enhance the security. Although the virtual inertia can improve the security levels in LI systems, it is not expected to impact the performance of ML-based DSA (e.g. data separability) as the virtual inertia is generally provided for very short time intervals [164]. In contrast to some commercial software packages such as DSA Tools or PSS/E, the CIG model used in this work cannot differentiate between the types of generation behind or consider the weather

conditions [165].Advanced frequency modes and whether these influence the dynamics were not studied, and this work is limited to the dynamic model used. The analysis of advanced dynamic models and their impact on the overall system is an interesting analysis to conduct in the context of ML-based DSA aiming at predicting system security. This work, however, provided only the first step toward more sophisticated studies on ML-based real-time DSA for lower inertia power systems.

6.4 Conclusion

This chapter aims at investigating whether findings from past research on Machine Learning approaches for DSA are still valid when moving from high to low inertia systems. ML is promising for real-time DSA as it can predict instantaneously the security, which becomes even more important in low inertia systems with shorter dynamic timescales. This chapter describes the full dynamical model of future low inertia systems, including the dynamics of CIGs, loads, lines and synchronous generators. Then, the effect of these new dynamics on the design of ML-based approaches for real-time DSA are investigated. This design was successively optimised over the last decades but unfortunately only studied on high inertia systems. The studies on LI and HI IEEE 14 and 68-bus systems demonstrated the increased promising role of ML approaches in future ML-based DSA for LI systems. The predictive accuracies of these approaches improved by up to 40% and 20% for the 14 and 68 system, respectively. ANNs and DTs show accuracies close to $\sim$ 99%. In the future, LI systems must be optimised for higher security, and regular updating of test systems is needed to develop future real-time DSA methods and tools.

Chapter 7

Conclusion

7.1 Summary of Thesis Achievements

Power systems are undergoing a fundamental transition from the conventional fossil-fuel-based paradigm to a decentralised and digitalised paradigm based on the massive integration of renewable and distributed energy sources and advanced communication and information technologies [13]. This transition is massively changing the way energy is transmitted, distributed and managed. In the past, the system inertia from fossil-fuel generators provided imminent flexibility to stabilise operations following disturbances. As DERs and RES are connected through power inverters, they do not provide inertia; instead, they introduce very fast coupling dynamics, which are rather challenging for the system's security management. The current strategy to address these emerging challenges follows the centralised paradigms from the past. To consider the new dynamical phenomena, the centralised strategy is to operate with very large static margins and invest in redundant grid infrastructure.

Digitalisation is one of the key drivers for a more efficient and decentralised management of the energy system. The advances in communication technologies and measurement devices render available a large amount of operational data and enable the centralisation of such data

storage and processing. This data centralisation enables the use of novel data analytics and more advanced processing techniques, such as AI, to fully exploit the new flexibility provided by the DERs at the local distribution level. Maximising the grid impact from this flexibility can enhance the system's security and significantly reduce the operational costs. However, at the same time, relying on data-based approaches increases the risk of cyber-attacks or inaccurate decisions, and robust counter-measures are therefore needed as an integral aspect of digitalisation efforts. Although a few digitalised approaches have reached the development stage, the real-time application of such approaches for power system's security assessment and control is still very challenging. The main challenges relate to the low interpretability of such approaches and their low robustness against changes in data distribution or system's topology [9, 11]. On one hand, operators that have relied on manual/visual situational awareness for decades, have now to start trusting these automated processes. On the other hand, topological changes are becoming very frequent in modern operations posing new threats to the real-time adoption of such automated algorithms. In this thesis, we propose novel operating data-driven approaches for the assessment and control of power system's security, which are more robust against topological changes as they combine the learning from data with the system physics that are always known in the domain of power systems. At the same time, such an intersection of physical modelling and knowledge extracted from data makes the proposed "hybrid" models much more interpretable, and therefore much more reliable from operators' point of view. Hence, informing the ML approaches with the known physics of the network is highly promising as it increases their generalisation ability, robustness and interpretability simultaneously [166].

The known physics can either be induced or learned when training the ML models. For models predicting the security, the known physics relate to the swing equations and other dynamics which can be learned through regularising the loss function, such as in Physics-Informed NNs (PINNs) [167]. For models predicting control actions, the known physics relate to the power flow equations (e.g., Kirchhoffs law) which can verify the feasibility of the control strategies. In this thesis, we first informed the ML model for security assessment with the network topology and then we informed deep learning models with system stability to control the post-fault

dynamics. As follows, we summarise these challenges, how we addressed them and conclude this thesis by stating future directions of this research. We repeat here the research objectives stated at the beginning of the work with the aim to highlight in greater detail the achieved research contributions.

1. *How can we inform ML-based DSA for power system control with the knowledge of network configuration in order to improve its robustness against increasingly frequent topological changes?*

In Chapters 3-4, we physically informed the ML-based approaches for DSA by inducing the physical knowledge of the network topology in the correlation structure between the input features and the dynamic security. This enables the selection of those features that are more relevant to classification even when a topology change occurs. The same features were then used to define a new metric for the detection of high-impact topology changes, as a result of which the DSA classifier needs to be retrained. The key novelty of such a metric is that no transient simulations are required for the computation and hence it can be computed in real-time. The use of such a metric enables the training of robust classifiers against topological changes, marking a fundamental step forward in the real-time adoption of ML-models for DSA.

2. *How can we use corrective control to actively enhance the system's dynamic security and reduce the operating costs?*

In Chapter 5, we showed how the optimal use of corrective control applications in combination with preventive strategies (not only as a backup strategy to when equipment fails) enhances the network utilisation and reduces the operating costs while maintaining adequate security levels. The proposed control approach is an example of learning the known physics for the control of dynamics. First, we informed a NN with the stability theory to learn the Lyapunov function of the post-fault system. The corresponding region of attraction of the learned function was then used to transform the differential stability constraints into algebraic constraints to include in the ACOPF, which can be then solved very quickly to compute preventive and corrective control actions in real-time.

