FEDERAL UNIVERSITY OF SANTA CATARINA
CENTER OF TECHNOLOGY
POSTGRADUATE PROGRAM IN ELECTRICAL ENGINEERING

Dionatan Augusto Guimarães Cieslak

Contributions on High Impedance Fault Classification in Microgrids Using Harmonic Synchrophasors

Florianópolis
2024

Dionatan Augusto Guimarães Cieslak

Contributions on High Impedance Fault Classification in Microgrids Using Harmonic Synchrophasors

Thesis submitted to the Postgraduate Program in Electrical Engineering of the Federal University of Santa Catarina as a requirement for obtaining the degree of PhD in Electrical Engineering.

Supervisor: Prof. Miguel Moreto, PhD
Co-supervisor: Prof. André E. Lazzaretti, PhD

Florianópolis
2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Cieslak, Dionatan Augusto Guimarães
Contributions on high impedance fault classification in microgrids using harmonic synchrophasors / Dionatan Augusto Guimarães Cieslak ; orientador, Miguel Moreto, coorientador, André Eugênio Lazzaretti, 2024.
109 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2024.

Inclui referências.

1. Engenharia Elétrica. 2. Sincrofasores harmônicos. 3. Falta de alta impedância. 4. Classificação de eventos. 5. Microrredes. I. Moreto, Miguel. II. Lazzaretti, André Eugênio. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Dionatan Augusto Guimarães Cieslak

Contributions on High Impedance Fault Classification in Microgrids Using Harmonic Synchrophasors

O presente trabalho em nível de doutorado foi avaliado e aprovado, em 24 de abril de 2024, pela banca examinadora composta pelos seguintes membros:

Prof. Adriano Peres de Morais, Dr.
Universidade Federal de Santa Maria - UFSM

Prof. Diego Issicaba, Dr.
Universidade Federal de Santa Catarina - UFSC

Prof. José Rubens Macedo Junior, Dr.
Universidade Federal de Uberlândia - UFU

Prof. Miguel Moreto, Dr.
Orientador
Universidade Federal de Santa Catarina - UFSC

Prof. André Eugênio Lazaretii, Dr.
Coorientador
Universidade Tecnológica Federal do Paraná - UTFPR

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Engenharia Elétrica.

ICP-Edu
Documento assinado digitalmente
Telles Brunelli Lazzarin
Data: 07/05/2024 11:25:20-0300
CPF: ***.400.719-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Telles Brunelli Lazzarin, Dr.
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

ICP-Edu
Documento assinado digitalmente
Miguel Moreto
Data: 29/04/2024 16:42:55-0300
CPF: ***.850.100-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Miguel Moreto, Dr.
Orientador

Florianópolis, 2024

To my dear parents.

ACKNOWLEDGEMENTS

First and foremost, I thank God for guiding me through all of the adversities that this journey has presented.

To my parents, Ildes and Geraldo, who have always shown me the worth and importance of knowledge seeking. Thank you so much for all of the difficult times when you sacrificed your preferences for my formal education. For all the support, my deepest thanks.

To my fiancée, Fernanda, for her unwavering support throughout the process, specially when I was discouraged.

To my supervisor, Prof. Miguel, for all of his teachings throughout my academic career, including my bachelor's, master's, and doctorate. His attitude has always been helpful and diligent, and it serves as a reference for me. It was a delight to have you as a mentor, and it is equally enjoyable to have you as a friend.

To my second supervisor, Prof. André, for his significant contributions to the success of this work. You are another point of reference for me, both professionally and personally. Your good sense in dealing with others is unquestionably a goal that I will strive toward.

My heartfelt gratitude to everyone who helped me to develop this work, whether directly or indirectly.

Finally, to the Federal University of Technology - Parana, Dois Vizinhos and Pato Branco (UTFPR), for the license granted to me to develop research full time.

"Do, or do not. There is no try."
Yoda

ABSTRACT

The advancements in monitoring technologies have made possible the deployment of Phasor Measurement Units (PMUs) in distribution networks. This is especially interesting if active distribution networks are considered, as the usage of these devices can significantly increase the amount of available information for supporting system operation. In addition, one of the most relevant impacts due to the integration of renewable energy sources into the current distribution networks is the presence of harmonic content in the system. These two aspects set challenges for dealing with the new era of large amounts of complex data in modern systems. In this sense, this work aims to contribute to the field of events classification in active distribution systems, emphasizing the classification of high impedance faults, through measurements of PMUs and harmonic synchrophasors. Using classification approaches based on features of the measured signals, as well as their respective time series, the work explores the state of the art of machine learning models in order to establish appropriate strategies for events classification. The impact of the quality of the measurements is also investigated within the scope of the classification, seeking to establish robustness in the application of PMU data. Real data are also evaluated in order to establish an adequate level of generalization of classification strategies, establishing a good alternative for application in real contexts.

Keywords: Harmonic Synchrophasors. Microgrids. High Impedance Fault. Event Classification.

RESUMO

Os avanços nas tecnologias de monitoramento possibilitaram a implantação de Unidades de Medição Fasorial (do inglês, PMUs) em redes de distribuição. Isto é especialmente interessante se forem consideradas redes de distribuição ativas, pois a utilização destes dispositivos pode aumentar significativamente a quantidade de informação disponível para apoiar a operação dos sistemas. Além disso, um dos impactos mais relevantes devido à integração de fontes de energia renováveis nas atuais redes de distribuição é a presença de conteúdo harmônico no sistema. Estes dois aspectos colocam desafios para lidar com a nova era de grandes quantidades de dados complexos em sistemas modernos. Nesse sentido, este trabalho visa contribuir com a área de classificação de eventos em sistemas de distribuição ativos, destacando a classificação de faltas de alta impedância, através de medições de PMUs e sincrofasores harmônicos. Utilizando abordagens de classificação baseadas em características dos sinais medidos, bem como suas respectivas séries temporais, o trabalho explora o estado da arte dos modelos de aprendizado de máquina a fim de estabelecer estratégias apropriadas para classificação de eventos. O impacto da qualidade das medições também é investigado no âmbito da classificação, buscando estabelecer robustez na aplicação dos dados da PMU. Dados reais também são avaliados a fim de estabelecer um nível adequado de generalização das estratégias de classificação, estabelecendo uma boa alternativa para aplicação em contextos reais.

Keywords: Sincrofasores Harmônicos. Microrredes. Falta de Alta Impedância. Classificação de Eventos.

RESUMO EXPANDIDO

INTRODUÇÃO

Nos últimos anos, tem havido um aumento significativo na complexidade das redes elétricas modernas devido à integração crescente de fontes de energia renovável. Isso tem levado a um aumento no conteúdo harmônico devido às características dessas fontes e também de cargas não-lineares, desafiando a estabilidade e confiabilidade do sistema. No entanto, essa complexidade também tem impulsionado avanços na estrutura de monitoramento, resultando em melhorias na operação e proteção dos sistemas elétricos. Com o surgimento de dispositivos de medição cada vez mais presentes nos sistemas elétricos modernos, incluindo Unidades de Medição de Fasorial (do inglês, PMUs) até mesmo em níveis de distribuição, há uma quantidade crescente de dados disponíveis para processamento e análise. No entanto, a multiplicidade de sistemas de medição apresenta desafios relacionados ao volume de informação a ser processada e analisada. Para lidar com essa crescente complexidade e volume de dados, têm surgido métodos baseados em dados para classificação de eventos em redes modernas. Esses métodos visam aproveitar a riqueza de informações disponíveis para melhorar a detecção e resposta a eventos como falhas ou perturbações na rede.

OBJETIVOS

O principal objetivo deste trabalho é avaliar o uso de sincrofasores harmônicos para distinguir falhas de alta impedância de outras perturbações comuns em microrredes, levando em consideração a penetração de fontes de energia renovável. Para alcançar esse objetivo, serão simuladas as classes mais comuns de eventos em microrredes, considerando a penetração de fontes de energia renovável, a fim de avaliar as características promissoras dos dados de sincrofasores. Além disso, pretende-se explorar o uso de séries temporais de sincrofasores e avaliar a robustez dos resultados diante de questões de qualidade de dados. Por fim, os resultados obtidos serão validados com dados reais, buscando assim contribuir para o avanço do conhecimento sobre o uso de sincrofasores na detecção e classificação de perturbações em microrredes.

METODOLOGIA

O trabalho apresenta a proposição de duas estratégias para realização da classificação de eventos em microrredes baseadas em dados obtidos de PMUs: uma delas baseada na extração de características do sinal medido, e a outra, baseada na própria série temporal. Para a abordagem baseada em extração de características, os dados de duas PMUs são utilizados de modo independente, sendo estes segmentados de acordo com a ocorrência do evento. Oito conjuntos de características são testados com seis modelos clássicos de aprendizado de máquinas. Já para a segunda estratégia, baseada em séries temporais, os dados das duas PMUs são combinados, dando origem a uma única série temporal que, não sofre nenhum processo de segmentação. Deste modo, dois modelos de aprendizado de máquina para classificação de séries temporais são avaliados.

RESULTADOS E DISCUSSÃO

Ambas as metodologias foram comparadas em termos da qualidade dos dados de PMU utilizados para fins de classificação. Mesmo apresentando bons resultados para a maioria das análises, quando considerando o melhor conjunto de características, combinado com o melhor modelo de aprendizado de máquina treinado, a metodologia baseada em características apresentou algumas fragilidades principalmente quando se deparou com dados faltantes nos registros das PMUs. Esses aspectos foram felizmente superados quando utilizando a abordagem baseada em séries temporais, mais precisamente, quando utilizando um modelo de estado da arte de aprendizado de máquina, conhecido com Rede Neural Transformer, cujos resultados superaram todos os demais modelos testados em todas as metodologias.

CONCLUSÕES E CONTRIBUIÇÕES

As conclusões desta tese destacam que uma única amostragem de fasor por ciclo de informação é suficiente para alcançar um equilíbrio satisfatório entre a quantidade de dados e a precisão geral para a maioria dos modelos de aprendizado de máquina considerados, viabilizando assim o uso de abordagens com PMUs comerciais de baixo custo. Além disso, foi observado que o uso de informações não combinadas de PMUs aumenta a dimensionalidade do problema sem necessariamente melhorar os resultados de classificação. O estudo também revelou que a presença de ruído nos sinais degrada os níveis de erro de estimação fasorial para harmônicas de ordem elevada, enquanto a perda de sincronismo é a questão de qualidade de dados menos prejudicial para fins de classificação. A análise mostrou que o atributo do mecanismo de atenção presente na rede neural Transformer se mostra vantajoso em todos os cenários de qualidade de dados, especialmente em dados ausentes, e que o uso de informações angulares combinadas melhorou os resultados de classificação.

As principais contribuições deste trabalho incluem o uso de sincrofasores harmônicos para classificação de eventos em microrredes, a avaliação dos requisitos de taxa de estimativa de fasor para fins de classificação de eventos e uma análise abrangente da qualidade dos dados de PMU, incluindo ruído, erro, dados ausentes e falta de sincronismo. Contribuições secundárias incluem a disponibilização de um repositório público de arquivos de simulação. Esses resultados têm o potencial de informar e orientar futuras pesquisas no campo da classificação de eventos em microrredes elétricas.

Keywords: Sincrofasores Harmônicos. Microrredes. Falta de Alta Impedância. Classificação de Eventos.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AC	Alternate Current
CBS	Capacitor Bank Switching
CIGRE	Conseil International des Grands Réseaux Electriques
CNN	Convolutional Neural Network
DC	Direct Current
DER	Distributed Energy Resources
DFT	Discrete Fourier Transform
DG	Distributed Generation
DGS	Distributed Generation Switching
DT	Decision Tree
GPS	Global Positioning System
HIF	High Impedance Fault
IGBT	Insulated-Gate Bipolar Transistor
kNN	k-Nearest Neighbors
LIF	Low Impedance Fault
LOV	Load Variation
LR	Logistic Regression
LSTM	Long Short-term Memory network
LV	Low Voltage
MLP	Multilayer Perceptron
MV	Medium Voltage
NOP	Normal Operation
PCC	Point of Common Coupling
PG	Single Phase
PMU	Phasor Measurment Unit
PP	Double Phase
PPG	Double Phase Grounded
PPP	Three Phase
PV	Photovoltaic Plant
PWM	Pulse Width Modulation
RF	Random Forest
RNN	Recurrent Neural Network
SCADA	Supervisory Control and Data Acquisition
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
TRE	Transformer Energization
TST	Transformer Network

TVE	Total Vector Error
WAMS	Wide-Area Measurement System
WT	Wind Plant

CONTENTS

1 INTRODUCTION

The evolution of contemporary power systems has undergone a notable transformation encompassing the realms of generation, transmission, and distribution of electrical energy. The early 20th century witnessed a fundamental shift towards centralized power generation predominantly reliant on conventional fossil fuels. These large-scale power generation facilities assumed a central role in electricity production, ensuring a consistent energy supply to burgeoning urban centers. This transition marked a substantive departure from preceding decentralized modes of energy production, facilitating the establishment of extensive and efficient transmission networks. However, as concerns pertaining to environmental sustainability and system resilience gained ascendancy, a discernible shift towards decentralization emerged. In consonance with this evolution, the concept of microgrids rises up, representing a progression of modern power systems, by integrating renewable energy resources, such as solar, wind, and hydroelectric power, into the existing grid framework. Microgrids, characterized by their localized and self-contained nature, have emerged as a critical element in the pursuit of a more sustainable and resilient energy infrastructure (BEVRANI; WATANABE; MITANI, 2014; GHORBANIAN et al., 2019).

Microgrids thus emerge as a pivotal innovation in this paradigmatic shift. These networks employ advanced control mechanisms and energy storage technologies to proficiently oversee energy generation and consumption within a delimited geographic domain. The significance of microgrids is underscored by their capacity to bolster grid resilience, mitigate transmission losses, and serve as a conduit for the seamless assimilation of renewable energy assets. Furthermore, they assume a central role in augmenting energy accessibility in remote or underserved communities, fostering energy self-sufficiency, and contributing substantively to the broader sustainability of the contemporary power landscape. These technological advances add more uncertainties to the operation, protection, and control procedures of the system. In addition, the constant deregulation of energy market is also a challenging aspect for the correct operation of the systems, since the greater participation of agents of distributed generation, DG, in the system evidences problems related to the inversion of the power flow, modification of the levels of short-circuit and greater presence of harmonic components in the system. The present and the future of power systems are increasingly dependent on how to integrate this complexity in operation with communication and information infrastructures to improve grid monitoring, control and management (KHETARPAL; TRIPATHI, 2020).

Such integration is strongly dependent on the quality of the information available in this sense, the adequate acquisition of electrical network data is mandatory. The Supervisory Control and Data Acquisition, SCADA, systems have long played an important role in power systems. The challenging issues for SCADA systems have changed as a result of new communication technologies and the need for quick access to power grid information. Low sampling rates (2–4 samples per cycle) and a lack of time synchronization are two issues that SCADA systems face. Synchrophasor technology appears to be a viable alternative in this regard with considerable higher sampling rates, when compared to SCADA technology. Since recent years, the tremendous development of information and communication technologies has allowed and increased the flexibility and capability in power systems monitoring, with large acquisition and fast broadcast of data. Among all these technologies, Wide-Area Measurement Systems (WAMS) using Phasor Measurement Units (PMUs) are one of the most preeminent tools for enhancing system's observation (BEVRANI; WATANABE; MITANI, 2014). Furthermore, given the dynamic behavior of a power system, PMUs can be quite useful in monitoring it (HOJABRI et al., 2019).

Historically, PMUs are broadly used to observe events in transmission systems. However, in recent years, the deployment of monitoring systems based on PMUs on the distribution level has grown significantly. A large spectrum of potential application in distribution system can be supported by the usage of PMUs. By the system's point of view, the usage of PMUs essentially compound diagnostics applications, that help operators and planners to better understand the past and present conditions of the distribution system, and control applications, that lens to notify specific actions to directly alter the operation of the distribution network. Only high-resolution, time-stamped voltage magnitude measurements are needed for some applications, while other applications may be also supported by using current measurements, frequency and also, phase angle data information (VON MEIER et al., 2017). When exploring the development of diagnostic approaches supported by PMUs, some applications stand out: event detection and classification; topology detection; model validation; DG characterization; microgrid operation; distribution state estimation and phasor-based control. (LIAO; STEWART; KARA, 2016; VON MEIER et al., 2017; SHARMA; SAMANTARAY, 2019; BHATTARAI et al., 2019; LIU, Yikui; WU, L.; LI, J., 2020).

Several issues appear when dealing with PMU data, such as noise, communication congestion, hardware failures, and transmission delays. Therefore, PMUs typically experience different data quality issues, such as noise contamination, missing data and synchronism error. In typical distribution networks, noise varies around 60 and 40 dB (ZHANG, Y. et al., 2020; GHIGA et al., 2018; ROSCOE et al., 2018), and when assessing a distribution system with PMU across the grid, filtered samples may retain part of that noise, impacting synchrophasor accuracy. Moreover, some operator reports establish that the problem of missing data can reach around 30% of information loss (YANG et al., 2020), synchronism error in measurements can reach around 60% (YAO et al., 2016). Such issues in data quality severely restrict some major PMU applications. In this sense, any event classification approach must be robust enough to deal with these specific scenarios of data quality issues. At transmission level, some investigations are dealing with the impact of data quality issues in event classification tasks (LIU, Yunchuan et al., 2022; LI, Z. et al., 2021; YUAN et al., 2021; DENG et al., 2019), however, when dealing with distribution level PMUs, these investigations are noticeably absent in the existing literature.

In the context of this new era of active energy systems with different types of DGs (conventional and renewable), loads (linear, non-linear, and unbalanced loads), switching events (capacitor banks, transformer energizing), and electrical faults (high impedance and low impedance faults), some impacts over the power quality can be imposed, creating a reducing in the lifetime of equipment used in these networks. As a result, whenever a disturbance occurs in the network, it must be detected and, if applicable, isolated as soon as possible. In general, electrical networks experience lots of disturbances, from different natures (faulty or non-faulty), that are hard to be avoided. Even in such adverse conditions, the system is expected to maintain stable operation. Understanding the difference between fault and non-fault events is essential for managing and maintaining systems' operation. Fault events, like sudden disruptions or equipment failures, can cause instability and require immediate attention to prevent system failures. Non-fault events, such as planned operational changes or demand variations, may not pose immediate threats but still need monitoring to maintain grid stability. Recognizing these distinctions allows microgrid operators to prioritize responses, allocate resources efficiently, and ensure operational resilience in dynamic energy environments (ALTAF et al., 2022). Moreover, with the integration of multiple power sources and loads, the impact of local power electronics converters associated with different distributed generators, distribution network operating modes and switching events, significantly increases the presence of harmonics, and as a result, the sensitivity on system protection functions must be properly enhanced and guaranteed (BISWAL et al., 2022; VINAYAGAM et al., 2022).

Among the possible disturbances, there is a class that stands out due to the difficulty in being detected, they are High Impedance Faults (HIFs). These faults are usually caused when power conductors break and touch the ground, or when some type of object (vegetation, for example) comes into contact with the conductors. HIFs set a class of events that stand between the idea of short and open circuit fault. While a short circuit is characterized by large currents originated from low impedance paths, the HIFs are basically the opposite: a disturbance with low currents originated from high impedance paths. Although the HIFs do not produce current magnitudes that overcome the protection devices thresholds, not interfering with the functionality of the system, this class of event does not fit the class of open-circuit faults. Thereby, in HIFs occurrences where the conductor does not break, the system current is not interrupted, and no subtle changes can be perceived by the protection scheme (GOMES et al., 2019).

The undetected behavior of HIFs are dangerous to the system operation because the adjacency of the fault remains energized putting the physical integrity of workers and others at risk. As well as damage to the surroundings of the fault, such as possible damage to equipment and fires caused by electrical arcs (GHADERI; GINN; MOHAMMADPOUR, 2017). For that reason, the accurate identification of different types of power system disturbances using advanced classification techniques is required to ensure safe and reliable operation of power systems in this new complex era of operation. Automatic detection and classification techniques using machine learning approaches are widely used for dealing with events and disturbances in power system networks with more precision than conventional classification methods.