3. *Can we assume that state-of-the-art ML-based DSA for power system control will still perform well in future low inertia systems?*

In Chapter 6, we questioned whether the underlying assumptions and conclusions of ML-based approaches for DSA will still be valid in future low inertia systems. We studied and analysed each sequential step of the most typical ML-based approach for DSA, i.e generation of the training database, data pre-processing, model training and validation, for the changed assumptions in a dynamic low inertia system. The key findings of such a study confirm that using ML makes significantly more sense in future low inertia systems.

7.2 Future Work

In the above discussion, we describe a few potential directions that we identify based on our research and that may further encourage the applicability of data-driven based approaches to the real-time assessment and control of power system security.

7.2.1 Data generation in future power systems

Considering the different topological variations in the training database (or changes in the data distributions more generally) by sampling many different system topologies or operating conditions is not feasible in the future, as a very flexible topological configuration is expected [34, 41]. Simultaneously, as the number of variables describing the physical system increases with the system's size, the sampling space becomes more and more multidimensional, but at the same time the feasibility of such a space is constrained by the physical equations.

Again, the future research direction for effective generation of training databases lies in the intersection between physical modelling and data statistics. One recent contribution in this direction is proposed in [152], where an outer approximation is used to convexify the original, non-convex space, and subsequently sample from this convex space by rejecting samples and

target sampling closer to the security boundary. Similarly, authors in [168] propose a novel split-based sequential sampling approach to generate diverse operation scenarios for training ML models. Then, a new volume-based coverage metric is defined to quantify the quality of samplers across the multidimensional space that are within the physical network limits.

7.2.2 Robust Machine Learning models for DSA

The state-of-the-art workflow for ML-based DSA has an offline stage where the model is trained, and a real-time stage where the ML model is used to assess the security of the next operating conditions [24]. Similarly to the data generation challenge, the offline trained model could be outdated in real-time operation, as in the time gap between the offline and online stages the system's topology or the operating scenario (and the data distribution) may often change. Such ML models exploit well the repetitive nature of DSA and learning from data, but most methods are incapable to consider the information about the power system physics and the dynamical system described by the ODEs.

A number of system parameters are assumed to be fixed when using ML for security classification. However these parameters, e.g. fault clearance, can affect the security of the system, therefore the sensitivity to such parameters should be taken into account when training the ML models. Similarly, the security rules that are trained only for one topology or for a single contingency may result in inaccurate assessment when used for a different topology/contingency. In this thesis, the robustness of ML models for DSA against different topological changes was investigated. However, only line disconnections were considered as topological changes and only three-phase line faults were simulated as contingencies. Due to the increasing operational uncertainty, more complex topological changes can nowadays be performed as remedial actions and new and more disruptive events can impact the grid, e.g. lost of generation or cascading failure. Therefore, a key future direction is to enhance the machine learning workflow by considering the system physics in the training in order to make it more robust against these new changes in the system.

Recent works in this research direction focus on novel "hybrid" models in which the known physical knowledge is considered into the training loss that describes the differences between learned and known physical equations. One of these hybrid methods is PINN, where the NN predicts the state variables from inputs and then the first-order derivatives of the predicted variables are evaluated using Automatic Differentiation (AD). These derivatives allows the inclusion of the physical regularisation in the training loss function. For instance, PINNs are used to predict the dynamics of a power system subjected to a fault in [167] and to estimate the non-linear parameters of power system dynamics in [169]. However, PINNs do not yet scale to larger power systems and more work is needed to develop them for real-time DSA applicability.

A different approach to incorporate physical information into standard ML models is proposed in [170], where a Graph Neural Network (GNN) represents the state of a physical system with particles, expressed as nodes in a graph, and computes dynamics via learned message-passing. This graph-network simulator could be used to predict the dynamics of a power system subjected to a fault, such as for PINNs in [167], and result in very promising performance in terms of robustness against topological changes as the graph indirectly captures the physical interconnectivity of the system.

7.2.3 Advanced data-driven control for system security

Other key research directions worthy of investigation lie in the consideration of dynamics within data-driven control approaches and in the full automation of such control approaches. The overarching vision for these control approaches is an integrated system of distributed components communicating with each other through an advanced communication network and new digital technologies and controllable in real-time without the need of operator's inputs [15]. Current practice in the control room is that operators decide the control actions to take based on their expertise. Operators are very reluctant to adopt automated processes as the task at hand is very complex and could result in severe consequences if the wrong control action is taken.

The future research direction lies again in how to make such automated control approaches more interpretable and more reliable, so that operators can trust them, and finally adopt them in real-time operations. Authors in [37] implement interpretable ML security rules for dynamic control by learning the safety margins from the classification regions of a decision tree. A similar approach is proposed in [171], where a NN rather than a DT is used for the same purpose. However, in such works authors also highlight the challenges of utilising the security rules for control purposes: an operating condition that is classified as safe according to the security rules can be found to be unsafe when the contingency is simulated. In this thesis, we propose a novel approach to capture the security conditions of the post-fault system within the pre-fault ACOPF using the Lyapunov stability theory. However, such an approach is not suitable for scenarios where the fault is unforeseen as no preventive control would be available. Training offline a NN to instantly predict the cost-optimal solution of the SDP based optimisation could be an interesting research direction to investigate, to improve the applicability of the proposed approach to real-time applications [115] where only corrective control would be available.

In the context of corrective control for dynamic stability, more recent approaches focus on the application of reinforcement learning for security control, such as in [172], or as corrective control for voltage stability as in [173]. Authors in [174] investigate how to inform the RL agent with the network topology using a graph convolutional network in order to make the RL-based control approach for voltage stability more robust against topological changes. Another promising research direction to investigate is to use topological changes for corrective control by considering the transient stability constraints within an optimal transmission switching problem [58, 59]. In this thesis, the additional storage capacity from EVs with V2G technology was used for corrective control, demonstrating the economic and operating value that such technology could provide. In a similar way, EVs can be utilised for different balancing and ancillary services, such as frequency response [175, 176]. The use of data analytics for the real-time provision of ancillary services is also worth investigating as it may result in significant cost reductions in future low inertia systems.

Bibliography

[1] P. Kundur, "Power system stability," *Power system stability and control*, pp. 7–1, 2007.

[2] J. L. Cremer, "Probabilistic Dynamic Security Assessment for Power System Control," Ph.D. dissertation, Imperial College London, 2020.

[3] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science & Business Media, 2006.

[4] C. Liu, F. Tang, and C. Leth Bak, "An Accurate Online Dynamic Security Assessment Scheme Based on Random Forest," *Energies*, vol. 11, no. 7, 2018.

[5] S. Izumi, H. Somekawa, X. Xin, and T. Yamasaki, "Estimation of regions of attraction of power systems by using sum of squares programming," *Electrical Engineering*, vol. 100, no. 4, pp. 2205–2216, 2018.

[6] N. Hatziargyriou, J. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal, P. Pourbeik, J. Sanchez-Gasca, A. Stankovic, T. Van Cutsem, V. Vittal, and C. Vournas, "Definition and classification of power system stability – revisited amp; extended," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3271–3281, 2021.

[7] Y. Jiang, R. Pates, and E. Mallada, "Dynamic Droop Control in Low-Inertia Power Systems," *IEEE Transactions on Automatic Control*, 2020.

[8] U. Markovic, O. Stanojev, P. Aristidou, E. Vrettos, D. S. Callaway, and G. Hug, "Understanding Small-Signal Stability of Low-Inertia Systems," *IEEE Transactions on Power Systems*, pp. 1–1, 2021.

[9] L. Duchesne, E. Karangelos, and L. Wehenkel, "Recent Developments in Machine Learning for Energy Systems Reliability Management," *Proceedings of the IEEE*, vol. 108, no. 9, pp. 1656–1676, 2020.

[10] R. Henriquez-Auba, J. D. Lara, D. S. Callaway, and C. Barrows, "Transient simulations with a large penetration of converter-interfaced generation: Scientific computing challenges and opportunities," *IEEE Electrification Magazine*, vol. 9, no. 2, pp. 72–82, 2021.

[11] P. Panciatici, G. Bareux, and L. Wehenkel, "Operating in the Fog: Security Management Under Uncertainty," *IEEE Power and Energy Magazine*, vol. 10, no. 5, pp. 40–49, Sep. 2012.

[12] G. Strbac, N. Hatziargyriou, J. P. Lopes, C. Moreira, A. Dimeas, and D. Papadaskalopoulos, "Microgrids: Enhancing the Resilience of the European Megagrid," *IEEE Power and Energy Magazine*, vol. 13, no. 3, pp. 35–43, 2015.

[13] F. Bellizio, W. Xu, D. Qiu, Y. Ye, D. Papadaskalopoulos, J. L. Cremer, F. Teng, and G. Strbac, "Transition to Digitalized Paradigms for Security Control and Decentralized Electricity Market," *Proceedings of the IEEE*, pp. 1–18, 2022.

[14] A. Rhodes, "Digitalisation of Energy: An Energy Futures Lab Briefing Paper," 2020.

[15] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, "Smart Transmission Grid: Vision and Framework," *IEEE transactions on Smart Grid*, vol. 1, no. 2, pp. 168–177, 2010.

[16] S. Brahma, R. Kavasseri, H. Cao, N. Chaudhuri, T. Alexopoulos, and Y. Cui, "Real-Time Identification of Dynamic Events in Power Systems Using PMU data, and Potential

Applications—Models, Promises, and Challenges," *IEEE transactions on Power Delivery*, vol. 32, no. 1, pp. 294–301, 2016.

[17] L. A. Wehenkel, *Automatic Learning Techniques in Power Systems*. Norwell, MA, USA: Kluwer Academic Publishers, 1998.

[18] S. Abapour, M. Nazari-Heris, B. Mohammadi-Ivatloo, and M. T. Hagh, "Game Theory Approaches for the Solution of Power System Problems: A Comprehensive Review," *Archives of Computational Methods in Engineering*, vol. 27, no. 1, pp. 81–103, 2020.

[19] J. Tao, M. Umair, M. Ali, and J. Zhou, "The Impact of Internet of Things Supported by Emerging 5G in Power Systems: A Review," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 2, pp. 344–352, 2019.

[20] N. Kulkarni, S. Lalitha, and S. A. Deokar, "Real time control and monitoring of grid power systems using cloud computing." *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 9, no. 2, 2019.

[21] W. Tushar, C. Yuen, T. K. Saha, T. Morstyn, A. C. Chapman, M. J. E. Alam, S. Hanif, and H. V. Poor, "Peer-to-peer energy systems for connected communities: A review of recent advances and emerging challenges," *Applied Energy*, vol. 282, p. 116131, 2021.

[22] M. L. Di Silvestre, P. Gallo, J. M. Guerrero, R. Musca, E. R. Sanseverino, G. Sciumè, J. C. Vásquez, and G. Zizzo, "Blockchain for power systems: Current trends and future applications," *Renewable and Sustainable Energy Reviews*, vol. 119, p. 109585, 2020.

[23] B. Donnot, I. Guyon, A. Marot, M. Schoenauer, and P. Panciatici, "Optimization of Computational Budget for Power System Risk Assessment," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 2018, pp. 1–6.