A challenge arises with regard to the monitoring of electrical systems: the technological evolution of electrical systems is unprecedented (new energy sources, increasingly active consumers, new devices and technologies, as well as new business models), and electrical networks are being instrumented with high resolution data detection capabilities and acquisition rate, orders that are much higher than those already seen. Such aspects result in an increase in the complexity and uncertainty of the information (bringing new challenges and opportunities), as well as a significant increase in the amount of information and a need for better data-based decision-making in the system operation (TU et al., 2017). Data analysis has become a kernel of nowadays electricity industry and an advantageous strategy for organizations looking to innovate and provide high levels of service quality and customer satisfaction. The analysis of a large amount of data thus becomes a transformational step towards the future of electric power networks (GUO et al., 2018; KEZUNOVIC et al., 2020).

From a technical point of view, the biggest challenges lie in the treatment of large volumes of data and in the choice of appropriate artificial intelligence and machine learning techniques. Machine learning approaches are considered to be the main components of data science, as they allow finding patterns in the data that provide understanding about the phenomenon described by the data and predictions about future events. These techniques are typically used in problems where conventional analytical techniques are inappropriate, for example, due to the large volume, dimensionality, heterogeneity, diversity of the data (HASSANI et al., 2021). Event classification problems in electric networks are commonly addressed using various machine learning techniques. Linear approaches remain prevalent, particularly for their efficacy in high-dimensional data classification tasks. Within the realm of neural networks, deep learning models based on neural networks have gained significant focus by their capacity of extracting features from complex data and evaluating sequential data. Ensemble and tree based methods are widely used for their interpretability and capacity to handle diverse data type. These techniques, often integrated with feature engineering and optimization methods, collectively offer robust solutions to event classification challenges in distribution networks (RAHMAN FAHIM et al., 2020). Furthermore, within the scope of data analysis in energy systems, there is a shortage of applications of state-of-the-art machine learning models, especially when looking at models based on deep learning, which have already shown themselves to be significantly robust in extremely complex areas, such as natural language processing, image processing, etc. (GALASSI;

LIPPI; TORRONI, 2020).

Harmonic synchrophasors usage in distribution networks represent an innovative approach to power system monitoring and control. By capturing harmonic components alongside fundamental frequencies, they provide a comprehensive view of grid dynamics, enabling accurate diagnosis of power quality issues and equipment faults. This real-time data facilitates proactive maintenance and rapid response to events, ensuring system reliability and stability. Furthermore, the availability of high-resolution synchrophasor data supports advanced analytics and the integration of emerging technologies like renewable energy sources and power electronic devices. Overall, harmonic synchrophasors offer enhanced visibility, diagnostic capabilities, and support for innovation in microgrid operations, driving advancements towards more resilient and sustainable energy systems (CISNEROS-SALDANA et al., 2024).

Considering the aforementioned arguments of rapid and constant development of the energy networks, more specifically by the arising of the microgrids, as well as the development of high-resolution and accurate sensors and their widely presence in the current networks, increasing the amount and the complexity of available information, it is noticed that there is a tremendous motivation in the field power systems data analytics, focusing on improving the methods of detection and classification of the events that the system may experience, with a special attention to events that are well known as difficult to be observed, by using harmonic synchrophasors information from PMUs.

In the big picture, the contributions of this work are on the analysis of PMUs data when used to promote event classification on microgrids. The gaps in the field are investigated to enhance the approaches using harmonic synchrophasors in distinguishing high impedance faults from other faulty and non-faulty events, which is, to the extent of our knowledge, a new approach in the literature. In this way, evaluations of how much data provided by PMU is needed to obtain the best trade-off between the amount of information and accurate classification of events in microgrids are performed, besides of how measurement errors may depreciate the classification task. Moreover, most common PMUs data quality issues are also investigated in order to observed their impact in the classification outcomes.

The following investigations are addressed to enhance supervisory and monitoring strategies in distribution networks analysis through the development of PMU-data-driven approaches of event classification. Moreover, the hypothesis considered in this work allow the application of the strategies for both an active distribution context as well a microgrid one. This work can be divided into two main fronts: the first (features based approach) is an investigation of the quality of data coming from PMUs for the task of classifying high impedance faults, when using a conventional data feature selection approach, when combined with the use of classical models of machine learning. The second (time series based approach), in turn, investigates the use of time series from PMUs for the same task, but now advancing to the frontier of deep learning models in order to evaluate the applicability of time series classification models for problems related to classification of events in microgrids.

The contributions of feature based approach are on the analysis of PMU data when used to promote event classification on microgrids. The gaps in the field are investigated to enhance the approaches using harmonic synchrophasors in distinguishing HIF from other faulty and non-faulty events, which is, to the extent of our knowledge, a new approach in the literature. Accordingly, this investigation focuses on answering the following research question: *What is the quality and how much information is needed to correctly discriminate the most common classes of events in an active distribution network using PMUs?* In this sense, the main contributions of this work are:

- An evaluation of how much data provided by PMU (harmonic synchrophasors) is needed to obtain the best trade-off between the amount of information and accurate classification of events in active distribution networks;
- An investigation of how measurement errors may depreciate the classification task, i.e., how robust the classification is in terms of data quality;
- A validation of the classification models with real data, in order to observe how generalist is the classification in terms of real world applications.

The time series approach incorporates harmonic synchrophasors, combined with an analysis of different types of data quality issues. Additionally, it employs cutting-edge time series classification approaches, demonstrating superior performance compared to existing literature, particularly in cases involving data failure. The applicability of the proposed method is demonstrated through real data application, thereby enabling its effective use in real-world scenarios. The main contributions of this research can be summarized as:

- To combine PMU data in terms of harmonic synchrophasors in order to evaluate the discriminant potential of harmonic patterns of most common events on microgrids.
- To improve the classification of high impedance faults in distribution systems with high level of basal harmonic contents.
- To evaluate the impact of data quality in terms of event classification. The influence of most common PMU data quality issues are investigated.
- To explore the benefits of attention mechanisms from state-of-the-art deep neural network, in order to enhance a methodology that do not require pre-processing and feature defining steps to correct classify events on microgrids, allowing the usage of combined raw data from multiple PMUs, including real synchronized data.

The usage of combined PMU time series for event classification is an innovative solution, moreover, by considering angle information from harmonic synchrophasors time series, a new approach of discriminative information is proposed.

1.1 GENERAL OBJECTIVES

The main objective of this work is to evaluate the usage of harmonics synchrophasors in order to distinguish High Impedance Faults from other common disturbances in microgrids, considering the penetration of renewable energy sources.

1.2 SPECIFIC OBJECTIVES

In order to reach the major objective, the following minor objectives are established:

1. To simulate a microgrid considering penetration of renewable energy sources, with faulty and non-faulty events;
2. To estimate harmonic synchrophasors from waveform of electrical quantities recorded;
3. To evaluate promising features from synchrophasor data;
4. To evaluate the robustness of the classification in terms of data quality;
5. To explore the benefits of attention mechanisms in order to enhance a methodology that do not require pre-processing and feature defining steps to correct classify events on microgrids;
6. To improve state-of-the-art classification scores by using harmonic synchrophasors;
7. To validate the classification models with real data, in order to observe how generalist is the classification in terms of real world applications.

Secondary contributions can be achieved by the publication of the simulation files, datasets, and codes in order to foster future investigation and benchmark. All simulation models and datasets are made publicly available under an open access policy[1], an initiative underexplored in other works.

1.3 THESIS' STRUCTURE

The following document is organized as follows: in Chapter 2 the fundamentals about the major topics considered in this research are presented, including a more detailed review about high impedance faults. In Chapter 3, a review of related works is presented in terms of event classification based on harmonic synchrophasors, with an emphasis on

[1] https://github.com/dionatancieslak/CIGRE-EuropeanMV

high impedance faults. In Chapter 4, the simulation environment is detailed in terms of data generation, processing and classification. The Chapter 6 analyses the usage of PMU data through a feature based approach of event classification, focusing on the impact of some data quality issues in the classification task. The Chapter 5 investigates the usage of harmonic synchrophasors combined with a cutting-edge time series classification approach, demonstrating superior performance compared to existing feature based literature approaches, particularly in cases involving data failure. Finally, Chapter 7 presents the conclusions of the investigations and the future works to be considered in the field. A mind map of the thesis' structure is presented in Figure 1.

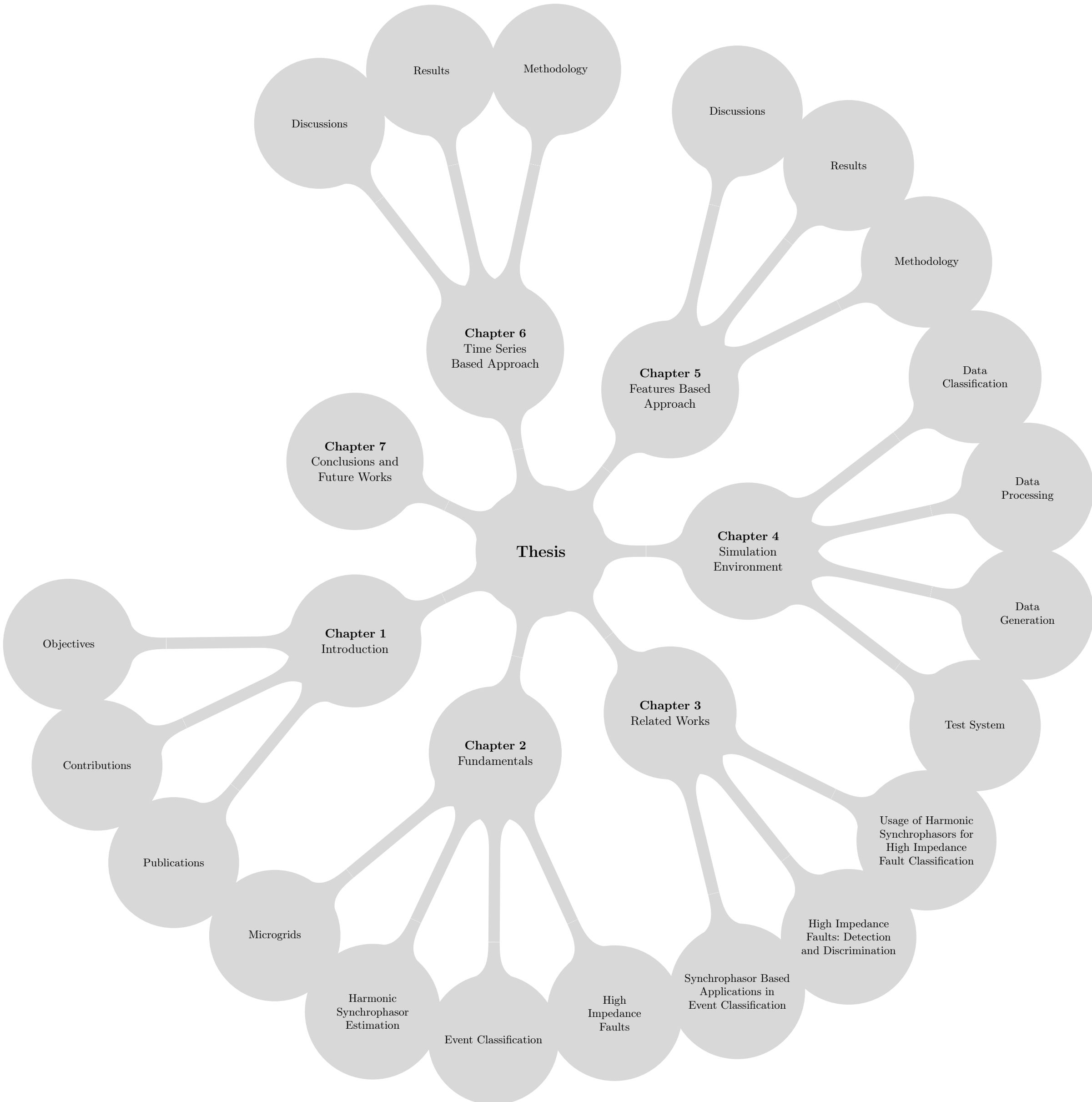


Figure 1 – Mind map of thesis' structure.

1.4 RELATED PUBLICATIONS

The Table 1 shows the list of publications produced during author's doctoral research.

Table 1 – List of developed papers during the doctoral.

Title	**Type**	**Status**	**Publisher**	**Related Chapter**
An Automated Methodology for Events Classification in Power Plants Based on DFR Data and Symmetrical Components	Conference	Published	Sociedade Brasileira de Automática	-
Event Classification in Microgrids using Harmonic Synchrophasors.	Conference	Published	Sociedade Brasileira de Automática	5
Usage of Harmonic Synchrophasors for High Impedance Fault Classification in Microgrids	Journal	Published	Journal of Control, Automation and Electrical Systems	5
High Impedance Fault Classification in Microgrids using a Transformer-Based Model with Time Series Harmonic Synchrophasors Under Data Quality Issues	Journal	Accepted	Neural Computing and Applications	6

2 FUNDAMENTALS

This chapter gathers some concepts that mark off the scope of the whole research. The considered boundaries in the energy systems are defined with the concept of the microgrids. The proposed approach is fully related to the concept of synchrophasors, therefore, its definition is addressed here. The aspects that guide the classification of events are also discussed, and the event of interest in this work is addressed in a greater level of detail.

2.1 ACTIVE DISTRIBUTION SYSTEMS AND MICROGRIDS

The rise in global energy demand is closely tied to the construction and growth of microgrids. Because their principal source is the usage of fossil fuels, the centralized models that propelled the globe until the mid-twentieth century would not be able to meet the new world energy standards on their own. The social-environmental impact of new major hydropower plants has also become an impediment from the standpoint of hydroelectric plants. In recent decades, several attempts have been made to create and improve renewable energy-based energy sources (HANNAH RITCHIE; ROSADO, 2020). It is shown in Figure 2[1], how rapidly the renewable energy is globally evolving along the years, as well as, the technologies look most promising in transforming the global energy mix.

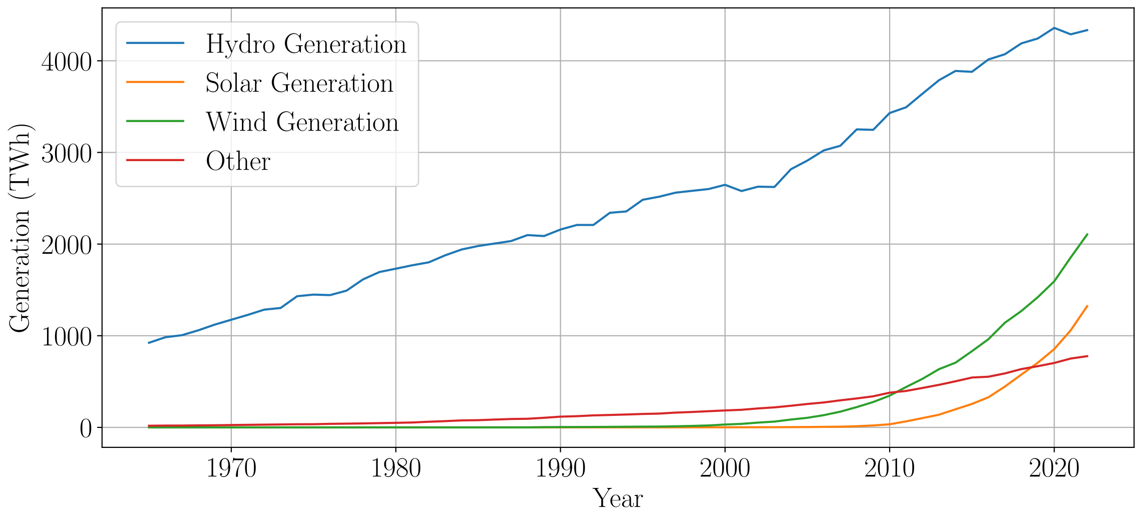


Figure 2 – Global renewable energy growth.

It took about 20 years for hydroelectric sources to double their generation potential and increase their capacity by about 1000 TWh; similarly, it took about 15 years for

[1] Extracted from: https://ourworldindata.org/renewable-energy

wind sources to reach the same generation level; and, around 20 years, solar sources had already reached the same generation level. This demonstrates how the last few years have been a breeding ground for new technologies that have allowed for a large increase in the percentage of new renewable sources in the global energy matrix. Furthermore, according to Stefanidou-voziki et al. (2022), renewable energy sources will account for roughly 80% of global energy output by 2050, up from around 15% in mid-2010. Therefore, new solutions such as distributed generation, renewable energy resources based microgrids, and energy storage systems emerged in recent decades as feasible solutions.

As mentioned before, with the growth of distribution generation sources, the concept of the microgrids takes place in the modern structure of energy networks. A microgrid can be defined as an active network of power distribution. It is usually a small size system, of low (LV) or medium voltage (MV), composed by one or more generation system, commonly based on renewable energy sources, and loads spread along its extension, as well as energy storage devices. Operationally speaking, present-day microgrids can run connected to, or isolated from a main grid, and the connection point to the main grid is called Point of Common Coupling, or PCC. Microgrids must be controllable as a single and independent system that is connected to the main grid. This control is carried out via a management center, which ensures that the microgrid operation is optimized.

There are numerous factors that make the development of microgrids appealing from a technical, economic, and environmental standpoint, the most important of which are: the improvement in the quality and reliability of energy supply due to the decentralization of the system; the best balance between load and generation; the reduction of energy transmission losses; the availability of electrical energy for hard-to-reach areas; the decrease in expenses with the expansion of generation, transmission, and distribution systems; the improvement in the quality and reliability of energy supply due to the decentralization of the system. Despite numerous positives that support their development, the microgrids face a number of challenges that remain as disadvantages or impossibilities, such as the high cost of insertion and installation of energy generation sources; the existence of technical difficulties in relation to aspects of control and, in particular, protection of microgrids; the lack of regulatory standards for operation; and the lack of funding (CHOWDHURY; CROSSLEY, 2009; SHAHZAD et al., 2023).

Of all the technical aspects that pertain to microgrids, for the scope of this research, harmonic content is the most popular. Harmonics pose a threat to the power grid's ability to operate reliably and consistently if no precautions are taken. As the vast majority of distributed generation sources have the connection to the microgrid via frequency inverters as a common element, basically every generation source is considered a harmonic source. Although numerous efforts have been made to mitigate such problem (SEN; KUMAR, 2018; ELMETWALY; ELDESOUKY; SALLAM, 2020), the existence of such components is a premise for the functioning of such architectures, therefore, any and all techniques of operation, protection, and control of microgrids that use information on the harmonic content of the signals of interest, must be able to operate under such conditions.

The advent of microgrids marks a significant milestone in modern power systems, offering a dynamic and sustainable approach to energy distribution. Their capacity to seamlessly integrate renewable resources and operate autonomously or in conjunction with the main grid underscores their potential as the future of electricity networks. However, the intricacies inherent to microgrid operation necessitate a nuanced understanding and vigilant monitoring. Ongoing research endeavors, focused on refining monitoring methodologies and event classification, are paramount in ensuring the reliable and resilient operation of microgrids. By addressing these challenges, we pave the way for a more robust and adaptive energy landscape, poised to meet the emerging needs of the society.

2.2 HARMONIC SYNCHROPHASOR ESTIMATION

For a long time, the concern in phase angles of voltage phasors is present in the routine of the power system engineers, taking into consideration that the power flow in a section of a system may be characterized very intimately proportional to the sine of the angle difference between voltages at the terminals. The first measures of transmission lines voltage angles differences exploring some sort of clock synchronization were reported in early of the 80s by (MISSOUT, 1981; BONANOMI, 1981), and the present era of phasor measurement technology has begun with research into transmission line computer relaying.