[24] I. Konstantelos, G. Jamgotchian, S. H. Tindemans, P. Duchesne, S. Cole, C. Merckx, G. Strbac, and P. Panciatici, "Implementation of a Massively Parallel Dynamic Security

Assessment Platform for Large-Scale Grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, 2017.

[25] 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 Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 74–84, 2018.

[26] J. D. Pinzón and D. G. Colomé, "Real-time multi-state classification of short-term voltage stability based on multivariate time series machine learning," *International Journal of Electrical Power & Energy Systems*, vol. 108, pp. 402–414, 2019.

[27] H. Khoshkhoo and S. M. Shahrtash, "Fast online dynamic voltage instability prediction and voltage stability classification," *IET Generation, Transmission & Distribution*, vol. 8, no. 5, pp. 957–965, 2014.

[28] L. Moulin, A. A. Da Silva, M. El-Sharkawi, and R. J. Marks, "Support Vector Machines for Transient Stability Analysis of Large-Scale Power Systems," *IEEE Transactions on Power Systems*, vol. 19, no. 2, 2004.

[29] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support Vector Machine-Based Algorithm for Post-Fault Transient Stability Status Prediction Using Synchronized Measurements," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1474–1483, 2010.

[30] I. Houben, L. Wehenkel, and M. Pavella, "Coupling of k-nn with decision trees for power system transient stability assessment," in *Proceedings of International Conference on Control Applications*. IEEE, 1995, pp. 825–832.

[31] Y. Zhou, Q. Guo, H. Sun, Z. Yu, J. Wu, and L. Hao, "A novel data-driven approach for transient stability prediction of power systems considering the operational variability," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 379–394, 2019.

[32] N. Amjady and S. F. Majedi, "Transient Stability Prediction by a Hybrid Intelligent System," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1275–1283, 2007.

[33] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees," *IEEE Transactions on Power Systems*, vol. 22, no. 4, Nov 2007.

[34] V. Krishnan, J. D. McCalley, S. Henry, and S. Issad, "Efficient Database Generation for Decision Tree Based Power System Security Assessment," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2319–2327, 2011.

[35] M. H. Vasconcelos, L. M. Carvalho, J. Meirinhos, N. Omont, P. Gambier-Morel, G. Jamgotchian, D. Cirio, E. Ciapessoni, A. Pitto, I. Konstantelos, G. Strbac, M. Ferraro, and C. Biasuzzi, "Online Security Assessment with Load and Renewable Generation Uncertainty: The iTesla Project Approach," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2016, pp. 1–8.

[36] S. Samantaray, I. Kamwa, and G. Joos, "Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame," *IET generation, transmission & distribution*, vol. 4, no. 12, pp. 1334–1348, 2010.

[37] J. L. Cremer, I. Konstantelos, S. H. Tindemans, and G. Strbac, "Data-Driven Power System Operation: Exploring the Balance Between Cost and Risk," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 791–801, 2018.

[38] C. Liu, K. Sun, Z. H. Rather, Z. Chen, C. L. Bak, P. Thøgersen, and P. Lund, "A Systematic Approach for Dynamic Security Assessment and the Corresponding Preventive Control Scheme Based on Decision Trees," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 717–730, 2014.

[39] J. L. Cremer, I. Konstantelos, and G. Strbac, "From Optimization-based Machine Learning to Interpretable Security Rules for Operation," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3826–3836, 2019.

[40] I. Kamwa, S. Samantaray, and G. Joós, "On the Accuracy Versus Transparency Trade-Off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 152–161, 2011.

[41] Y. Xu, Z. Y. Dong, J. H. Zhao, P. Zhang, and K. P. Wong, "A Reliable Intelligent System for Real-Time Dynamic Security Assessment of Power Systems," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1253–1263, Aug 2012.

[42] L. Zhu, C. Lu, Z. Y. Dong, and C. Hong, "Imbalance Learning Machine-Based Power System Short-Term Voltage Stability Assessment," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2533–2543, 2017.

[43] M. Sun, I. Konstantelos, and G. Strbac, "A Deep Learning-Based Feature Extraction Framework for System Security Assessment," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5007–5020, 2019.

[44] H. Wang, Q. Chen, and B. Zhang, "Transient stability assessment combined model framework based on cost-sensitive method," *IET Generation, Transmission & Distribution*, vol. 14, no. 12, pp. 2256–2262, 2020.

[45] T. Guo and J. V. Milanović, "Probabilistic Framework for Assessing the Accuracy of Data Mining Tool for Online Prediction of Transient Stability," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 377–385, 2013.

[46] J. L. Cremer and G. Strbac, "A machine-learning based probabilistic perspective on dynamic security assessment," *International Journal of Electrical Power & Energy Systems*, vol. 128, p. 106571, 2021.

[47] B. Donnot, B. Donon, I. Guyon, Z. Liu, A. Marot, P. Panciatici, and M. Schoenauer, "LEAP nets for power grid perturbations," in *ESANN 2019 - 27th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 2019.

[48] B. Donnot, I. Guyon, M. Schoenauer, P. Panciatici, and A. Marot, "Introducing machine learning for power system operation support," in *IREP Symposium*, 2017.

[49] K. A. Loparo and F. Abdel-Malek, "A Probabilistic Approach to Dynamic Power System Security," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 6, pp. 787–798, 1990.

[50] M. He, J. Zhang, and V. Vittal, "Robust Online Dynamic Security Assessment Using Adaptive Ensemble Decision-Tree Learning," *IEEE Transactions on Power Systems*, vol. 28, no. 4, Nov 2013.

[51] M. Sugiyama and M. Kawanabe, *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. MIT press, 2012.