The fundamental of phasor measurements is the phasor estimation, which consists in represents a time-domain signal, into its polar form, i.e, with magnitude and angle. Phasor estimation can be seen as a signal processing technique to synchrophasor measurements and one of the most common manners to perform this representation is through the Discrete Fourier Transform (DFT), which is a method of calculating the Fourier transform of a few samples taken from an input signal $x(t)$. The Fourier transform is calculated at discrete steps in the frequency domain, just as the input signal is sampled at discrete instants in the time domain (PHADKE; THORP, 2008).

Considering a sinusoidal signal $x(t)$ with frequency f that can be represented by a Fourier series, and using the relationship of the Fourier series coefficients with the DFT, the phasor representation of the i_{th} harmonic components of $x(t)$ is given by:

$$X^i = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} x(n\Delta T) e^{-\frac{j2\pi in}{N}}. \tag{1}$$

In Eq. 1, N is the number of samples that represents $x(t)$ by a fixed sampling interval ΔT and $2\pi/N = \theta$ is the sampling angle measured in terms of the period of the fundamental frequency component. So, using the notation $x(n\Delta T) = x_n$, the i_{th} estimated phasor is:

$$X^i = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} x_n \big[\cos(in\theta) - j\sin(in\theta)\big]. \tag{2}$$

In Eq. 2, the fundamental frequency phasor is obtained when $i = 1$, and, all the desired harmonic phasors are obtained when i is changed accordingly. By splitting this into sums of sines and cosines, the phasor expression becomes:

$$X_i = X_{ic} - jX_{is}. \tag{3}$$

Once phasor calculation is a continuous process which will update the phasor estimate as new samples are acquired, it is necessary to apply some algorithm to perform the updates. Among the possible techniques, the *recursive algorithm* shows great computation efficiency, and it is usually the choice in many applications. The recursive algorithm consists of applying the DFT over one cycle of the fundamental frequency, Eq. 4, and then correcting it recursively for every additional sample as in Eq. 5. This technique allows to tracking the changes in the phasor over time, providing valuable insights into the behavior of the signal. Considering that there is no frequency variation, DFT provides exact phasors.

$$\begin{aligned} X_k^i &= \frac{\sqrt{2}}{N} \sum_{n=k}^{k+N-1} x_n e^{-jni\theta}, & (4)\\ X_{k+1}^i &= e^{-ji\theta} X_k^i + \frac{\sqrt{2}}{N}\left(x_{N+k} - x_k\right) e^{-jki\theta}. & (5)\end{aligned}$$

where x_k is the k_{th} sample of the signal, X_k is the k_{th} estimated phasor, N is the number of samples per cycle of fundamental frequency, $\theta = 2\pi/N$ is the sampling angle, k is the sample index and i is the harmonic index ($i = 1$ means fundamental frequency).

The visualization of the phasor estimation process of a signal $x(t)$ is presented in Figure 3. A synthetic sinusoidal signal, with 50 Hz of frequency, defined by Eq. 6 is considered and the harmonic content up to third order is estimated, where the magnitude $|X_i|$ is on a generic measurement unit, and the angle $\angle X_i$ is in degrees.

$$x(t) = \cos(2\pi ft + \pi/9) + 2\cos(4\pi ft + 9\pi/2) + 3.5\cos(6\pi ft + 9\pi/3). \tag{6}$$

Once DFT technique is widely embedded to commercial PMUs (PHADKE; THORP, 2008; HOJABRI et al., 2019), therefore, the harmonic phasor estimation plays a crucial role in modern electric system monitoring. It offers real-time and precise information on voltage and current waveforms across the grid, majorly because these devices are build-up based on Global Positioning System (GPS) technology enabling operators to quickly identify irregularities and maintain grid stability in a dynamic environment. Therefore, synchrophasor technology enhances situational awareness, allowing for timely and informed decision-making. By providing a comprehensive view of the grid's dynamic behavior,

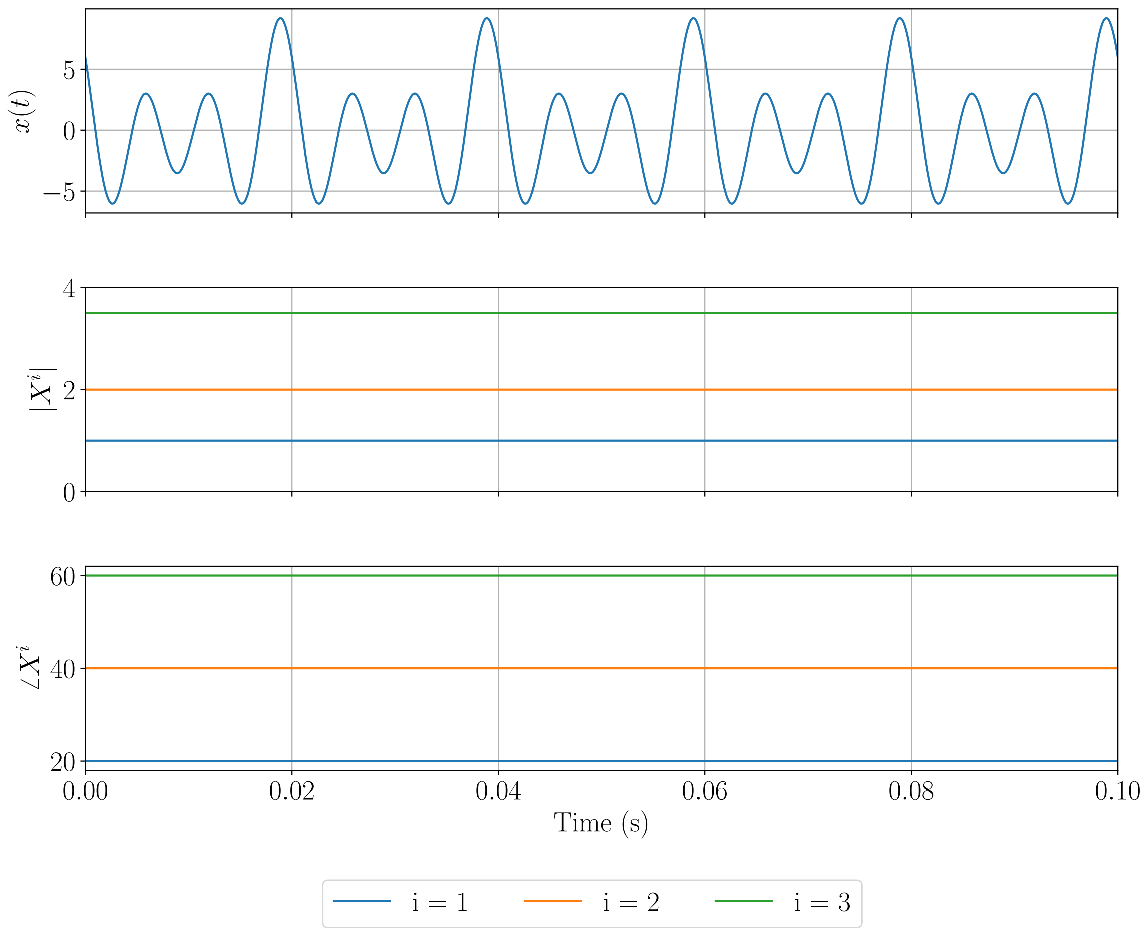


Figure 3 – Example of phasor estimation.

synchronized measurements are instrumental in preventing cascading failures, optimizing energy generation and distribution, and ultimately strengthening the reliability and resilience of electrical networks.

2.3 EVENT CLASSIFICATION

In the early stages of electrical power systems, protection systems heavily relied on manual interventions where engineers conducted visual inspections and manually operated switches to detect faults. The advent of electromechanical relays in the early to mid-20th century marked a significant milestone. These devices autonomously identified abnormal conditions like overcurrent or voltage imbalances, markedly expediting response times and enhancing system reliability. The subsequent introduction of microprocessor-based relays in the late 20th century revolutionized event classification. With their ability to simultaneously analyze multiple parameters, these relays brought a new level of precision to the detection and classification of events. The integration of communication networks into power systems further accelerated this evolution. Real-time monitoring and control capabilities were enhanced, enabling seamless information exchange between devices and central control centers, thereby transforming the coordination and response capabilities of protection systems (AZEROUAL et al., 2022).

In the domain of distribution networks, grid resilience and reliability are paramount, particularly in the face of unforeseen events such as extreme weather conditions or equipment failures. Swift response to faults, facilitated by event classification, minimizes disruptions and bolsters grid resilience. Furthermore, the surge in renewable energy integration underscores the critical role of event classification. With the intermittent nature of solar and wind generation, accurate classification becomes essential in stabilizing voltage and frequency levels, ensuring seamless integration into the grid (CHANG et al., 2023). Event classification acts as a safety net, promptly isolating faults and averting the spread of disturbances, thus mitigating the potential for widespread outages and equipment damage. Additionally, on an economic front, accurate event classification safeguards costly equipment, prolonging the lifespan of critical assets, and ultimately reducing maintenance expenditures (FAZAL et al., 2023).

As we step into the era of active distribution networks, characterized by the integration of diverse Distributed Energy Resources (DERs) and advanced monitoring systems, event classification assumes even greater prominence. It is paramount in the efficient management of bidirectional power flows, a hallmark of these modern networks. This management is critical for maintaining grid stability and ensuring the seamless incorporation of DERs. Moreover, event classification enhances situational awareness, providing operators with a comprehensive view of grid conditions. This, in turn, empowers them to take proactive measures, further fortifying grid resilience (ATENCIA-DE LA OSSA; OROZCO-HENAO; MARÍN-QUINTERO, 2023). Additionally, in the context of demand response programs and grid flexibility initiatives, accurate event classification is pivotal. It identifies opportunities for load shedding or shifting, contributing significantly to overall grid optimization. In this dynamic landscape, characterized by rapid technological advancements, event classification stands as a linchpin in the continuous evolution and effective operation of electric power systems (HAMANAH et al., 2023).

The evolution of event classification methods in electric power systems has seen a remarkable transition from traditional rule-based approaches to sophisticated data-driven methods, particularly with the advent of machine learning models. These methods are instrumental in ensuring grid stability and reliability, and they can be broadly categorized into two primary approaches: feature-based classification and time series-based classification. Feature-based event classification involves extracting specific attributes or parameters from the data related to an event, which are then used as input to a classification algorithm. This approach is efficient for cases where distinct features directly correlate with event types and often involves lower computational complexity compared to time series analysis. However, it relies on the assumption that relevant features have been appropriately identified and may struggle with events that are complex and dynamically evolving, not easily characterized by predefined features. In contrast, time series-based event classification delves into the temporal sequence of data, considering the behavior and patterns over time. This approach is particularly adept at capturing temporal dependencies and patterns, making it suitable for dynamic events. However, it may be computationally intensive, especially for extended or high-frequency time series data, and requires meticulous preprocessing and the selection of appropriate time series analysis techniques. Hybrid approaches, combining features and time series analysis, are also employed to provide a comprehensive classification solution, with the choice between these methods hinging on factors such as the nature of events, available data, computational resources, and interpretability requirements (OLIVEIRA; BOLLEN, 2023; OUBRAHIM et al., 2023).

In summary, event classification holds significant importance in modern electric networks. It ensures a reliable power supply, especially with the increasing use of renewable energy sources and advanced grid technologies. Ongoing research in this area is crucial for refining event classification methods and addressing emerging challenges. This continuous effort will play a vital role in sustaining the future of electricity.

2.4 HIGH IMPEDANCE FAULTS

In active distribution networks, particularly in microgrid configurations, events including low impedance faults, capacitor bank switching, transformer energization, inverters switching, and loads switching are commonplace. Low impedance faults, originating from short circuits or insulation breakdowns, can disrupt the normal flow of current and voltage, potentially resulting in equipment damage or introducing harmonic distortions into the system. Capacitor bank switching, a routine operation in distribution systems, may introduce transients and harmonic components due to rapid changes in reactive power. Transformer energization, while essential for voltage conversion, can lead to inrush currents and subsequent harmonic content in the system. Furthermore, the operation of inverters in microgrid environments, especially during transitions between grid-connected and disconnected modes, can introduce voltage and current distortions, influencing power quality. Load switching events, common in dynamic microgrid environments, can cause sudden changes in current and voltage levels, and also, potentially leading to harmonics generation (FUCHS; MASOUM, 2008).

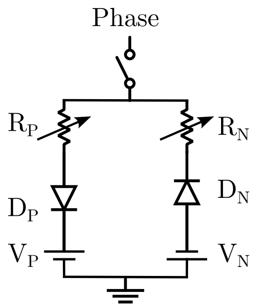
Among the possible events in microgrids, the High Impedance Faults (HIFs) stand out due to the difficulty in being detected. These faults are usually caused when power conductors break and touch the ground or when some object (e.g., vegetation) comes into contact with the conductors. Low-level fault currents characterize such events due to high grounding impedance (GOMES et al., 2019). Different from most power system faults, HIFs firstly menace human safety and could also lead to environmental issue before endanger electrical equipment. Consequently, once that protection devices are unable to detect these class of disturbance, it is an important task to detect it using suitable approaches. The aspects that influence the intensity of HIFs are multiple, but majorly, ground surface material, surface humidity, weather conditions, feeder configuration and voltage levels. For example, in one of the pioneer's works that dealt about evaluation of HIFs, Emanuel et al. (1990) showed that a higher surface humidity leads to higher magnitudes of fault current. Furthermore, once that HIFs can occur on various arrangements of environment, each arrange results in different voltage-current profiles, making the identification of such events, that correspond to up to 20% of the faults in distribution networks, even challenger (THERON; PAL; VARGHESE, 2018).

Emanuel's model is based on laboratory measurements and theoretical components. It is shown in Figure 4a how the arc is modeled using two DC sources, connected as anti-paralleled by two diodes. In order to reach more realistic approximations of the phenomenon of non-linearity, time-varying components can be considered (EMANUEL et al., 1990). In this model, when the instantaneous phase voltage is greater than V_P the positive cycle of fault current is developed and the fault current flows towards the ground. On the other hand, when V_N is greater than the phase voltage, the negative cycle of fault current flows reversely. This aspect guarantees both the asymmetrical characteristics and the randomness of the phenomenon (CUI; EL-ARROUDI; WENG, 2019). The switch on Figure 4a is responsible for the inception of the event in the healthy phase as well as for allowing the possibility to consider the intermittence in the disturbance, i.e., the HIF can be ceased and initiated as many times as possible, emulating multiple touches of the faulty conductor with the high impedance path.

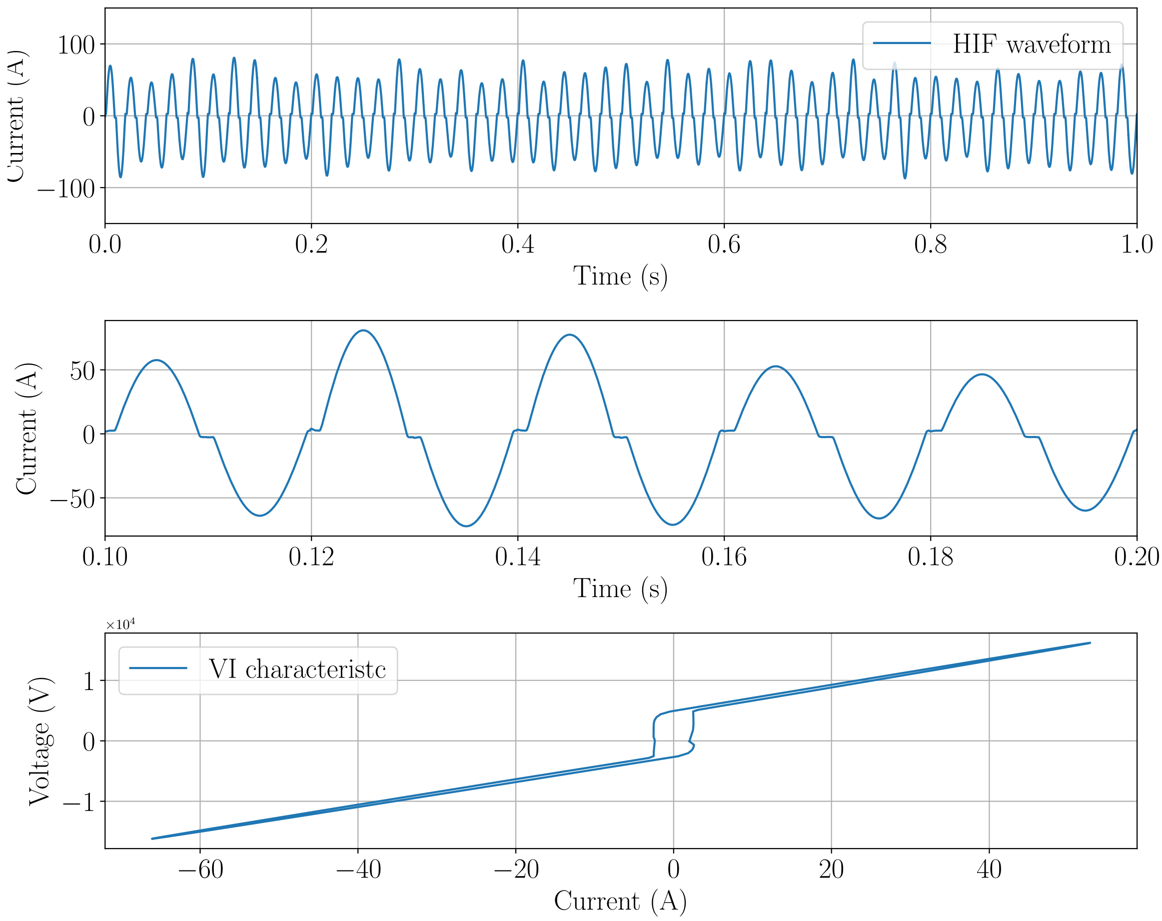
The most significant characteristic of HIFs is that they often occurr with the presence of electric arcs. These arcs appear once the magnitude of voltage of the surface contacted conductor surpasses the break-down voltage of the surrounding. Besides the arc presence and low magnitude currents, there are several other physical characteristics, including the intermittence of the arc, the asymmetry in current waveform, the buildup, and shoulder current, non-stationary current, randomness, non-linearity, low frequency components in voltage waveform and both low and high frequency components in current waveform (GHADERI; GINN; MOHAMMADPOUR, 2017).

As can be seen in Figure 4b, the current signal of a HIF has a certain level of discontinuity and randomness. Yet, the V-I relationship shows the outstanding non-linearity of fault impedance. The waveforms in Figure 4b were obtained through a parametric change in model proposed by (CUI; EL-ARROUDI; WENG, 2019), later explained in Section 4.

In modern electric systems, dealing with harmonic content is crucial. Events like capacitor bank switching, transformer energization, and inverter switching introduce harmonics, making HIF detection and classification more challenging in front of other harmonic-manifesting events. Recognizing the significance of identifying HIFs in such environment is vital. Accurately detecting these faults not only protects the network but also ensures the



(a) HIF model scheme.



(b) Typical behavior of HIF waveforms.

Figure 4 – Model and typical waverforms of a HIF.

overall system's reliability and safety. Therefore, using advanced monitoring techniques and protective measures is essential in managing the risks associated with HIFs in today's electric grids.

3 RELATED WORKS

This chapter presents a literature review that seeks to establish a chain of ideas that relate the application of data from PMUs, that is, harmonic synchrophasors, in the task of events classification in electrical energy systems, emphasizing a critical class of events, which are high impedance faults.

3.1 SYNCHROPHASOR BASED APPLICATIONS IN EVENT CLASSIFICATION

Advancements in power system monitoring, protection and control are strict related to the development of PMUs' technology. These devices are widely present in transmission system about four decades. In recent years, PMUs are also gradually gaining ground at distribution level, aiming to enhance the characterization of events on these systems, albeit the deployment of these systems is still in an embryonic stage, especially because of its high investment cost. However, the advancements in measuring systems is one of the keys for developing of the emerging active distribution systems. These emerging systems are envisioned to include a deeper penetration of distributed energy resources and active participation of consumers as well as smart appliances and electrical vehicles are also representative components in the future's systems. When compared to transmission systems, voltage phasor variations at distribution system buses are significantly smaller. Combined with the involvement of multiple sources in the system, these small-scale variations are essential for better control and protection and their monitoring and detection are extremely relevant (LIU, Yikui; WU, L.; LI, J., 2020; JOSHI; VERMA, 2021).

The most prominent causes that have been easing PMUs installations at distribution system are the technical and technological advancements, like higher measurement rates, but mostly important, the decrease of capital cost when compared to transmission system's PMUs. Another cause that forces the application of PMUs at distribution level is the growing needs in system's situation awareness, once the systems are experimenting a rapid growth of distribution energy resources, this fact increases the requests in power quality, reliability, and resiliency. Several works summarize relevant applications for pilot distribution systems containing PMUs, indicating the inevitable tendency to implement these equipment in distribution systems (LIU, Yikui; WU, L.; LI, J., 2020).

An example of using PMUs to enhance event detection and classification is proposed by (MIRANDA et al., 2019), where the authors use frequency information to detect events in a real scenario, using data from Brazil wide-area monitoring project called Medfasee. The approach uses images from frequency data and through a convolutional neural network, events such generator tripping, line tripping, load disconnection and inter-area oscillations are detected and classified. The authors suggest that in interconnected systems, where multiple events can occur simultaneously, the usage of individualized algorithms in each PMU or in groups of them. Also, when there is a history of events in a system, the authors mention the possibility of using transfer of learning to improve the accuracy of the classification. Another perspective of using PMU data is shown in (SHAW; JENA, 2020). The scheme is based on monitoring the frequency through a WAMS. Raw frequency as well as its rate of change and voltage phase angle difference across some points of the system are used as features. By comparing the features with standardized thresholds, events like oscillatory events, step change events (like generation trip), impulse events (class that represents sudden surge or drop of power or frequency in the system), and islanding events are detected and classified.

A holistic framework for novel machine learning based applications analyzing both historical and online synchrophasor data streams is proposed in (KUMMEROW et al., 2020). To detect disturbances automatically, various methods such as dimension reduction, anomaly detection, and time series classification are applied. The framework, which can be integrated into existing control centers, provides useful information for the automated online detection and classification of critical system states, as well as the activation of appropriate countermeasures to ensure secure system operation. Different application modules for efficient data processing, *post-mortem* analysis, and online recognition make up the platform. In comparison to existing approaches, this framework enables a comprehensive situational awareness to recognize known disturbance events that are part of simulated contingencies, as well as new disturbance events that are discovered after the fact through efficient analysis of large historical measurements.

In order to enhance event detection and characterization in power systems using PMU data, (PAVLOVSKI et al., 2021) perform a comprehensive investigation using convolutional neural networks. An advanced convolutional neural network model, focused on univariate measurements, introduced ways to process such data, identifying distinct components, and detecting events without the need for exhaustive system specifics. PMU measurements of current, voltage and frequency are broken down into fixed-length intervals, feeding the classification models. The authors key findings reveal that multi-channel hierarchical convolutional networks were superior in performance when compared to single-channel ones, benefiting significantly from domain expert involvement. The model's accuracy particularly improved when trained with expert-reviewed data for extended periods.

In (YUAN; WANG, Z.; WANG, Y., 2022), a recent novel for processing and classifying events in power systems using synchrophasor readings from 440 PMUs is presented. It employs graph-specialized neural networks combined with autoencoders for unsupervised interaction learning and event classification. Through pre-processing and a Markov-based feature reconstruction, the method converts time-series PMU data into image-like formats. The core of this innovative approach involves inferring data-driven interactive relationships among different PMUs and seamlessly integrating this knowledge into an autoencoder architecture. This not only streamlines the optimization of both graph inference and the classification model but also incorporates a dilated inception model designed to autonomously capture multi-scale event features, minimizing parameter requirements. When tested on a comprehensive real-world dataset, the method exhibited remarkable reliability in interaction inference and demonstrated superior classification accuracy over established methods.

Another method used for detecting and classifying events in power systems is proposed by (AALAM; SHUBHANGA, 2023). It relies on non-training based signal processing techniques and its main aim is to create a tool that accurately pinpoints, locates, and categorizes events using data from multiple PMUs. The method utilizes the wavelet transform and the standard deviation technique. The wavelet transform calculates detail coefficients, which are then used to determine the wavelet energy parameter, reflecting the characteristics of these coefficients, serving as an indicator for event detection. The standard deviation method identifies irregularities in normally distributed signals. By assessing the standard deviation of signals like phase angle difference and rate of change of frequency, events can be identified by comparing these values against predefined thresholds. The efficacy of these techniques is showcased using simulated and real data. Furthermore, an event localization algorithm is introduced, which classifies disturbances as either local or widespread events, based on the number of PMUs involved in the event detection stage. Additionally, a method to identify a loss-of-synchronism condition using phase angle difference signals across transmission lines is proposed as part of the event detection process.

Lots of other researches have been conducted with the usage of PMUs at distribution level. Most part of the investigations are based on the idea that the system is a microgrid, i.e, the environment counts with multiple power sources and, also, it has basically the capability to operates both connected or unconnected to a major grid. The mainly applications addressed are: *event detection and classification*, *voltage and frequency monitoring*, *islanding detection* and *low impedance and high impedance faults detection and location* (BHATTARAI et al., 2019; HOJABRI et al., 2019; SHAHSAVARI et al., 2019; LIU, Yikui; WU, L.; LI, J., 2020; GRANDO; LAZZARETTI; MORETO, 2021; JOSHI; VERMA, 2021; EHSANI; AMINIFAR; MOHSENIAN-RAD, 2022; FUENTES-VELAZQUEZ et al., 2022; NGUYEN et al., 2023; WANG, Shaorui et al., 2023).

PMUs have significantly advanced event classification in electric systems, offering both a macro and micro perspectives of the systems' operation. Their ability to provide synchronized measurements across vast regions ensures real-time monitoring and rapid event classification, capturing even the slightest discrepancies in voltage and current phasors. Moreover, PMUs excel in providing complex harmonic patterns, which becomes particularly vital in microgrids where inverter connections introduce foundational harmonic disturbances.This comprehensive view, encompassing both the synchronized overview and the detailed harmonic analysis, ensures that events are classified with utmost accuracy, recognizing even the subtle nuances that might indicate looming issues in the system. When it comes to events that are traditionally challenging to detect, such as high impedance faults, the high-resolution data from PMUs becomes invaluable. These faults, due to their subtle nature, often escape conventional detection systems. However, the granularity and synchronism of PMU data can identify and highlight these elusive events, thereby greatly enhancing the reliability and safety of electric systems. Recognizing the significance of this subject, the subsequent section will delve into a detailed review of techniques that are centered around PMU data and its efficacy in identifying high impedance faults.

3.2 HIGH IMPEDANCE FAULT - DETECTION AND DISCRIMINATION

HIF detection and classification have long posed challenges to the electric power industry, often stretching back several decades. These types of faults are notorious for their elusive nature, presenting themselves with characteristics that don't neatly fit the profile of traditional short-circuit faults. Traditional protection systems and methods have often struggled to detect them due to their low current characteristics and the myriad of conditions under which they can occur, leading to potential safety hazards and system vulnerabilities. The development of PMUs technology, with their ability to capture high-resolution, synchronized data across the grid, offers a fresh perspective on this age-old problem. Their rich-informative data, combined with advanced analytics, can uncover the subtle signatures of high impedance faults, thus enabling preciser classification. As a result, the integration of PMU data into detection systems is progressively bridging the gap in the timely and accurate identification of these problematic faults (LOPES, G. N. et al., 2023).

In the last decade, researchers have presented numerous HIFs identification algorithms using a combination of computational intelligence methods, along with the proper signal processing techniques, most of them based on a combination of time and frequency analysis for identification of faulty and non-faulty events in the system (MOHAMED, 2013; ALI et al., 2014; ROUTRAY; MISHRA; ROUT, 2015; SOHEILI et al., 2016; SEKAR; MOHANTY, 2017; SILVA et al., 2018; LIMA; BRITO; SOUZA, 2019; SARWAR et al., 2020; RAI, K. et al., 2021; GAO, J. et al., 2022; SOLANKEE; RAI, A.; KIRAR, 2023).

A frequency domain approach is proposed by (SOHEILI; SADEH; BAKHSHI, 2018) that employs the relative relation between the third, fifth, and seventh current harmonic measured at the substation's level. The methodology was tested in the IEEE 13-bus system and evaluated with real data and presented promising results, differentiating HIFs from non-faulty switching events in the presence of non-linear loads in the system. Features are obtained after a fast Fourier transformation in the substation current signal three-phase by comparing the ration between harmonic components for each class, then, by defining correct thresholds, a comparison is carried out, and the event is detected. A sensitivity analysis is driven to evaluate the method in front of variations in fault resistance, location, load switching, presence of spikes in the original signal as well as noise. For both set-ups, the algorithm has shown immunity and stability in the outcome.

A feature extraction method based on discrete wavelet transform is proposed by (SILVA et al., 2018), that is combined with an evolving neural network to recognize patterns of electrical current data. When compared with non-evolving techniques, such as multi-layer networks, probabilistic neural networks and support vector machines, an evolving approach showed to be quite appropriate to HIF detection, since this class of event is known as a time-varying problem. The behavior of the system under fault conditions becomes evident and clearly observed from the second order detail coefficient. One HIF model was used to compose the training data set while a second HIF model was used to validate the method. When comparing the evolving network with non-evolving approaches, better performance was attained because of evolving layer that is able to change on fly, thus adding new neurons that imply in new prototypes of fault patterns, making the retraining no necessary.

Another discrete wavelet transform method was developed by (VEERASAMY et al., 2018). Authors propose an adaptive neuro-fuzzy that considers extracted features based on the standard deviation values of the detail and approximation coefficients up to fifth order from three-phases fault current signal. They simulated a radial 13.8 kV distribution power network composed by 5 feeders connected to a grid source. Also, the HIFs were simulated based on the simplified (EMANUEL et al., 1990) two-diode model. The authors just evaluated the HIF detection technique in front of symmetrical and unsymmetrical low impedance faults, with fault resistance variation, and normal operation of the system. The method compared both a classic fuzzy inference system and an adaptive neuro-fuzzy inference system and results showed that while a classic fuzzy system reached about 85% of correct discrimination, the adaptive neuro-fuzzy system achieved 100% of accuracy.

A method based on mathematical morphology that used current signals observed from the distribution feeder to detect HIFs is suggested by (KAVASKAR; MOHANTY, 2019). Mathematical morphology can be defined as a theory of spatial structures based on set theory and integral geometry. It is a non-linear and time domain signal processing technique. Based on some grouped kernel functions, the features are extracted as the result of a filtering process of the original signal, where the noise is removed, and faulty signatures are preserved. When compared with other transient events, the HIFs present some differences that allow the discrimination with a simple rule based algorithm. Several scenarios of fault location, fault inception time, and pre-fault loading were investigated to verify the performance of the suggested technique. The proposed method's findings are good, rapid to detect, secure, and dependable under a variety of transient settings.

The authors in (CHAKRABORTY; DAS, 2019) proposed a novel application based on smart meters for HIF detection in distributions systems that uses the amount of even harmonics present in the measured voltage in several points of the system. HIFs are known by containing both even, odd and inter-harmonics. Moreover, today's distribution system contains a large variety of power electronic components that generate a significant amount of odd harmonic components during steady state operation but usually do not produce harmonic components during steady state. Because of these characteristics, the presence, or absence of specific harmonic components can be used for differentiating HIF from steady state scenarios. Based on that, each smart meter calculates an index to measure the even harmonic components present in the voltage waveform and HIF is detected if the index calculated by any meter crosses a threshold for a specific time. The proposed present a satisfactory performance in presence of HIFs, voltage sag-swells, capacitor and load switching, transformer and load energization, power electronic loads, arc furnace loads and distributed generation. The proposed method has been implemented on a commercial energy meter and showed its viability.

A two-level artificial neural network method responsible for identifying, locating HIFs as well as discriminate them from low impedance faults, in a distribution network is developed by (LEDESMA et al., 2020). The method is based on synchronized measurements of three-phase currents, obtained from several points in the system. The main objective of the proposed method is locating and identifying the phase and fault by areas of HIFs. The authors introduce a concept of observable areas between meters, so, the evaluated system is divided into plenty of these areas, then, an artificial neural network is trained with faulty and non-faulty data from each of these areas. The features are obtained from magnitude and angle and symmetrical components of the current signal. The method was evaluated in different scenarios, as load variation, HIF resistance variation, the presence of distributed generation and system reconfiguration. The robustness of the proposed approach was tested, and a good performance was acquired, for all the tested conditions, achieving 100% accuracy in most cases of identification step and error lower than 1% in the location step.

In the research of (WANG, Shiyuan; DEHGHANIAN, 2020), an artificial intelligence solution based on an improved HIF model and a modified wavelet transform for feature extraction, also, a compact convolutional neural network-based event detection technique is tested with extracted features. Phasor records are obtained from the more upstream terminal from a radial 13.8 kV designed distribution system. Features are extracted from current signals using continuous wavelet transform considering some specific modifications in order to obtain more information extracted and waveform feature redundancy in the scalograms. Essentially, the neural network classifies the scalograms obtained through the pseudo-continuous wavelet transform based on each class of the following events: HIFs, load changes and normal operation. The method was tested in noisy scenarios and showed a stable performance with an overall accuracy was about 99.9%. The authors defend that the proposed approach could be embedded within existing PMUs or other intelligent electronic devices that are capable to record and process power waveforms.

A method for detection of HIFs in solar photovoltaic integrated power system using long short-term memory recurrent neural network is developed by (VEERASAMY et al., 2021). Tests were conducted in a 25 kV IEEE 13-buses system, considering a 300 kW solar photovoltaic plant. Substation three-phase currents were measured with 20 kHz sample rate and the energy of the approximation and detail coefficients from discrete wavelet transform were used as features. The events were considered as normal and non-normal. The overall accuracy of the classifier was about 92%. The spikes observed in the wavelet coefficients caused by the transient of each event were the discriminant characteristics that allowed the segregation of classes. The adopted classifier based on recurrent neural network was compared with other classifier, as k-nearest neighbors algorithm, decision tree, support vector machine and Naive Bayes classifier, and showed the best performance indices.

High-order statistics are used by (SOUSA CARVALHO et al., 2021) for HIF detection and classification based on extracting relevant features from the signals measured in a substation. The considered events are associated with their high-order statistics based - cumulants. In probability theory and statistics, the cumulants of a probability distribution are a set of quantities that provide an alternative to the moments of the distribution, for example, the first cumulant is the mean, the second is the variance. The results explore the potential representation of the features of the events. After the features' determination, Fisher's discriminant ratio is applied to select the cumulants with the greatest potential to perform the separation between classes. At the end, a multi-layer perceptron artificial neural network with three inputs that correspond to the best scored high-order cumulants is trained in order to recognize the patterns of each event type. When compared with other approaches available, the proposed method showed relevant benefits, as demanding low sampling frequency and a lower number of features needed. The higher-order statistics approach achieved around 99.75% of global accuracy and proved to be a promising tool for the classification of disturbances in distribution networks.

The authors in (BHATNAGAR; YADAV; SWETAPADMA, 2022) proposed a combination of discrete wavelet transform and fuzzy inference system for HIF detection and classification. A modified IEEE 13-bus system was adopted to validate the proposed scheme. The method considers current signals at substation's level that are processed using discrete wavelet transform to obtain appropriate input features. The evaluation of the method considers low and high impedance faults, various switching events as well as distributed generation penetration and evolving faults and takes into account the effect of noise. Features are obtained by the standard deviation of the first level approximate coefficients. The fuzzy inference system was evaluated through variations in the following parameters: distributed generation (solar and wind), inception angle, fault resistance, evolving faults and noisy signals and showed to be robust to all these variations. At the end, the method was evaluated with IEEE 33-bus system and become efficient in detecting faults in more complex distribution systems. The approach is 100% accurate for both events considered, showing that fuzzy based techniques are promising for event classification because of their capability to deal with uncertain data.

The proposed work in (LOPES, G. et al., 2022), consists of an approach for the detection of HIFs based on three-phase measurements of current in the substation and transformation of the current signal in order to obtain the harmonic spectrum. The method identifies HIFs in other transient scenarios of short duration: switching capacitor banks, linear loads and side branches, as well as transients with longer duration: energizing transformers and connecting non-linear loads. Since the fault current of the HIFs, regardless of position, has relevant harmonic content at the fundamental frequency, second and seventh harmonics, an energy profile is obtained for each of these components. The energy is calculated between two consecutive data windows of the signal and compared to a threshold for detecting transients in the signal. If transients are detected, the algorithm can distinguish a HIF from other events, otherwise the threshold is updated. The method is evaluated both with simulated data (IEEE 34-bus system) and with real data (obtained for different types of soil), and the results obtained were quite satisfactory. The overall performance of the method reached an accuracy of 96% and a minimum of 90% (as the noise level increases, the accuracy of the method decreases). It was also found that the farther the fault point is from the measurement point, the lower the accuracy of the method (although it still has a precision greater than 90%). The authors also mention that their method, based on Stockwell transform, does not require a computational effort so superior to the effort of the discrete-time Fourier transform (widely used commercially in protection devices), therefore, the presented approach is promising for real-time applications.

In the recent research developed by (GAO, J.-H. et al., 2023), the authors endeavor into HIF diagnosis have heralded the introduction of a two-phase diagnostic approach, chiefly focusing on fault triggering and fault detection, drawing inspiration from prevalent applications of semantic segmentation in medical analysis and the continuous electrocardiogram wave segmentation. Primarily, the model interprets the transient process of potential fault instances and pinpoints the exact inception moment. To further refine the fault detection, the zero-sequence voltage's long-term data undergoes extraction of salient features: signal envelope and Hilbert marginal spectrum, obtained via the Hilbert-Huang Transform. By transforming these extracted features into image form, HIFs are distinguished according their harmonic distortions and randomness patterns. Essentially, this novel approach transitions the traditional waveform analysis into an image segmentation task, thereby optimizing the accuracy and efficiency of HIF diagnosis. This research not only pushes the boundaries of HIF diagnostics but also paves the way for broader applications of advanced neural networks in power system diagnostics.

Recent advancements in HIFs detection in power systems have showcased a wide range of innovative methodologies. Frequency domain analysis techniques have been leveraged to distinguish them based on harmonic content, underscoring the significance of extracting harmonic spectrum from three-phase current and voltage measurements. Time-frequency transforms have emerged as a robust approach for feature extraction, in the same way, mathematical morphology stands out as a non-linear, time-domain processing tool, proving efficient in differentiating HIFs from other transient events. Other studies have dived deep into the realm of high-order statistics, introducing the power of their indexes in classifying disturbances. On the machine learning front, there's been a tilt towards modern artificial neural networks, which have shown superior performance compared to other classifiers. Collectively, these studies paint a vivid picture of the dynamic and multifaceted efforts dedicated to refining the reliability and precision of HIF detection and classification in modern power systems.

Due to their physical characteristics, HIFs present subtle manifestations in electrical quantities and to observe these aspects are not an easy task. Most of the approaches presented in literature take use of current information to identify the presence or absence of the event. The usage of phasor measures are standing out, mainly boosted by the increased presence PMUs in medium and low-level voltages. The willingness of low cost PMUs in a distribution system, for example, allows multiple points of monitoring, creating a sort of mesh of sensors and, once synchronized measures are available, this arrangement of meters can improve the system's operation. The features used to discriminate HIFs of other faulty and non-faulty events are many, from physical information (like energy of the signal), to non-physical in formation (features based on statistical extraction). Consequently, feature engineering is a promissory path while solving events classification in electrical networks. Lastly, all the effort in properly process the measured data, as well as the suitable extraction of information of the data demand proper and powerful tools to process this information. In this sense, artificial intelligence techniques are consolidated and have been proved to be more and more suitable to deal with large amount of complex data.