[52] F. Bellizio, J. L. Cremer, and G. Strbac, "Machine-learned security assessment for changing system topologies," *International Journal of Electrical Power & Energy Systems*, vol. 134, p. 107380, 2022.

[53] Miao He, Junshan Zhang, and V. Vittal, "A Data Mining Framework for Online Dynamic Security Assessment: Decision Trees, Boosting, and Complexity Analysis," in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, Jan 2012, pp. 1–8.

[54] 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 Transactions on Power Systems*, vol. 24, no. 2, pp. 832–839, 2009.

[55] G. Granelli, M. Montagna, F. Zanellini, P. Bresesti, R. Vailati, and M. Innorta, "Optimal network reconfiguration for congestion management by deterministic and genetic algorithms," *Electric Power Systems Research*, vol. 76, no. 6-7, pp. 549–556, 2006.

[56] R. Bacher and H. Glavitsch, "Loss Reduction by Network Switching," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 447–454, 1988.

[57] G. Schnyder and H. Glavitsch, "Integrated Security Control Using an Optimal Power Flow and Switching Concepts," *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 782–790, 1988.

[58] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, "Optimal Transmission Switching," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1346–1355, 2008.

[59] C. Crozier, K. Baker, and B. Toomey, "Feasible region-based heuristics for optimal transmission switching," *Sustainable Energy, Grids and Networks*, vol. 30, p. 100628, 2022.

[60] A. Marot, B. Donnot, C. Romero, B. Donon, M. Lerousseau, L. Veyrin-Forrer, and I. Guyon, "Learning to run a power network challenge for training topology controllers," *Electric Power Systems Research*, vol. 189, p. 106635, 2020.

[61] F. R. S. Sevilla, Y. Liu, E. Barocio, P. Korba, M. Andrade, F. Bellizio, J. Bos, B. Chaudhuri, H. Chavez, J. Cremer *et al.*, "State-of-the-art of data collection, analytics, and future needs of transmission utilities worldwide to account for the continuous growth of sensing data," *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107772, 2022.

[62] J. Zhao et al., "Roles of Dynamic State Estimation in Power System Modeling, Monitoring and Operation," *IEEE Transactions on Power Systems*, pp. 1–1, 2020.

[63] A. Fouad, F. Aboytes, V. Carvalho, S. Corey, K. Dhir, and R. Vierra, "Dynamic Security Assessment Practices in North America," *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 1310–1321, 1988.

[64] F. Capitanescu, J. M. Ramos, P. Panciatici, D. Kirschen, A. M. Marcolini, L. Platbrood, and L. Wehenkel, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731–1741, 2011.

[65] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor *et al.*, "Definition and Classification of Power System Stability IEEE/CIGRE Joint Task Force on Stability Terms and Definitions," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1387–1401, 2004.

[66] J. D. McCalley, V. Vittal, and N. Abi-Samra, "An Overview of Risk Based Security Assessment," in *1999 IEEE Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No. 99CH36364)*, vol. 1. IEEE, 1999, pp. 173–178.

[67] J. M. G. Alvarez, "Critical Contingencies Ranking for Dynamic Security Assessment Using Neural Networks," in *2009 15th International Conference on Intelligent System Applications to Power Systems*. IEEE, 2009, pp. 1–6.

[68] J. McCalley, S. Asgarpoor, L. Bertling, R. Billinion, H. Chao, J. Chen, J. Endrenyi, R. Fletcher, A. Ford, C. Grigg *et al.*, "Probabilistic Security Assessment for Power System Operations," in *IEEE Power Engineering Society General Meeting, 2004.* IEEE, 2004, pp. 212–220.

[69] L. Duchesne, E. Karangelos, and L. Wehenkel, "Using Machine Learning to Enable Probabilistic Reliability Assessment in Operation Planning," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–8.

[70] M. Ni, J. D. McCalley, V. Vittal, and T. Tayyib, "Online Risk-Based Security Assessment," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 258–265, 2003.

[71] F. Xiao, J. D. McCalley, Y. Ou, J. Adams, and S. Myers, "Contingency Probability Estimation Using Weather and Geographical Data for On-line Security Assessment," in *2006 International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2006, pp. 1–7.

[72] L. Vanfretti and F. R. S. Sevilla, "A three-layer severity index for power system voltage stability assessment using time-series from dynamic simulations," in *IEEE PES Innovative Smart Grid Technologies, Europe*. IEEE, 2014, pp. 1–5.

[73] Y. Pipelzadeh, R. Moreno, B. Chaudhuri, G. Strbac, and T. C. Green, "Corrective Control With Transient Assistive Measures: Value Assessment for Great Britain Transmission System," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1638–1650, 2016.

[74] R. Moreno, D. Pudjianto, and G. Strbac, "Transmission Network Investment With Probabilistic Security and Corrective Control," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3935–3944, 2013.

[75] I. Konstantelos, M. Sun, S. H. Tindemans, S. Issad, P. Panciatici, and G. Strbac, "Using vine copulas to generate representative system states for machine learning," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 225–235, 2019.

[76] M. Sajjadi, M. Seyedhosseini, and T. Tasdizen, "Disjunctive normal networks," *Neurocomputing*, vol. 218, pp. 276–285, 2016.

[77] M. Seyedhosseini and T. Tasdizen, "Disjunctive normal random forests," *Pattern Recognition*, vol. 48, no. 3, pp. 976–983, 2015.

[78] Z. Wan, Z. Yu, L. Shu, Y. Zhao, H. Zhang, and K. Xu, "Intelligent optical performance monitor using multi-task learning based artificial neural network," *Optics express*, vol. 27, no. 8, pp. 11 281–11 291, 2019.

[79] F. Bellizio, J. L. Cremer, M. Sun, and G. Strbac, "A causality based feature selection approach for data-driven dynamic security assessment," *Electric Power Systems Research*, vol. 201, p. 107537, 2021.