3.3 USAGE OF HARMONIC SYNCHROPHASORS FOR HIGH IMPEDANCE FAULT CLASSIFICATION

In modern electric energy systems, particularly in complex environments like smart grids, the study of harmonic behavior has become crucial. As many authors have emphasized over the past few decades, the integration of power inverters and other advanced technologies has led to a significant presence of harmonic contents in these systems. These harmonics, arising from nonlinear voltage-current characteristics, are more than mere disturbances; they are distinctive signatures that offer deep insights into the system's operation. However, accurately classifying events within these harmonic-rich environments poses a formidable challenge. In the intricate landscape of smart grids, where numerous variables and dynamic conditions intersect, distinguishing between normal operational variances and potential issues becomes increasingly complex. This difficulty underscores the vital role of sensors in modern electric systems, once them capture detailed and continuous data streams that are essential for understanding and maintaining correct operation.

In the context of sensors technology, the advent of PMUs has marked a transformative era in electrical system management. By providing real-time, granular data, sensors enable a more nuanced understanding of the system's state, enhancing the ability to detect and classify events accurately amidst the harmonic noise. This improved detection capability is pivotal for the robustness and real applicability of the system, ensuring that the operations are not only efficient but also resilient to potential disruptions. With these advancements, the goal is not just to manage the complexities of harmonics in smart grids but to turn these challenges into opportunities for optimizing system performance and reliability.

In the perspective of using PMU data for event classification at distribution level, most approaches are feature-based methods, i.e., a feature extraction procedure is applied before the classification. The primary concept behind these methods is to capture signal characteristics that identify a certain class of signals. Regrettably, using non-automatic feature extraction methods can lead to a time-consuming process, requiring the involvement of experts to ensure that pertinent information is not lost (SUSTO; CENEDESE; TERZI, 2018). Several recent works focused on event diagnosis are presented in (PARAMO; BRETAS; MEYN, 2023; LIU, Yang et al., 2023; ZHANG, Y. et al., 2020; WANG, Shiyuan; DEHGHANIAN, 2020; WEI et al., 2020; CUI; EL-ARROUDI; WENG, 2019).

Approaches that do not demand pre-processing and feature pre-definition are typically based on time series classification. Most common architectures used in the context of power systems are the Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long-short Term Memory Networks (LSTMs) (SHADI; AMELI; AZAD, 2022; RAI, K. et al., 2021; RAI, P.; LONDHE; RAJ, 2021; ZHANG, Y. et al., 2020; SIROJAN et al., 2018), and such approaches are characterized by an inner feature definition, based on their correspondent architectures. With the recent innovations in deep learning models, specifically with the introduction of attention mechanisms (BAHDANAU; CHO; BENGIO, 2014), new models show superior performance over classic machine learning algorithms. Time Series Transformer Networks (TSTs) (ZERVEAS et al., 2021) have attention layers that allow the model to dynamically learn temporal features of a sequence by focusing on its time steps (VASWANI et al., 2017). Modern breakthroughs in natural language processing (translation, summarizing, sentiment analysis, language generation) and computer vision (image classification, object recognition, image generation) have been made possible by the development of attention mechanisms (GALASSI; LIPPI; TORRONI, 2020; KHAN et al., 2022). In the context of event diagnosis in power systems, methods based on transformer neural architecture are proposed by (THOMAS; SHIHABUDHEEN, 2023; THOMAS et al., 2023).

Models derived from natural language processing, like TST, represent a relatively novel area of exploration in the realm of computational techniques. Given their novelty, the application of these models within the context of electrical systems remains largely uncharted, with only a limited number of studies venturing into this domain. Despite the nascent stage of research in this area, the preliminary results emerging from these studies are notably promising. Particularly, the application of these models to problems involving time series analysis in electrical systems has shown potential. This emerging evidence suggests that these methodologies, traditionally confined to linguistics and text analysis, may offer innovative approaches for interpreting and managing complex time-dependent data in electrical engineering. Such a cross-disciplinary application not only broadens the scope of natural language models but also opens new avenues for enhancing analytical capabilities in electrical system studies (BAHDANAU; CHO; BENGIO, 2014; VASWANI et al., 2017).

The concern to enhance the generality and improve outcomes in HIFs detection follows the scientific community up since at lest three decades. The continuous advancements in computer's performance have been allowed lots of improvements in plenty areas, like simulations, monitoring devices, signal processing and development of artificial intelligence tools. By taking a look in recent researches, it is possible to observe a trend in using data-based techniques through mining and featuring such information. Also, the increasing in available data results in challenges that adhere to big data concepts. Then, it is easily noted that to deal and solve some problems in current energy systems, more and more some knowledge are demanded. In this context, the proposed work intends to process and explore synchronized phasor measurements in order to engineer a group of suitable features that allows to discriminate HIFs from other possible events in the operation of modern active distribution networks.

4 SIMULATION ENVIRONMENT: DATA GENERATION, PROCESSING AND CLASSIFICATION

This chapter brings details about the simulated network used to obtain realistic events data. The first step of the method is based on data generation, that consists in modelling a simulation environment that represents a microgrid, as well as automating the simulation routine by choosing adequate parameters that contemplate a large variability in interest event characteristics. Once all the data is available, they are processed in the context of synchrophasor estimation to obtain useful datasets to event classification tasks that will be presented in the next chapter.

4.1 TEST SYSTEM AND THE BENCHMARK

It is widely recognized that one of the major challenges of the twenty-first century is the massive use of renewable and distributed energy resources worldwide. The availability of methods and techniques that enable their economic, robust, and environmentally responsible integration is critical to the success of this transition. All over the world, industry, universities, and research institutes are engaged to develop these methods and techniques. However, test systems that facilitate the analysis, validation and also the comparison of developed methods and techniques are lacking. In this sense, it is widely accepted the usage of well-known systems in order to develop an environment of transparency and reference. Striving to address this lack of distributed energy systems benchmarks, the Council on Large Electric Systems (in French, Conseil International des Grands Réseaux Electriques - CIGRE), which is a global community committed to the collaborative development and sharing of power system expertise proposed in 2014, the CIGRE Task Force C6.04.02, named "*Benchmark Systems for Network Integration of Renewable and Distributed Energy Resources*", that is a common basis for testing containing both European and North American benchmark network (CIGRE, 2014).

In order to emulate a distribution system in a trustworthy way, it was opted to use the medium voltage distribution network benchmark, with European configuration and as fewer adjustments as possible, i.e, the configuration and parameters proposed by CIGRE were narrowly followed, except by the fact that some renewable distributed generation units were added to the system, in order to broach the new era of energy systems, with high penetration of renewable energy.

The European MV distribution system has a three-phase feeder that can be operated both meshed or radially. The feeder includes numerous laterals at which MV/LV transformers can be connected. The nominal voltage is 20 kV and frequency is 50 Hz. Basically, every node is a load node or a generation node and details about them are presented in the sequence. Normally, efforts are made to balance the various low voltage laterals along the MV branches, but some unbalances are still typically experienced in practice. However, unbalance is not explicitly included in the European benchmark. Maybe one of the most significant changes in the original structure of MV benchmark is the complete usage of overhead lines. In the original configuration, the system has predominantly underground cables, meantime for HIF investigation, it makes more sense to deal with overhead lines, once that most HIFs' characteristics are based on conductor-ground touch. In this sense, the system is adapted to contain overhead lines. Another nuanced modification performed is concerned to the system's grounding. In this sense a grounding transformer at the end of the substation is added to the network. Finally, detailed distributed generation units are inserted in the system: these units are coupled through inverters controlled by maximum power point tracking. The aforementioned system is fully modelled and implemented in Matlab/Simulink environment.

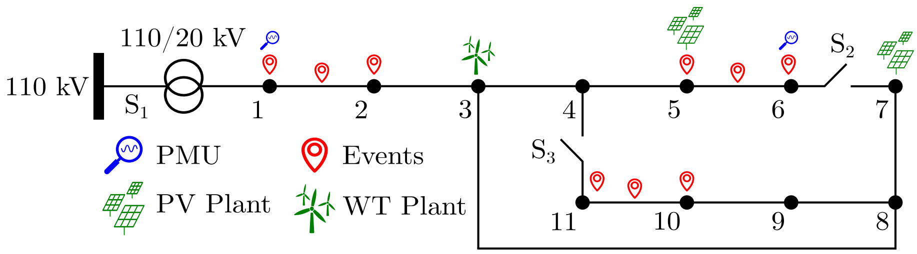


Figure 5 – European MV distribution system.

The European MV distribution network considered to all the following developments is shown in Figure 5. It consists in an 11-nodes system connected to the main grid through a power transformer, as mentioned, the network operates in 20 kV and 50 Hz. The switch S_1 is liable to change the operational mode of the grid (from connected to islanded and vice-versa). Also, switches S_2 and S_3 change the topology of the network, from radial to mesh. All the other relevant details and parameters information needed to reproduce the experiments are extracted from the CIGRE technical brochure and presented in the following.

4.1.1 Main Grid and the Point of Common Coupling

The main grid (110 kV) is modeled as a balanced three-phase voltage source with internal R-L impedance and internally grounded, operating in swing mode. The source is assumed to have a three-phase short-circuit level of 5000 MVA and a short-circuit ratio (resistance-reactance relation) of 0.1. Levels' connection is performed by a 25 MVA coupling transformer (T_C), it has a grounded wye-delta configuration. As mentioned before, in order to create a low-impedance path for fault currents, a 100 MVA grounded wye-delta grounding transformer (T_G) is connected between the coupling transformer and the bus at MV level. Finally, the PCC is illustrated by the switch S_1 in Figure 5, and is modelled as a three-phase circuit breaker that is normally closed to ensure a connected operation.

The parameters of the transformers both from coupling and grounding are shown in Table 2, where sub-indexes 1 and 2 in the resistance and inductance columns, refer to high and low level side, respectively. The other transformers used to connect the distributed generation units to the system as well as some specific loads arrangements are presented in the corresponded subsections.

Table 2 – Parameters of the transformers of the main grid and PCC.

Transformer	Connection	Voltage Level (kV)	Power (MVA)	$R_{1,2}$ (pu)	$L_{1,2}$ (pu)
T_C	Yg-Δ	110/20	25	0.001	0.02652
T_G	Yg-Δ	20	100	0.025	0.00238

4.1.2 Overhead Lines

Overhead lines are made of bare aluminum and are liable for conducting the energy through the MV system, different from the benchmark that some branches are built with underground cables. For that, the distances are maintained, and the respective parameters are accordingly adjusted. The overhead lines parameters are estimated following the π model, where, for a balanced three-phase transmission line model, the parameters are lumped. The line parameters R, L and C are specified as positive- and zero-sequence parameters that take into account the inductive and capacitive couplings between the three-phase conductors, as well as the ground parameters. The overhead lines' parameters RLC are presented in Table 3.

Table 3 – Parameters of the overhead lines.

Resistance (Ω/km)		Inductance (H/km)		Capacitance (nF/km)	
R_0	R_1	L_0	L_1	C_0	C_1
0.6581	0.5132	0.0051	0.0012	4.0744	10.0971

The proposed distribution system has branches with lengths no longer than 50 m, a fact that simplifies the estimation of the electrical parameters of the model π with hyperbolic corrections. The length of each branch from Figure 5 is presented in Table 4.

Table 4 – Extension of the overhead lines.

Node from	Node to	Length (km)	Node from	Node to	Length (km)
1	2	2.82	7	8	1.67
2	3	4.42	8	9	0.32
3	4	0.61	9	10	0.77
4	5	0.56	10	11	0.33
5	6	1.54	11	4	0.49
6	7	0.24	3	8	1.30

4.1.3 Load Nodes

Loads of the European benchmark are considered to be all symmetric, totaling around 25 MVA. There are two classes of loads: residential and commercial/industrial and are all modelled as three-phase parallel RLC loads, with neutral grounded and considering that for a constant frequency the impedance is constant. The parameters are detailed in Table 5 (only node 2 has no load). It is observed that there is at node 1 a high concentration of load, actually, around 80% of the total load is located there.

Table 5 – Parameters of the symmetric loads.

Node	Apparent Power (kVA)	Power Factor	Apparent Power (kVA)	Power Factor
	Residential		Commercial/Industrial	
1	15300	0.98	5100	0.95
3	295	0.97	265	0.85
4	445	0.97	-	-
5	750	0.97	-	-
6	565	0.97	-	-
7	-	-	90	0.85
8	605	0.97	-	-
9	-	-	675	0.85
10	490	0.97	80	0.85
11	340	0.97	-	-

4.1.4 Generation Nodes

In the European benchmark brochure, the system is considered to have different sorts of generation units as well as energy storage devices. In order to observe fundamentally the aspects of harmonic contents influences in the system, only photovoltaic PV and wind turbine WT units are considered in the following steps, once that their connections with the distribution system occur through inverter-based interfaces and, such topology is well known as a source of harmonic to the system. Three units are modelled and added to the system, located at buses 3, 5 and 7, and their details are presented in Table 6.

Table 6 – Parameters of the distributed generators.

Node	Type	Label	Active Power (MW)
3	Wind Plant	WT	9
5	Photovoltaic Plant	PV-1	0.1
7		PV-2	

These three distributed generation units are connected to the 20 kV network through power transformers, each one with the parameters shown in Table 7.

4.1.4.1 Details of the Photovoltaic Plant

For each PV plant is considered a 100 kW model, connected to the 20 kV distribution network via a DC-DC boost converter and a three-phase three-level voltage source

Table 7 – Parameters of the transformers of the distributed generators.

Transformer	Connection	Voltage Level (kV)	Power (MVA)	$R_{1,2}$ (pu)	$L_{1,2}$ (pu)
T_{PV}	Yg-Δ	0.260/20	0.1	0.001	0.030
T_{WT}	Yg-Δ	0.575/20	10.5	8.3E-4	0.025

converter. The model implements the maximum power point tracking in the boost converter using the incremental conductance and integral regulator technique. The model contains an array of 66 strings in parallel, delivering the maximum power at a sun irradiance of $1000\,\mathrm{W} \cdot \mathrm{m}^{-2}$. The voltage increase is performed by a 5 kHz DC-DC boost converter that convert the natural voltage (273 V DC at maximum power) to 500 V DC. The three-phase three-level voltage source converter converts the 500 V DC voltage to 260 V AC, while keeps unity power factor. The converter control system uses two control loops: an external control loop which regulates DC link voltage around 250 V and an internal control loop which regulates the active and reactive current components. The model also contains a capacitor bank of 10 kvar, liable for filtering the harmonics produce by the voltage converter. Finally, a three-phase coupling transformer with 100 kVA and 260/20 kV is used to connect the PV unit accordingly to the distribution system level.

4.1.4.2 Details of the Wind Plant

The WT plant considered is actually a 9 MW wind farm, containing 6 units of 1.5 MW with a voltage output of 575 V each, that uses a detailed model of a doubly-fed induction generator connected to the 20 kV network by coupling transformers of 1.75 MVA. Wind turbines consist of a wound rotor induction generator and an AC/DC/AC IGBT-based PWM converter. The stator winding is connected directly to the 50 Hz grid while the rotor is fed at variable frequency through the AC/DC/AC converter. This configuration allows extracting maximum energy from the wind for low wind speeds by optimizing the turbine speed, while minimizing mechanical stresses on the turbine during gusts of wind. The wind speed is maintained constant at $15\ \mathrm{m{\cdot}s^{-1}}$ while the control system uses a torque controller in order to maintain the speed at 1.2 p.u with the wind turbine regulated to produce no reactive power.

In the following, details about simulation process are presented. As mentioned before, all the simulations are developed in Matlab/Simulink environment.

4.2 DATA GENERATION

This section presents the design of the data simulation. In order to allow the reproducibility, the major technical aspects as electrical parameters and simulation conditions are detailed of each considered class of electric event.

The nodes of the system presented in Figure 5 are modelled as three-phase voltage-current measurement devices, that measures instantaneous three-phase voltages and currents in the circuit. Voltages are selected to be measured as phase-to-ground, while the measured current is the current that flows through the node. All the data are sampled with a frequency of 12.5 kHz, which represents 250 samples per cycle. This parameter is based on the sampling rate of commercial PMUs. The complete scenario, considering three distributed generation units with detailed models is simulated during 1.2 s, in the discrete mode, with a sampling time of 5 µs. In virtue of the presence of DGs in the simulation routines, the first instants of measures suffer with a high presence of harmonic content and transitory fluctuations. Therefore, once that the aim of the data processing is to extract value information from harmonic contents of each class of event, the beginning of the simulation affects the process of choosing suitable information, for that, the first 0.3 s of the simulations are discarded.

4.2.1 Structural Aspects of the Simulations

The scope of this research is to extract meaningful information from synchrophasors, i.e., PMU data, in order to enhance the diagnosis and classification of events in active distribution networks. In addition to the fact that HIFs hardly ever trigger the protection devices, these events are frequently mistaken with other common events in daily operation of an electric network, for instance, capacitor banks switching and load switching (SOHEILI; SADEH; BAKHSHI, 2018; KAVASKAR; MOHANTY, 2019). In order to examine how informative and discriminative harmonic synchrophasors can be, the following 7 classes of major events are considered in the simulations: capacitor bank switching (CBS), distributed generation switching (DGS), high impedance fault (HIF), low impedance fault (LIF - single, double, and three-phase), load variation (LOV), normal operation (NOP) and transformer energization (TRE).

As mentioned, HIFs are disturbances that mostly occur during a contact with the ground, so, in the most part of the researches this class of events are considered to be a single-phase event (GHADERI; GINN; MOHAMMADPOUR, 2017). Based on that, and considering that low impedance faults do not follow the same characteristic, all the simulated events are adjusted to their common occurrence, as detailed in Table 8, where PG means single-phase (phase to ground) event; PP, double-phase (phase to phase) event; PPG, double-phase-grounded event and PPP, three-phase. This arrangement creates subclasses in the class of events labeled as low impedance faults, so the original 7 classes now become 10 classes of events.

Other aspect that is considered for simulating the range of events is the location of them. For that, the events of interest are simulated in multi locations in the system: bus 1; between bus 1 and bus 2; bus 2; bus 5; between bus 5 and bus 6; bus 6; bus 10; between bus 10 and bus 11 and bus 11. The points defined as *between two nodes* are considered as the midpoint of the nodes, in this way, 9 are the possible points of occurrence of events. As can be seen, the 10 classes can occur in 9 different locations in the system. Each example takes

Table 8 – Class of events and their respective occurrence phases.

Class of Event	**Label**	**Type**
Capacitor Bank Switching	CBS	PPP
Distributed Generation Switching	DGS	PPP
High Impedance Fault	HIF	PG
Low Impedance Fault	LIF	PG, PP, PPG, PPP
Load Variation	LOV	PPP
Normal Operation	NOP	PPP
Transformer Energization	TRE	PPP

an average of 45 minutes to be simulated and, in order to build a homogeneous dataset, each class of event is simulated 30 times at each location, where for every simulation, random values of class-specific parameters are picked up, totaling 2700 examples, with a simulation elapsed time of around 2025 hours, in two computers Intel i5-3230M @ 2.6 GHz with 8 GB of RAM, totalling more than 30 GB of simulated data.

4.2.2 Parametric Variation of the Simulations

The parametric information used in the simulations is presented in Table 9, remembering that in each simulation, the parameters are randomly selected among each interval of values, where Q_C is the capacitor bank power, P_G is the generation power, V_P, V_N, R_P, and R_N are the voltage sources and resistances used in the HIF model ((CUI; EL-ARROUDI; WENG, 2019)), R_f is the fault resistance, R_g is the ground resistance, S_L is the load power, and S_T is the transformer power. The inception angle, i.e, the instant of time which the event is switched into the system is defined considering that in a fundamental cycle of 20 ms the 360° round is completed, the step angle is around 55.5 µs/°, in this way, a random value of time is selected in the inception angle interval. In order to have a controlled environment of simulation, a fixed stamp of time is selected, here $t = 0.7$ s, as the beginning of all the simulations and then, a deviation of time t_{inc} is added to it, emulating the idea of different inception angles.

Table 9 – Parameters considered in the simulations.

Class	**Set of Inception Angle**	**Set of Parameters**	**Set of Values**	**Event Location**
CBS	$[0°, 180°]$	Q_C	$[250, 500]$ kvar	$[1, 6, 11]$
DGS	$[0°, 180°]$	P_G	$\pm\{33.3\%, 66.6\%, 100\%\}$ MW	3
HIF	$[0°, 180°]$	$\{V_P, V_N\}$	$[2, 5]$ kV	$[1, 1-2, 2, 5, 5-6, 6, 10, 10-11, 11]$
		$\{R_P, R_N\}$	$[150, 250]\,\Omega$	
LIF	$[0°, 180°]$	$\{R_f, R_g\}$	$[0.01, 1]\,\Omega$	$[1, 1-2, 2, 5, 5-6, 6, 10, 10-11, 11]$
LOV	$[0°, 180°]$	S_L	$\pm\{15\%, 30\%, 45\%\}$ kVA	1
NOP	-	-	-	-
TRE	$[0°, 180°]$	S_T	$[300, 1000]$ kVA	$[1, 6, 11]$

In the following, all the considered classes are described and visualized considering a PMU at bus 1, i.e, at substation level (though PMUs are installed also at other nodes).

4.2.2.1 Class of event: CBS

The class CBS is implemented as a three-phase parallel RLC load, considering just the injection of capacitive reactive power. The beginning of the event is based on the definitions of the inception angle and the amount of reactive power (Q_C) injected is defined and presented in Table 9. After its beginning, the event lasts up to the end of the simulation. Figure 6 illustrates the waveforms of a CBS.

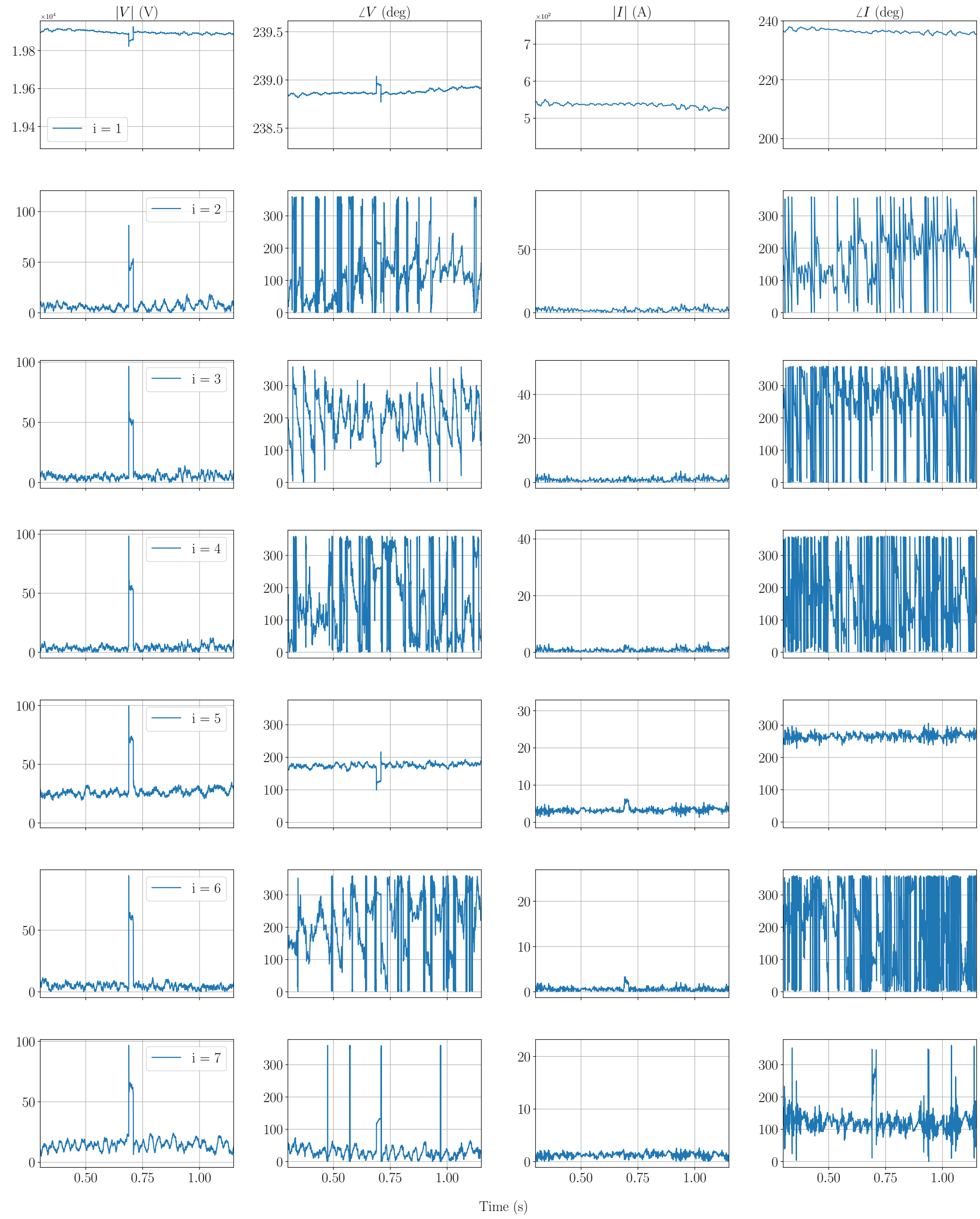


Figure 6 – Harmonic phasors of a capacitor bank switching.

4.2.2.2 Class of event: DGS

The impact of power injection or rejection is considered in class DGS. The primary renewable energy source, WT, was considered in the analysis, for that, different levels of WT power (P_G) are switched on or off, emulating steps of power in the inverters. The amount of power considered in the simulations is shown in Table 9. An example of an energy generation decrease is depicted in Figure 7.

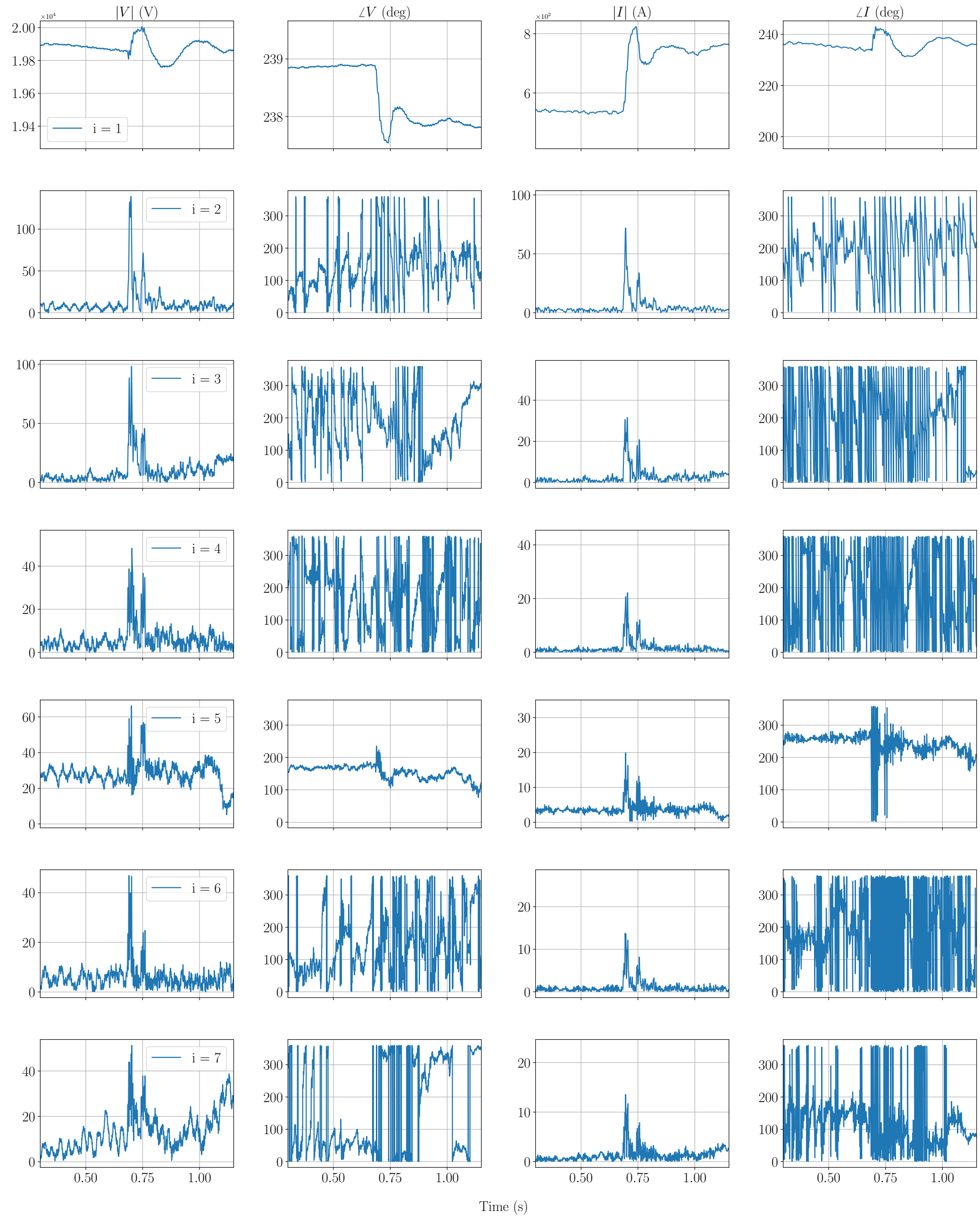


Figure 7 – Harmonic phasors of a energy generation decrease.

4.2.2.3 Class of event: HIF

The major aspects of a HIF are well represented by the model proposed by (EMANUEL et al., 1990) and widely accepted until nowadays. Based on this model, and on the adjustments proposed by Cui, El-Arroudi, and Weng (2019), the parameters adopted to the considered HIF model (see Figure 4a) are presented in Table 9. The values of the voltages (V_P and V_N) and resistances (R_P and R_N) change every 10 ms and 1 ms, respectively. For each possible location of occurrence of the event in the system, half of examples do not consider intermittence and other half, do. The estimated phasors of a HIF is depicted in Figure 8.

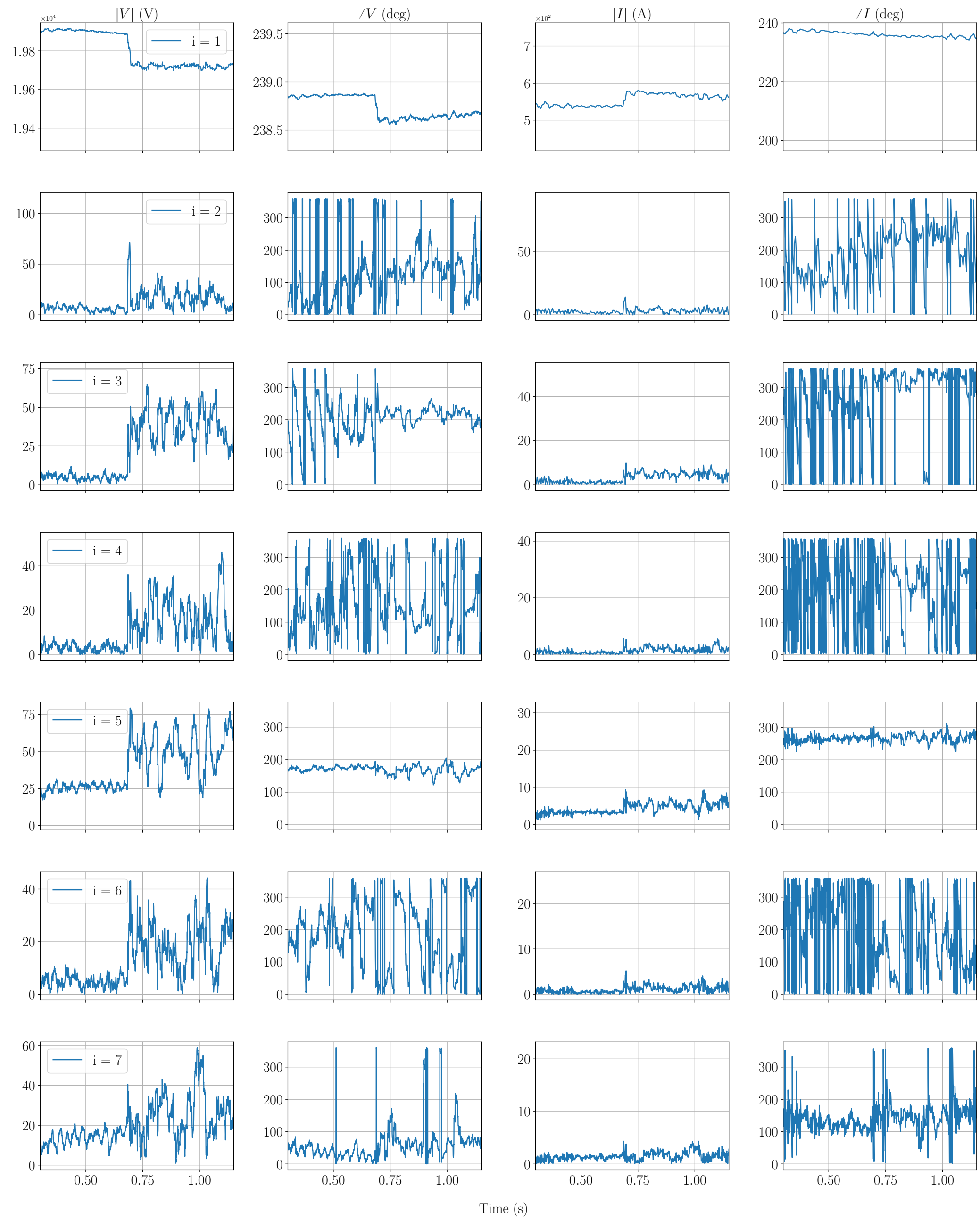


Figure 8 – Harmonic phasors of a high impedance fault without intermittence.

4.2.2.4 Class of event: LIF

These events are implemented as three breaker blocks that can be individually switched on and off to program single, double and three-phase short-circuits. Different from all other evaluated events, LIFs do not last all the simulation, the extinction occurs in order to emulate the trigger of the protection system, therefore, LIFs last $10t_0$ at maximum, where t_0 is the fundamental period of the system. Other two parameters that are considered in the simulation of these events are fault and ground resistance (R_f and R_g) and the interval that defines both resistances are presented in Table 9. An example of a single-phase LIF is presented in Figure 9.

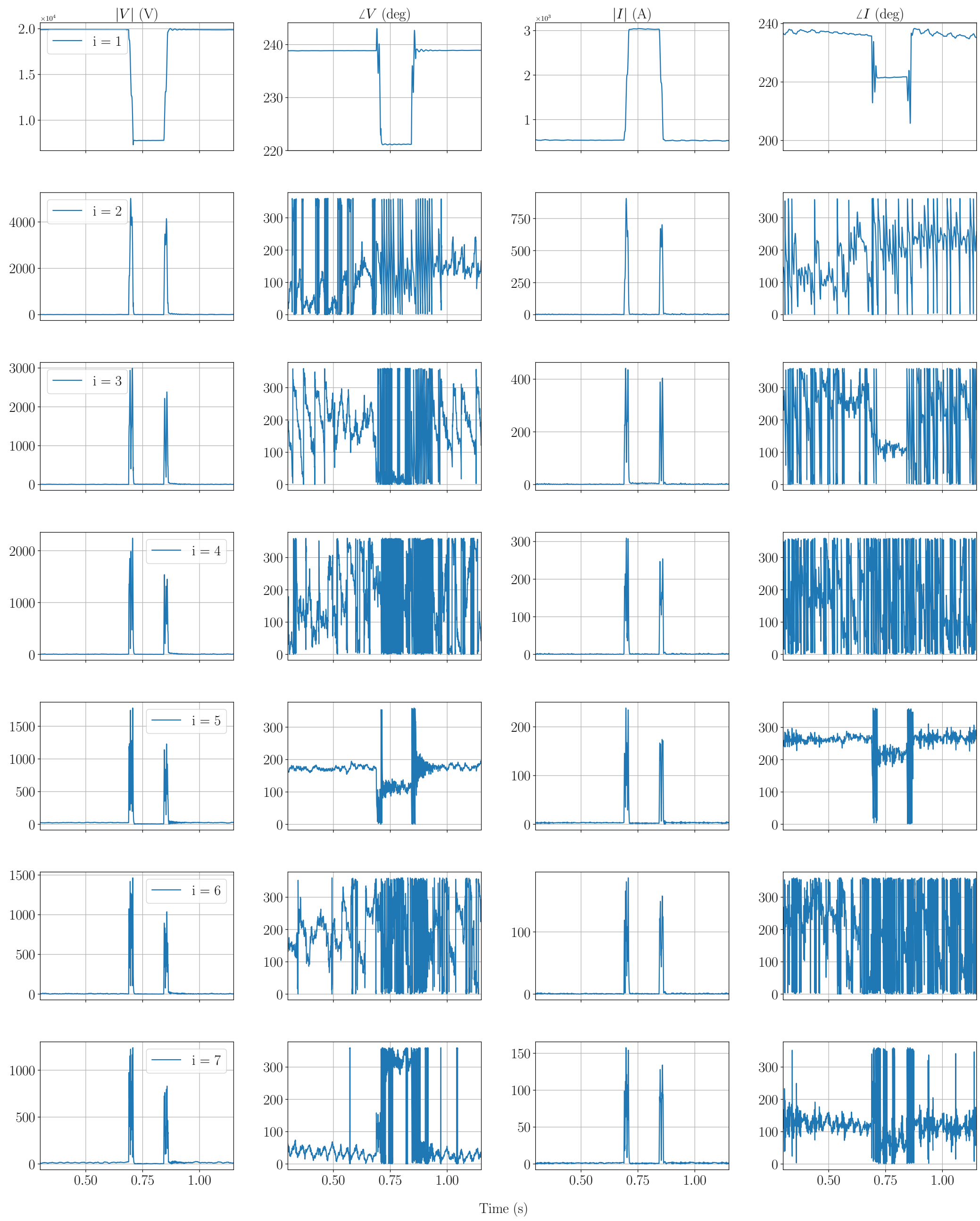


Figure 9 – Harmonic phasors of a single-phase low impedance fault.

4.2.2.5 Class of events: LOV

As observed in Table 5, around 80% of the total load is located at bus 1, therefore, all variations considered in the class LOV are located in this bus. It is considered that the load can be both increased or decreased in the bus, therefore, three changes in the load bus (S_L) is considered: 15%, 30% and 45%. In this way, to emulate such events, the corresponding amount of load is switched on or off, considering the inception angle definitions. Figure 10 illustrates a decrease of load in the system.