[80] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe, "Direct importance estimation for covariate shift adaptation," *Annals of the Institute of Statistical Mathematics*, 2008.

[81] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[82] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature Selection: A Data Perspective," *ACM Computing Surveys*, vol. 50, 01 2016.

[83] R. Zhang, Y. Xu, Z. Y. Dong, and D. J. Hill, "Feature selection for intelligent stability assessment of power systems," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–7.

[84] K. Yu, X. Guo, L. Liu, J. Li, H. Wang, Z. Ling, and X. Wu, "Causality-based Feature Felection: Methods and Evaluations," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–36, 2020.

[85] B. Schölkopf, "Causality For Machine Learning," *CoRR*, vol. abs/1911.10500, 2019.

[86] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of statistical planning and inference*, vol. 90, no. 2, pp. 227–244, 2000.

[87] D. Koller and M. Sahami, "Toward Optimal Feature Selection," Stanford InfoLab, Tech. Rep., 1996.

[88] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed Energy Resources Topology Identification via Graphical Modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2682–2694, July 2017.

[89] S. Bolognani, "Grid Topology Identification via Distributed Statistical Hypothesis Testing," in *Big Data Application in Power Systems*. Elsevier, 2018, pp. 281–301.

[90] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, May 1968.

[91] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[92] N. Friedman, "Building Classifiers Using Bayesian Networks," in *In Proceedings of the thirteenth national conference on artificial intelligence*. AAAI Press, 1996, pp. 1277–1284.

[93] C. Ratanamahatana, "CloNI: Clustering Of Square Root Of N-interval Discretization," *WIT Transactions on Information and Communication Technologies*, vol. 29, 2003.

[94] C. F. Aliferis, I. Tsamardinos, and A. Statnikov, "HITON: A Novel Markov Blanket Algorithm for Optimal Variable Selection," in *AMIA Annual Symposium Proceedings*, 2003.

[95] I. Tsamardinos, C. F. Aliferis, and A. R. Statnikov, "Algorithms for Large Scale Markov Blanket Discovery," in *FLAIRS Conference*, 2003.

[96] J. Shen, L. Li, and W.-K. Wong, "Markov Blanket Feature Selection for Support Vector Machines," in *AAAI*, vol. 8, 2008, pp. 696–701.

[97] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," *J. Mach. Learn. Res.*, vol. 5, Dec. 2004.

[98] E. Schaffernicht and H.-M. Gross, "Weighted Mutual Information for Feature Selection," in *International Conference on Artificial Neural Networks*. Springer, 2011, pp. 181–188.

[99] N. Friedman and M. Goldszmidt, "Discretizing Continuous Attributes While Learning Bayesian Networks," in *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., 1996, pp. 157–165.

[100] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical Representation of Distribution System Loads Using Gaussian Mixture Model," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 29–37, 2009.

[101] M. B. Cain, R. P. O'neill, A. Castillo *et al.*, "History of Optimal Power Flow and Formulations," *Federal Energy Regulatory Commission*, vol. 1, pp. 1–36, 2012.

[102] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoost," *Statistics and its Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[103] M. Chen, Q. Liu, S. Chen, Y. Liu, C.-H. Zhang, and R. Liu, "XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System," *IEEE Access*, vol. 7, pp. 13 149–13 158, 2019.

[104] J. Geeganage, U. Annakkage, T. Weekes, and B. A. Archer, "Application of Energy-Based Power System Features for Dynamic Security Assessment," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1957–1965, 2014.

[105] I. N. Kassabalidis, M. A. El-Sharkawi, R. J. Marks, L. S. Moulin, and A. A. Da Silva, "Dynamic Security Border Identification Using Enhanced Particle Swarm Optimization," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 723–729, 2002.

[106] C. Donnat and S. Holmes, "Tracking Network Dynamics: A Survey Using Graph Distances," *Annals of Applied Statistics*, vol. 12, no. 2, pp. 971–1012, 06 2018.

[107] G. Nalbantov, P. Groenen, and C. Bioch, "Nearest Convex Hull Classification," Tech. Rep., 2006.

[108] A. Murray, M. Kyesswa, P. Schmurr, H. K. Çakmak, and V. Hagenmeyer, "A Comparison of Partitioning Strategies in AC Optimal Power Flow," *CoRR*, vol. abs/1911.11516, 2019.

[109] S. Blumsack, P. Hines, M. Patel, C. Barrows, and E. C. Sanchez, "Defining Power Network Zones from Measures of Electrical Distance," in *2009 IEEE Power & Energy Society General Meeting*, 2009, pp. 1–8.

[110] J. L. Cremer, I. Konstantelos, G. Strbac, and S. H. Tindemans, "Sample-Derived Disjunctive Rules for Secure Power System Operation," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018, pp. 1–6.

[111] PJM, "Manual 03: Transmission Operations," 2012. [Online]. Available: https://www.pjm.com/-/media/documents/manuals/m03.ashx

[112] M. Rostami and M. Boolochi, "How DERs Could Change Grid Topology and Affect System Performance," *IEEE Smart Grid Newsletters*, 2020.

[113] M. Panteli and P. Mancarella, "The Grid: Stronger, Bigger, Smarter?: Presenting a Conceptual Framework of Power System Resilience," *IEEE Power and Energy Magazine*, vol. 13, no. 3, pp. 58–66, 2015.

[114] F. Bellizio, J. L. Cremer, and G. Strbac, "Transient Stable Corrective Control Using Neural Lyapunov Learning," *IEEE Transactions on Power Systems*, 2022. *Under Review.*

[115] N. Guha, Z. Wang, M. Wytock, and A. Majumdar, "Machine Learning for AC Optimal Power Flow," *arXiv preprint arXiv:1910.08842*, 2019.

[116] R. Canyasse, G. Dalal, and S. Mannor, "Supervised Learning for Optimal Power Flow as a Real-Time Proxy," in *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2017, pp. 1–5.

[117] S. Abhyankar, G. Geng, M. Anitescu, X. Wang, and V. Dinavahi, "Solution techniques for transient stability-constrained optimal power flow–part i," *IET Generation, Transmission & Distribution*, vol. 11, no. 12, pp. 3177–3185, 2017.

[118] G. Geng, S. Abhyankar, X. Wang, and V. Dinavahi, "Solution techniques for transient stability-constrained optimal power flow–part ii," *IET Generation, Transmission & Distribution*, vol. 11, no. 12, pp. 3186–3193, 2017.

[119] V. Miranda, J. Fidalgo, J. Lopes, and L. Almeida, "Real time preventive actions for transient stability enhancement with a hybrid neural network-optimization approach," *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 1029–1035, 1995.

[120] M. La Scala, M. Trovato, and C. Antonelli, "On-line Dynamic Preventive Control: An Algorithm for Transient Security Dispatch," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 601–610, 1998.

[121] Q. Jiang and Z. Huang, "An Enhanced Numerical Discretization Method for Transient Stability Constrained Optimal Power Flow," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1790–1797, 2010.

[122] M. Pavella, D. Ernst, and D. Ruiz-Vega, *Transient stability of power systems: a unified approach to assessment and control*. Springer Science & Business Media, 2012.

[123] A. Michel, A. Fouad, and V. Vittal, "Power System Transient Stability Using Individual Machine Energy Functions," *IEEE Transactions on Circuits and Systems*, vol. 30, no. 5, pp. 266–276, 1983.

[124] P. Varaiya, F. F. Wu, and R.-L. Chen, "Direct Methods for Transient Stability Analysis of Power Systems: Recent Results," *Proceedings of the IEEE*, vol. 73, no. 12, pp. 1703–1715, 1985.

[125] T. L. Vu and K. Turitsyn, "Lyapunov functions family approach to transient stability assessment," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1269–1277, 2015.

[126] P. Giesl and S. Hafstein, "Review on Computational Methods for Lyapunov Functions," *Discrete and Continuous Dynamical Systems-Series B*, vol. 20, no. 8, pp. 2291–2331, 2015.

[127] V. Petridis and S. Petridis, "Construction of Neural Network Based Lyapunov Functions," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 5059–5065.

[128] N. Noroozi, P. Karimaghaee, F. Safaei, and H. Javadi, "Generation of Lyapunov Functions by Neural Networks," in *Proceedings of the World Congress on Engineering*, vol. 2008, 2008.

[129] S. M. Richards, F. Berkenkamp, and A. Krause, "The Lyapunov Neural Network: Adaptive Stability Certification for Safe Learning of Dynamical Systems," in *Conference on Robot Learning*, 2018, pp. 466–476.

[130] Y.-C. Chang, N. Roohi, and S. Gao, "Neural Lyapunov Control," *arXiv preprint arXiv:2005.00611*, 2020.

[131] W. Cui and B. Zhang, "Lyapunov-Regularized Reinforcement Learning for Power System Transient Stability," *IEEE Control Systems Letters*, vol. 6, pp. 974–979, 2022.

[132] T. Huang, S. Gao, X. Long, and L. Xie, "A Neural Lyapunov Approach to Transient Stability Assessment in Interconnected Microgrids," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021, p. 3330.

[133] T. Zhao, J. Wang, X. Lu, and Y. Du, "Neural Lyapunov Control for Power System Transient Stability: A Deep Learning-based Approach," *IEEE Transactions on Power Systems*, 2021.

[134] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*. John Wiley & Sons, 2008.

[135] G. Li and S. Rovnyak, "Integral Square Generator Angle Index for Stability Ranking and Control," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 926–934, 2005.

[136] B. Severino and K. Strunz, "Enhancing Transient Stability of DC Microgrid by Enlarging the Region of Attraction Through Nonlinear Polynomial Droop Control," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 11, pp. 4388–4401, 2019.

[137] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*. Wiley-interscience New York, 1972, vol. 1.

[138] S. Gao, J. Avigad, and E. M. Clarke, "δ-Complete Decision Procedures for Satisfiability over the Reals," in *International Joint Conference on Automated Reasoning*. Springer, 2012, pp. 286–300.

[139] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco, "Implementation of a Large-Scale Optimal Power Flow Solver Based on Semidefinite Programming," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3987–3998, 2013.

[140] J. H. Chow, *Time-scale modeling of dynamic networks with applications to power systems*. Springer, 1982, vol. 46.

[141] A. Shakoor, G. Davies, G. Strbac, D. Pudjianto, F. Teng, and D. Papadaskalopoulos, "Roadmap for Flexibility Services to 2030: A Report to the Committee on Climate Change," 2017.

[142] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.

[143] J. Q. Tortós and V. Terzija, "Controlled islanding strategy considering power system restoration constraints," in *2012 IEEE Power and Energy Society General Meeting*. IEEE, 2012, pp. 1–8.

[144] D. P. Wadduwage, C. Q. Wu, and U. Annakkage, "Power system transient stability analysis via the concept of Lyapunov Exponents," *Electric Power Systems Research*, vol. 104, pp. 183–192, 2013.

[145] I. Genc, R. Diao, V. Vittal, S. Kolluri, and S. Mandal, "Decision Tree-Based Preventive and Corrective Control Applications for Dynamic Security Enhancement in Power Systems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1611–1619, 2010.

[146] M. Joos and I. Staffell, "Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany," *Renewable and Sustainable Energy Reviews*, vol. 86, pp. 45–65, 2018.