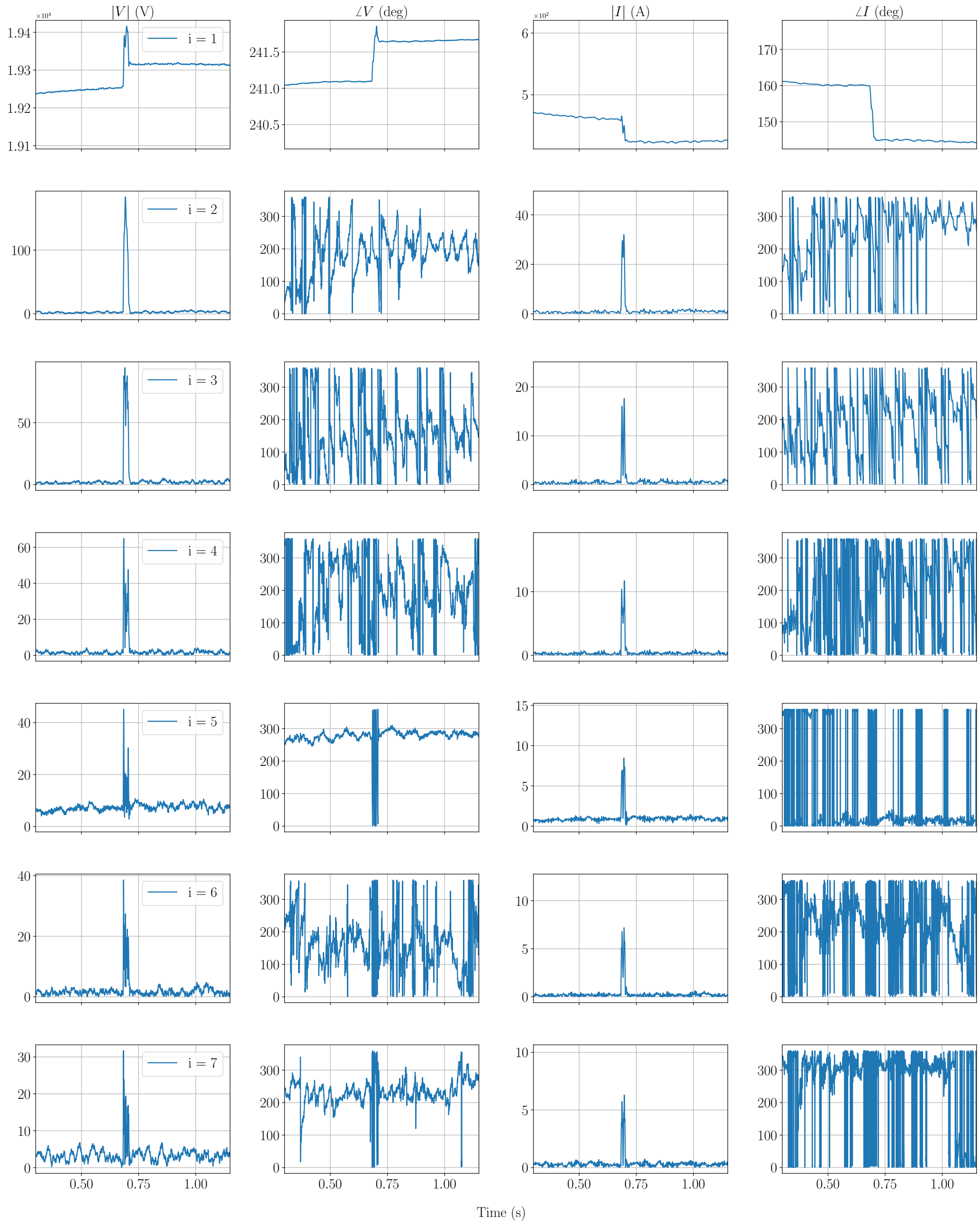


Figure 10 – Harmonic phasors of a load decrease.

4.2.2.6 Class of events: TRE

The class TRE causes the appearance of inrush currents in the system generated by the initial magnetization of the core and winding. In order to evaluate the impact of different levels of inrush current, the different levels of power (S_T) of the transformer are considered when the secondary is connected without load, as detailed in Table 9. An illustration of a TRE is depicted in Figure 11.

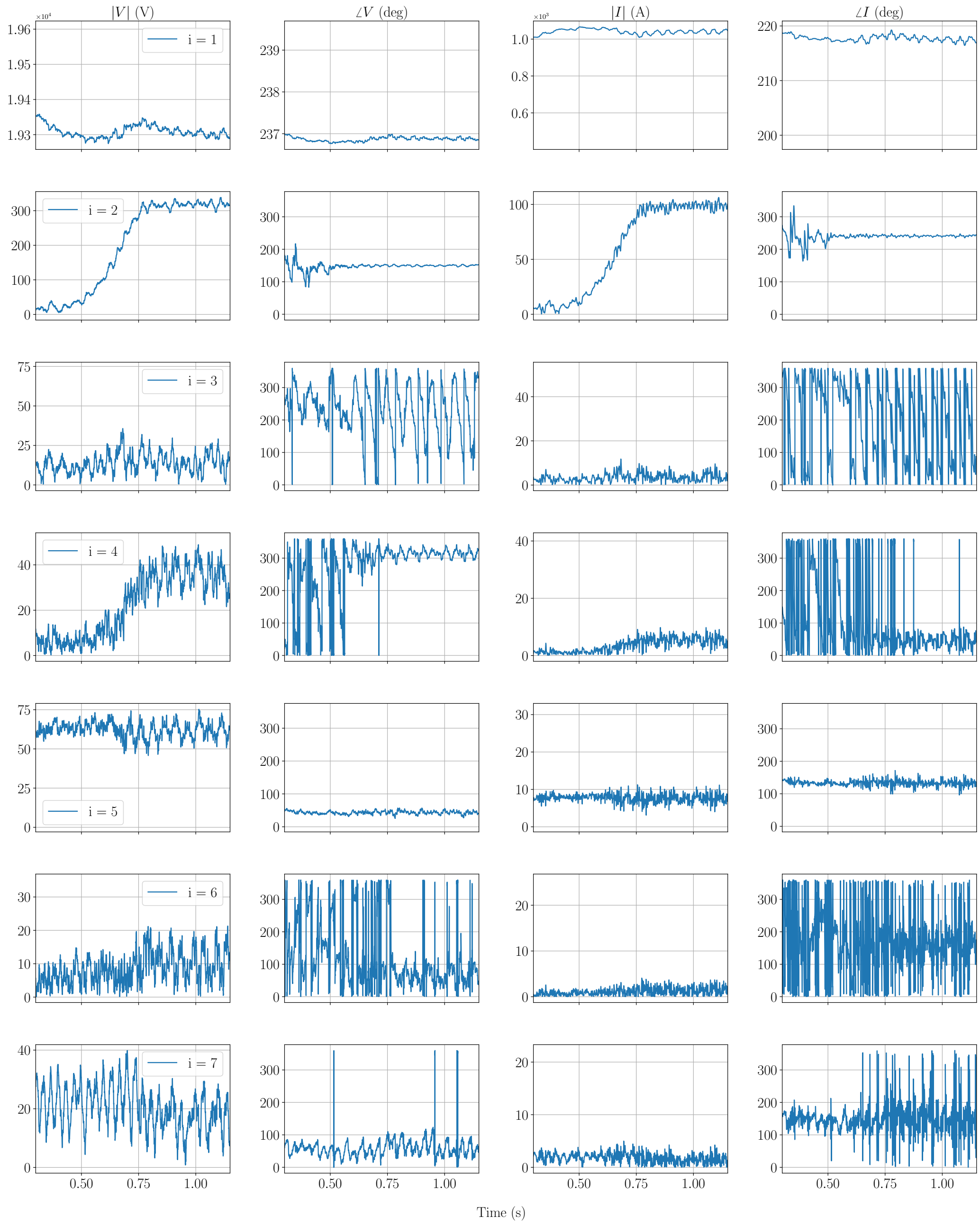


Figure 11 – Harmonic phasors of a transformer energization.

4.3 COMMON DATA PROCESSING

This section presents the common aspects of signal processing adopted in the data that will be used in the classification step. This fact leads to a fair comparison between the considered strategies of event classification, i.e, all the classification models used experience the same pre-processing data, leading to a reasonable evaluation of metrics and model's robustness.

4.3.1 Estimation Process

The current and voltage waveforms previously obtained are processed in terms of harmonic phasor estimation, emulating a class M PMU. The estimation is performed by applying the recursive algorithm of the DFT as in Eqs. 4 and 5, where, by considering an estimation rate of 50 phasors per cycle of 50 Hz it results in a rate of 250 harmonic phasors per second. In this work, some scenarios of estimation are considered to evaluate the relationship between the number of synchrophasors per second needed to discriminate the classes of interest, i.e, how much information is needed to achieve good results in a classification problem based on PMU data. The impact of the estimation rate is depicted in Figure 12, where higher rates impact on more detail of the information.

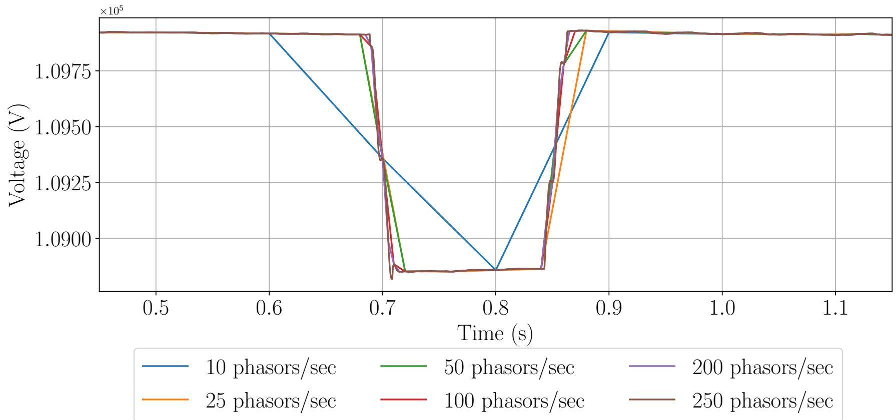


Figure 12 – Impact of estimation rate.

4.3.2 Data Quality Issues

4.3.2.1 Measurement Error

In electric networks, noise is typically more common in low and medium-voltage systems. The noise in distribution systems can be considered white noise (ZHANG, J. et al., 2020). Moreover, in distribution systems, the signal-to-noise ratio (SNR), which represents the level of a desired signal to the level of background noise, is about 60 to 70 dB (ROSCOE et al., 2018), and defined as:

$$SNR = 10\log_{10}\left(\frac{\int_0^T x(t)^2\,dt}{\int_0^T n(t)^2\,dt},\right) \tag{7}$$

where the formula express the ratio between the power of the signal $x(t)$ and the noise $n(t)$ (white noise is characterized by having a zero mean and a constant variance), given a time period T. Considering a distribution system containing PMU across the grid, although the acquired samples will be filtered first, the noise level can still be high enough to pollute the synchrophasor estimation.

Another perspective of error when dealing with synchrophasor data is the difference between a reference phasor and an estimated one. The IEEE C37.118.1 standard defines it as the TVE, which simplifies the compliance specification of the device as well as the magnitude and angle error. In summary, the TVE combines all error sources that may involve a phasor estimation (IEEE..., 2011), as is defined as:

$$\mathrm{TVE}_n = \sqrt{\frac{(\hat{X}_r(n) - X_r(n))^2 + (\hat{X}_i(n) - X_i(n))^2}{(X_r(n))^2 + (X_i(n))^2}}. \tag{8}$$

where $\hat{X}_r(n)$ and $\hat{X}_i(n)$ are the real and imaginary part of the estimated phasor and $X_r(n)$ and $X_i(n)$ the values from the exact (or reference) phasor at the instant of the n. The TVE combines all error sources that may involve a phasor estimation. In this sense, a complete noise analysis makes it possible to measure the TVE of all the interest harmonic contents. Once the IEEE C37.118 standard does not address the TVE for harmonics higher than the fundamental, observing the impact of noise in terms of TVE across a larger spectrum of harmonic contents can enlighten how a classification task based on such information is affected and also, what are the less influenced contents.

Different levels of noise over a HIF voltage waveform are depicted in Figure 13, where it is possible to observe that for the most common scenarios of SNR (around 40 dB), the TVE of the fundamental does not overpass the standard requirement of 1%. As long as the harmonics are increasing, the impact of the noise causes a high level of TVE over the phasors: around 250%, 70%, and 130% of average TVE over the harmonics $i = 3$, 5, and 7, respectively. In a most severe SNR scenario (20 dB), the average TVE reaches even bigger percentiles: 2700%, 700%, and 1100% for the harmonics $i = 3$, 5, and 7. A nuance of the TVE interpretation is the concept of Euclidean distance between estimated and exact phasor, in this sense, when a noisy scenario is considered (SNR $=$ 20 dB in Figure 13), once the distribution of noise over the signal occurs randomly, it is possible that in a specific sample of the time series, the distance increases significantly, becoming this specific TVE an outlier. Moreover, once the SNR is added to the voltage waveform, the estimation process via recursive DFT does not promote the mitigation of the power of the noise over the harmonic, i.e., the impact of noise is more relevant in harmonics greater than the fundamental, once the magnitude of those signals is smaller. This fact also contributes to an increase in the estimation error.

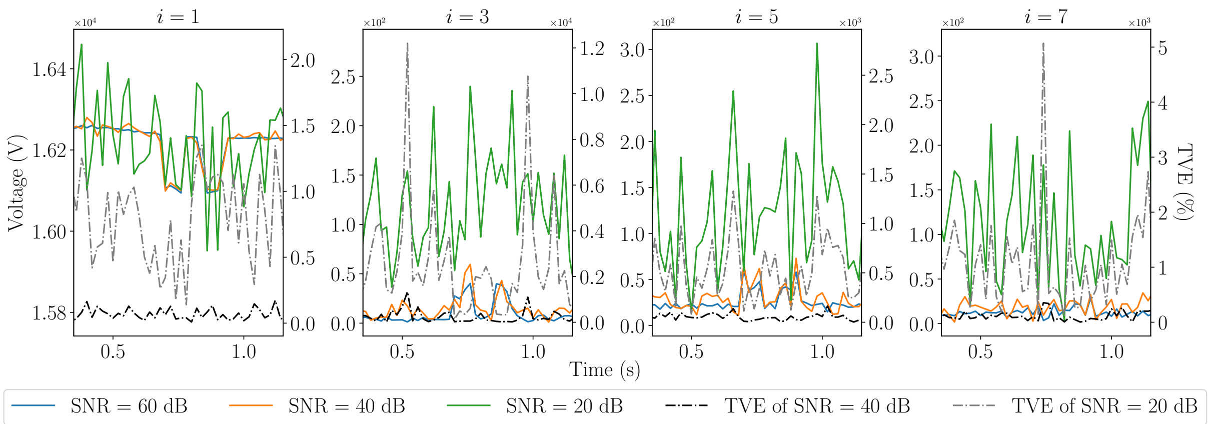


Figure 13 – TVE according to different levels of SNR - HIF event - 50 phasors per second.

When dealing with PMUs, other nature of error is given when considering synchronized measurements. Supposing that in a bus k, PMU_a and PMU_b result on equal measures V. However, under the loose time synchronization scenario, such equality disappears. Such discrepancy between measures can be seen as a phase angle shift (YE; FARAJOLLAHI; MOHSENIAN-RAD, 2022), δ in Eq. 9, defined in terms of a multiple of the fundamental cycle of the original signal.

$$V_k^a = V_k^b e^{j\delta}. \tag{9}$$

Some scenarios of synchronism error are depicted in Figure 14. Once the nature of the angular error is constant, such lack of synchronism is added to the angle signal by correlating time information (number of fundamental cycles) with an angular offset δ.

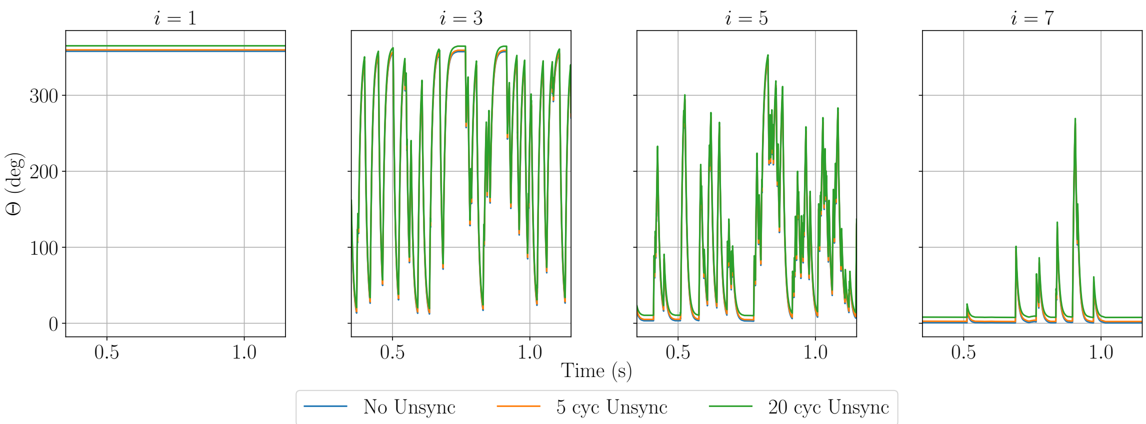


Figure 14 – Impact of synchronism error - HIF event - 50 phasors per second.

4.3.2.2 Missing Data

The information loss in PMUs occur both as a random distribution of missing data over the time series as well as more continuous lack of information, like when a PMU does not measure some cycles of the signal. In this sense, a missing data scenario can be emulated by forcing the PMU to lose some data, i.e., to force the appearance of zeros in the time series, ideally in a random distribution (LIU, Yunchuan et al., 2022; LI, Z. et al., 2021; YUAN et al., 2021). An example of the impact of missing data in a measurement is presented in Figure 15.

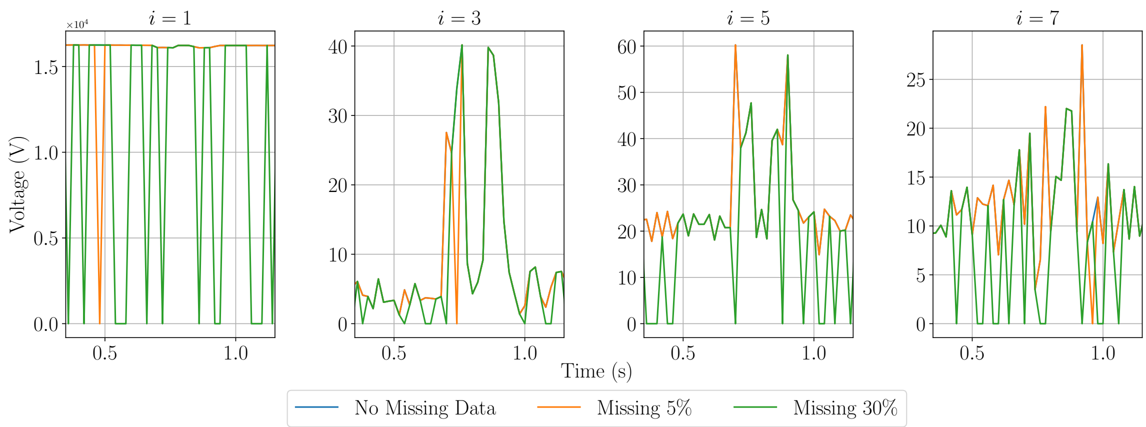


Figure 15 – Impact of missing data - HIF event - 50 phasors per second.

4.4 COMMON DATA CLASSIFICATION

The major classification metrics widely used are depicted in this section. These metrics are extremely useful when comparing machine learning models' performance. Moreover, given the complexity of the HIF phenomenon and its classification methods, a broader evaluation framework is crucial. When protective measures respond to HIF, they may inadvertently affect infrastructures like communication systems, health facilities, and transit infra. Utility providers should weigh safety and sensitivity in their risk evaluations, and advanced HIF classification methods further necessitate additional performance metrics.

A common approach to evaluating the performance of a classification model is through the confusion matrix. It provides a detailed breakdown of true and false predictions made by the classifier when the true labels are known. The confusion matrix (CM) is defined in Equation 10, where TP is the true positive and the model predicted positive, and the true label was also positive; TN is the true negative and the model predicted negative, and the true label was also negative; FP is false positive, and the model predicted positive, but the true label was negative; FN is the false negative and the model predicted negative, but the true label was positive (RASCHKA; LIU, Y. H.; MIRJALILI, 2022).

$$CM = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \tag{10}$$

Some performance metrics can be derived from the confusion matrix, such as:

- Accuracy: it measures the proportion of correctly predicted observations to the total observations:
$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

- Dependability: it refers to the reliability and consistency of the model's predictions across different datasets or conditions:
$$DP = \frac{TP}{TP + FN} \tag{12}$$

- Safety: it refers to the model's ability to make predictions without causing harm or making high-risk errors:
$$SF = \frac{TN}{TN + FN} \tag{13}$$

- Security: it often pertains to the model's resilience to adversarial attacks, where malicious inputs are crafted to deceive the model into making incorrect predictions:
$$SC = \frac{TN}{TN + FP} \tag{14}$$

- Sensibility: it refers to the reasonableness or validity of a model's predictions:
$$SB = \frac{TP}{TP + FP} \tag{15}$$

- F1-score: is the harmonic mean of precision (sensibility) and recall (dependability):
$$FS = 2\left(\frac{TP}{TP + FP + FN}\right) \tag{16}$$

Having provided a comprehensive overview of the problem and a detailed explanation of the data generation, processing and classification, the sequence of this document will present significant contributions to the field of high impedance fault classification in active energy distribution systems. The following chapters will not only introduce innovative strategies but will also include robust analyses to ensure the practical applicability of these proposed methodologies. The aim is to bridge theoretical concepts with real-world scenarios, thereby enhancing the efficiency and reliability of active energy distribution systems.