[147] B. Kroposki, B. Johnson, Y. Zhang, V. Gevorgian, P. Denholm, B.-M. Hodge, and B. Hannegan, "Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy," *IEEE Power and energy magazine*, vol. 15, no. 2, pp. 61–73, 2017.

[148] J. Fang, H. Li, Y. Tang, and F. Blaabjerg, "On the inertia of future more-electronics power systems," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 7, no. 4, pp. 2130–2146, 2018.

[149] F. Milano, F. Dörfler, G. Hug, D. J. Hill, and G. Verbič, "Foundations and challenges of low-inertia systems," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–25.

[150] T. Guo and J. V. Milanović, "Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 377–385, 2013.

[151] S. You, Y. Zhao, M. Mandich, Y. Cui, H. Li, H. Xiao, S. Fabus, Y. Su, Y. Liu, H. Yuan, H. Jiang, J. Tan, and Y. Zhang, "A review on artificial intelligence for grid stability assessment," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2020, pp. 1–6.

[152] F. Thams, A. Venzke, R. Eriksson, and S. Chatzivasileiadis, "Efficient database generation for data-driven security assessment of power systems," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 30–41, 2020.

[153] H. Sawhney and B. Jeyasurya, "A feed-forward artificial neural network with enhanced feature selection for power system transient stability assessment," *Electric power systems research*, vol. 76, no. 12, pp. 1047–1054, 2006.

[154] T. Zhang, M. Sun, J. L. Cremer, N. Zhang, G. Strbac, and C. Kang, "A Confidence-Aware Machine Learning Framework for Dynamic Security Assessment," *IEEE Transactions on Power Systems*, 2021.

[155] R. Liu, G. Verbič, J. Ma, and D. J. Hill, "Fast stability scanning for future grid scenario analysis," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 514–524, 2018.

[156] P. N. Papadopoulos and J. V. Milanović, "Methodology for online identification of dynamic behavior of power systems with an increased amount of power electronics interface units," *CSEE Journal of Power and Energy Systems*, vol. 5, no. 2, pp. 171–180, 2019.

[157] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load Modeling—A Review," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5986–5999, 2018.

[158] B. Stott, "Power system dynamic response calculations," *Proceedings of the IEEE*, vol. 67, no. 2, pp. 219–241, 1979.

[159] M. V. Reddy, B. P. Muni, and A. Sarma, "Enhancement of Voltage Profile for IEEE 14 Bus System with Inter Line Power Flow Controller," in *2016 Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy (PESTSE)*. IEEE, 2016, pp. 1–5.

[160] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science & Business Media, 2006.

[161] J. D. Lara, C. Barrows, D. Thom, D. Krishnamurthy, and D. Callaway, "Powersystems. jl—a power system data management package for large scale modeling," *SoftwareX*, vol. 15, p. 100747, 2021.

[162] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward, "Sundials: Suite of nonlinear and differential/algebraic equation solvers," *ACM Transactions on Mathematical Software (TOMS)*, vol. 31, no. 3, pp. 363–396, 2005.

[163] D. C. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg, "Electric Load Forecasting Using an Artificial Neural Network," *IEEE transactions on Power Systems*, vol. 6, no. 2, pp. 442–449, 1991.

[164] U. Tamrakar, D. Shrestha, M. Maharjan, B. P. Bhattarai, T. M. Hansen, and R. Tonkoski, "Virtual inertia: Current trends and future directions," *Applied Sciences*, vol. 7, no. 7, p. 654, 2017.

[165] P. Pourbeik, J. J. Sanchez-Gasca, J. Senthil, J. D. Weber, P. S. Zadehkhost, Y. Kazachkov, S. Tacke, J. Wen, and A. Ellis, "Generic Dynamic Models for Modeling Wind Power Plants and Other Renewable Technologies in Large-Scale Power System Studies," *IEEE Transactions on Energy Conversion*, vol. 32, no. 3, pp. 1108–1116, 2017.

[166] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021.

[167] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, "Physics-Informed Neural Networks for Power Systems," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[168] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Split-Based Sequential Sampling for Realtime Security Assessment," *Available at SSRN 4059939*, 2022.

[169] J. Stiasny, G. S. Misyris, and S. Chatzivasileiadis, "Physics-Informed Neural Networks for Non-Linear System Identification for Power System Dynamics," in *2021 IEEE Madrid PowerTech*. IEEE, 2021, pp. 1–6.

[170] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia, "Learning to Simulate Complex Physics with Graph Networks," in *International Conference on Machine Learning*. PMLR, 2020, pp. 8459–8468.

[171] Y. Chen, Y. Shi, and B. Zhang, "Optimal Control Via Neural Networks: A Convex Approach," *arXiv preprint arXiv:1805.11835*, 2018.

[172] D. Ernst, M. Glavic, and L. Wehenkel, "Power Systems Stability Control: Reinforcement Learning Framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, 2004.

[173] J. Zhang, C. Lu, C. Fang, X. Ling, and Y. Zhang, "Load Shedding Scheme with Deep Reinforcement Learning to Improve Short-term Voltage Stability," in *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 13–18.

[174] R. R. Hossain, Q. Huang, and R. Huang, "Graph Convolutional Network-Based Topology Embedded Deep Reinforcement Learning for Voltage Stability Control," *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 4848–4851, 2021.

[175] C. O'Malley, M. Aunedi, F. Teng, and G. Strbac, "Value of Fleet Vehicle to Grid in Providing Transmission System Operator Services," in *2020 Fifteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*. IEEE, 2020, pp. 1–8.

[176] S. S. Ravi and M. Aziz, "Utilization of Electric Vehicles for Vehicle-to-Grid Services: Progress and Perspectives," *Energies*, vol. 15, no. 2, p. 589, 2022.