Building upon the aforementioned context, the following chapters delve into the application of harmonic synchrophasors for the purpose of event classification in microgrids, with a particular focus on high impedance faults. The investigations will encompass a comparative analysis of two principal strategies: the first employing feature-based models, and the second utilizing time-series-based models. The objective is to conduct a thorough comparison of these methodologies, elucidating the inherent advantages and disadvantages of each approach. Central to this comparative analysis is the adoption of a uniform methodology across both strategies, ensuring a consistent and fair evaluation. The outcomes of this extensive analysis are anticipated to make a substantial contribution to the field of event classification in electrical systems. This endeavor not only seeks to enhance understanding of the specific models and their applicability but also aims to inform future research directions and practical implementations in this vital area of electrical engineering.

5 FEATURES BASED CLASSIFICATION APPROACH

The classification of events in distribution systems based on specific characteristics of PMU signals is not as widespread, in addition, more detailed analyzes about the quality of the data required for the task are not easily found in the literature. Therefore, there is a gap to be filled in the field of PMU data analysis for applications in event classification in distribution networks.

5.1 METHODOLOGY

A feature based approach strictly depends on a successful feature definition. The act of extracting features from a time series most of the time takes to a dimensionality reduction once the features are mostly based on a combination of a sequence of information in a considered time interval. In this presented strategy, the PMU data is segmented (see Figure 16), based on the inception of each event. Other aspect that defines the segmentation window is the fact that the events considered in the approach correspond to switching elements in the system, producing short-duration electromagnetic transients (generally lasting a few cycles), returning into a steady-state period after switching or being ceased by the protection system.

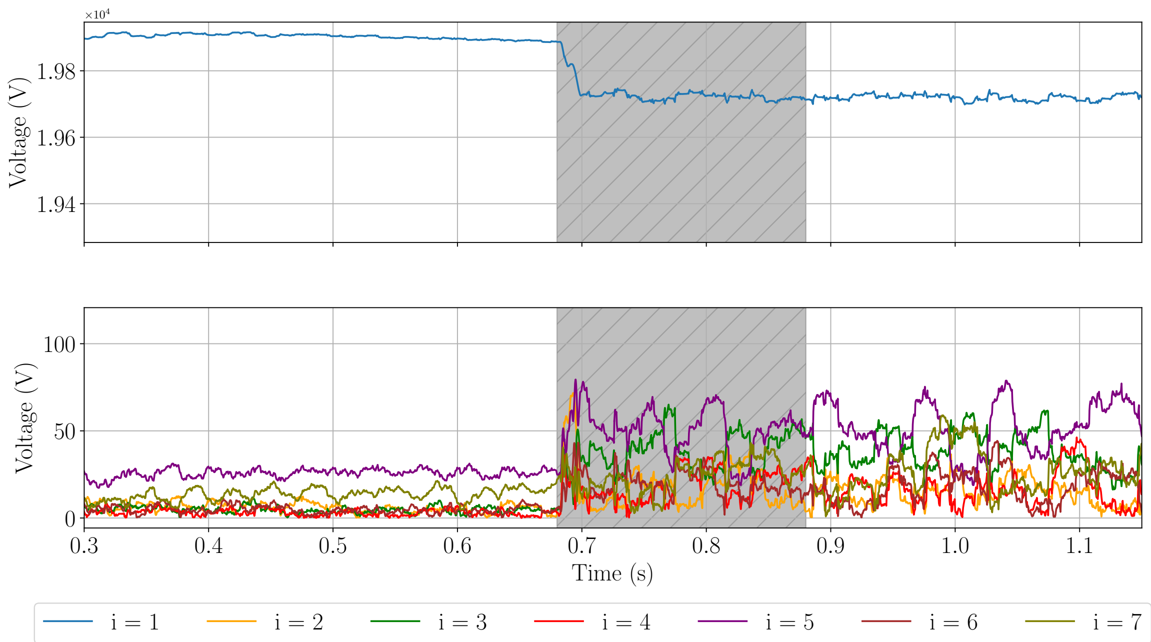


Figure 16 – Segments of a synchrophasor - HIF event - 250 phasors per second.

In this sense, this preset segmentation separates the record into three sections: pre, during, and post transient and the mean value of each segment is calculated, resulting in a set of features $\mathbf{H}_i$ that contains the mean value of each segment of magnitude and angle of each harmonic synchrophasor of each PMU, for the current and voltage waveforms, where i indicates the desired harmonic content:

$$\mathbf{H} = \begin{bmatrix} |\mathbf{I}|^i_{pre} \\ \angle\mathbf{I}^i_{pre} \\ |\mathbf{I}|^i_{tra} \\ \angle\mathbf{I}^i_{tra} \\ |\mathbf{I}|^i_{pos} \\ \angle\mathbf{I}^i_{pos} \\ |\mathbf{V}|^i_{pre} \\ \angle\mathbf{V}^i_{pre} \\ |\mathbf{V}|^i_{tra} \\ \angle\mathbf{V}^i_{tra} \\ |\mathbf{V}|^i_{pos} \\ \angle\mathbf{V}^i_{pos} \end{bmatrix}_{i=2,\ldots,7} \tag{17}$$

For example, each PMU yields 3 mean values (one for each segment), for 2 measures (current and voltage) for 2 phasor information (magnitude and angle) for each desired harmonic content, totalling 12 features for each i_{th} harmonic content. This is, by considering from fundamental up to the 7th harmonic, each PMU provides 84 features.

In order to evaluate the presented methodology, a reference method is considered and analyzed. The selected case study is the one proposed by (SOHEILI; SADEH; BAKHSHI, 2018). The choice has been made for some significant aspects: first, the authors use DFT to extract the features from the waveform signals. This fact matches the usage of phasor estimation for extracting information from electrical waveforms. Second, the authors evaluate the discriminative potential of specific harmonic contents of the signals to classify some classes of events in an electrical network that naturally contains some level of harmonics. It also matches our objective of evaluating the impact of the basal harmonic contents produced by renewable energy sources in the event classification task when considering a large spectrum of frequency features. Finally, the authors evaluate how the noise can affect the classification task. This observation also matches our interest in investigating the impact of error in the data for distinguishing certain classes of disturbances.

The selected method employs the relative relation between the third, fifth, and seventh current harmonic to distinguish high impedance faults from other events. The approach is tested with data simulated from the radial distribution network, IEEE 13-bus. Based on this, some amendments are considered in our analysis to adequate the author's methodology to the CIGRE benchmark system. In this sense, the major considerations are that instead of non-linear loads used in the reference method, inverter-connected energy sources are considered to add natural harmonic levels to the microgrid. In addition, other classes of events are considered (see Table 9).

Another interest of this research is to evaluate promising features from synchrophasors data concerning the most useful harmonic information provided by PMU. Based on that, the potential of combinations of harmonic contents are explored to evaluate their discriminative potential. The explored scenarios of each PMU are detailed in Table 10, where I and V are the evaluated current and voltage signals, M and A are the magnitude and angle of the phasors, for each PMU.

Table 10 – Feature evaluation scenarios.

Scenario	**Set of Features**	**Input**	**Information**	**Amount of Features**
Case of Study	$\{H_5/H_3,\ H_7/H_3\}$	{I}	{M}	6
#1	$\{H_5/H_3,\ H_7/H_3\}$	{V, I}	{M}	12
#2	$\{H_5/H_3,\ H_7/H_3\}$	{V, I}	{M, A}	24
#3	$\{H_2,\ H_4,\ H_6\}$	{V, I}	{M}	18
#4	$\{H_2,\ H_4,\ H_6\}$	{V, I}	{M, A}	36
#5	$\{H_3,\ H_5,\ H_7\}$	{V, I}	{M}	18
#6	$\{H_3,\ H_5,\ H_7\}$	{V, I}	{M, A}	36
#7	$\{H_2,\ H_3,\ H_4,\ H_5,\ H_6,\ H_7\}$	{V, I}	{M}	36
#8	$\{H_2,\ H_3,\ H_4,\ H_5,\ H_6,\ H_7\}$	{V, I}	{M, A}	72

From Table 10, it is possible to see that the sets of features consider both odd and even harmonic contents and discard the fundamental content, and as long as the scenarios are evolving, the number of features increases as well. This fact helps to elucidate if the number of features influences the overall outcome of the classification task, moreover, by disregarding the fundamental harmonic, the evaluation focus only on the discriminative potential of the more subtle information from PMUs.

5.1.1 Data Classification

5.1.1.1 Data Preparation and Machine Learning Models

The inputs are normalized between -1 and 1. Then, the dataset is split, with 80% of the data used for training and 20% for the test. In addition, all the models are evaluated via a 5-fold cross-validation procedure. Since the purpose of this work is to investigate the potential of synchrophasor data in event classification and not in developing machine learning models, a few classic algorithms are tested to observe the potentiality of data in the task of multi-class classification. Therefore, the considered machine learning algorithms are Decision Tree (DT), k-Nearest Neighbors (kNN), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM) (GOODFELLOW; BENGIO; COURVILLE, 2016; BURKOV, 2019). Finally, all the considered machine learning models are tuned via Bayesian Optimization (WU, J. et al., 2019). All the steps of the classification are developed based on the open-source libraries Scikit Learn and Pandas, based on Python.

5.1.1.1.1 Decision Tree

The DT model is a non-parametric approach used for classification tasks. It segments the dataset into subsets based on a series of conditional control decisions. The tree structure comprises nodes representing the data attributes, branches representing the decision rules, and leaf nodes representing the outcome. Its main advantages are interpretability and simplicity, allowing for easy visualization of the decision-making process. However, it can be prone to overfitting, especially in cases with numerous features (BURKOV, 2019).

5.1.1.1.2 k-Nearest Neighbors

The kNN is an instance-based learning method where the class of a new instance is predicted based on a majority vote of its k closest neighbors from the training set. The distance between instances is usually calculated using Euclidean distance, although other measures can be used depending on the context. This model is simple yet effective, requiring no explicit training phase. However, its performance can degrade with high-dimensional data due to the curse of dimensionality (BURKOV, 2019).

5.1.1.1.3 Logistic Regression

A LR despite its name, is a linear model used for classification tasks. It predicts the probability that a given input belongs to a certain class using the logistic function. The model is fitted by estimating the coefficients that maximize the likelihood of the observed data. Logistic regression is easy to implement and interpret, but it assumes linearity between the dependent variable and the independent variables, which can be a limitation in complex datasets (BURKOV, 2019).

5.1.1.1.4 Multi-Layer Perceptron

MLP is a class of feedforward artificial neural network. It consists of at least three layers of nodes: an input layer, one or more hidden layers, and an output layer. Each node, or neuron, uses a non-linear activation function, which allows the network to capture complex patterns in the data. MLPs are powerful models capable of learning non-linear relations but require careful tuning of parameters and can be prone to overfitting (BURKOV, 2019).

5.1.1.1.5 Random Forest

The RF is an ensemble learning technique that builds multiple decision trees during training and outputs the class that is the mode of the classes of individual trees.

It introduces randomness in the construction of trees to ensure model robustness and prevent overfitting. This method is known for its high accuracy, capability to handle large datasets with higher dimensionality, and maintaining performance even if a significant proportion of the data is missing (BURKOV, 2019).

5.1.1.1.6 Support Vector Machine

SVM is a powerful and versatile classification technique. It works by finding the hyperplane that best divides a dataset into classes. The core idea is to maximize the margin between data points of different classes. SVM is effective in high-dimensional spaces and relatively immune to overfitting, especially in high-dimensional spaces. However, its performance heavily depends on the choice of the kernel, and it can be less effective on datasets with overlapping classes (BURKOV, 2019).

5.2 RESULTS

5.2.1 Impact of Estimation Rate

In this first analysis, the models are trained considering a noise of 60 dB on the original waveforms. By specifying the window in the estimation process (Equation 4 and 5), the number of phasors per second is inherently defined. In this sense, the accuracy of the machine learning models is evaluated using different estimation rates: considering commercial PMU rates (1 phasor per cycle or 50 phasors per second in 50 Hz) and also, lower and higher rates of sampling. All the models are trained and tested considering the steps mentioned above, and the classification outcomes are shown in Figure 17. The features scenarios are the ones from Table 10. The models consider measurements from bus 1 and 6 (see Figure 5).

The features scenarios are the ones from Table 10. Therefore, besides evaluating how the estimation process can affect the classifier's performance, it is also considered the best arrangement of features that improves the classification and how the number of features impacts the global accuracy.

In the reference method, a comparison between the temporal evolution of the features and specific thresholds is performed to identify if some classes occur. From Figure 17, the presented analysis has shown that the features proposed by (SOHEILI; SADEH; BAKHSHI, 2018) (*Case of Study*), can be discriminative when evaluating some common events in electrical systems. Considering the urge for increasingly robust solutions in the electrical systems' operation, this method triggered the interest in evaluating if other harmonic components can improve the state-of-art results, as well as how the choice of suitable features can be improved by applying feature selection techniques. The usage of voltage and angle information of the harmonic synchrophasors is also a point of interest in the following discussions.

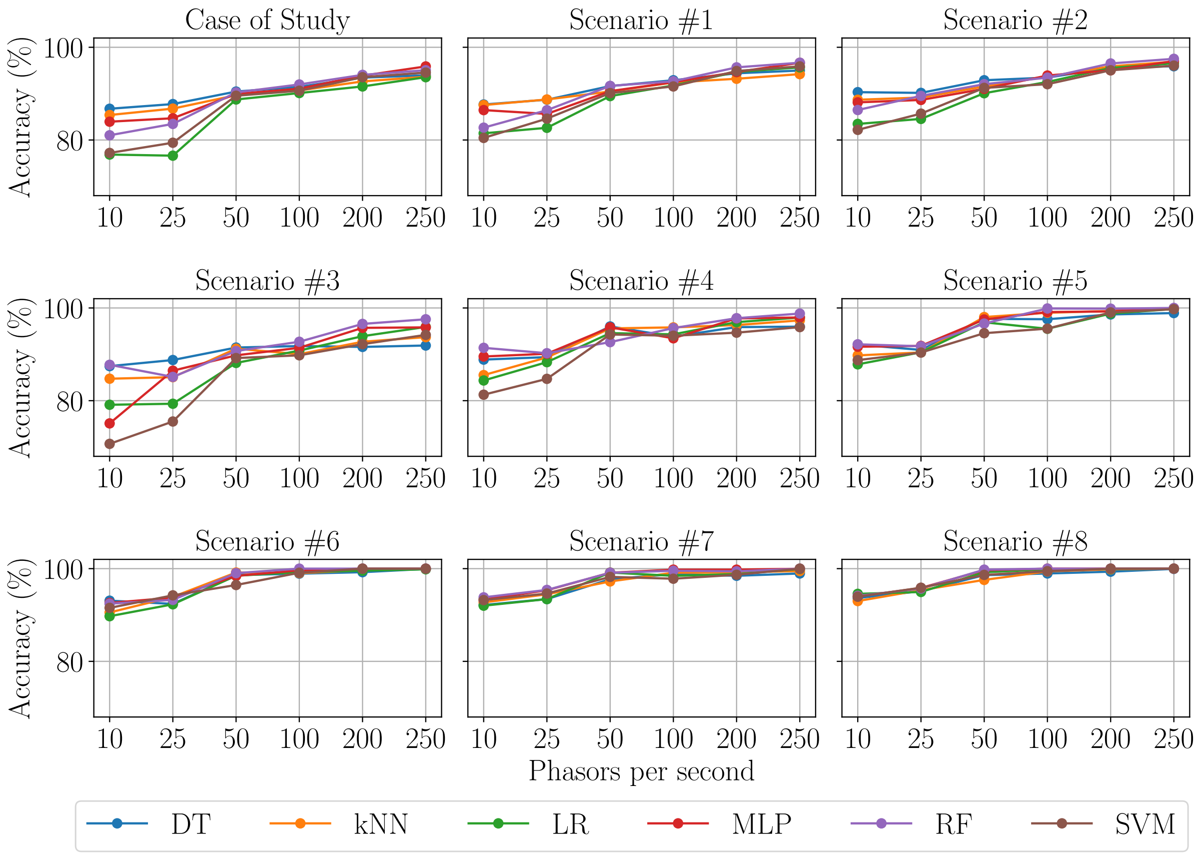


Figure 17 – Accuracy as a function of estimation rate.

PMU data sampled at lower rates presents a lower discriminative potential of classes. For example, for data based on an estimation of 10 phasors per second, the accuracy reaches a maximum of 95% of correct matches. When considering 25 phasors per second, the results suffer a subtle enhancement. However, the best models still present around 5% of misclassifications. With an estimation greater than 50 phasors per second, all the models reached the highest accuracies, thereby, commercial PMU rates are suitable for the classification purpose, showing the best trade-off between classification accuracy and detail of information.

Some feature scenarios are proposed to investigate how discriminative the harmonic contents can be. For that, even and odds harmonics are evaluated separately and also combined. Furthermore, the impact of adding the phasor angle information is also investigated. The first aspect that can be observed is that the overall accuracy is marginally improved as long as more features are considered. All the models enhance their performance when the classification problem involves more features. These aspects point out that once the measurements are available, their usage help to improve the capability of the system's monitoring. By looking at the results from the models that use only the magnitude information of voltage and current harmonic synchrophasors, the usage of odd harmonics instead of even ones results in more accurate predictions, corroborating what is observed in the literature.

The number of features by combining even and odds contents does not significantly improve the accuracy of the models, whereas the number of features doubles up. With the addition of angle information into the classification task, most models naturally present a better performance when compared with the scenarios that only count with the magnitude information. The improvement pattern when comparing the accuracy of the models that use separately even and odds harmonic contents is maintained, i.e., the accuracy of the classifications that use odd contents overcomes the accuracy of the classifications that use even contents. Increasing the number of suggested features marginally improves global accuracy. However, it is important to point out that the proposed set of features' accuracy is substantially high, surpassing the mark of 95% of correct predictions. Consequently, any subtle improvement is a contribution, mainly when considering that using synchrophasor angle information is non-explored in the majority of current classification methods available.

5.2.2 Impact of Data Quality Issues

The noise analysis is performed based on two approaches: first, different levels of SNR are added to the waveforms, and then the phasors are estimated. Second, phasors are estimated from noise-free waveforms and then, levels of TVE are added to the phasors. It is worth highlighting that the following TVE analysis considers an estimation error across all the harmonic contents, following the concepts presented in the IEEE standard (IEEE..., 2011). The following results consider an estimation process of 50 phasors per second in 50 Hz.

The SNR analysis is carried out considering regular noisy scenarios, around 60 dB of SNR, and critical ones, around 20 dB. Then, the overall accuracy of the models is observed as a function of the noise, and the results are shown in Figure 18. All the models are trained and tuned considering 60 dB of noise in the data.

By evaluating the accuracy curves from Figure 18, it is possible to observe that scenarios that contain only the magnitude information of the harmonic synchrophasors (scenarios #1, #3, #5, and #7) tend to present an overall accuracy lower than the scenarios that contain both magnitude and angle information. This fact corroborates the idea that the angle information enhances accuracy. On the other hand, severe levels of SNR (around 20 dB) degenerate the performance of the classification. Nevertheless, the classification accuracy is well-behaved for usual conditions of SNR in the distribution system, once the overall accuracy for most the scenarios of features is around 90% when the SNR is around 60 and 50 dB. The impact of the noise in the classification process can be better visualized by looking at the Figures 19a and 19b, where the confusion is presented considering the RF model and the scenario #6 of features. It is possible to see that, as mentioned, for low levels of noise, the accuracy is quite robust and basically there is no mismatch in the classification. However, as far as the noise increases, the accuracy

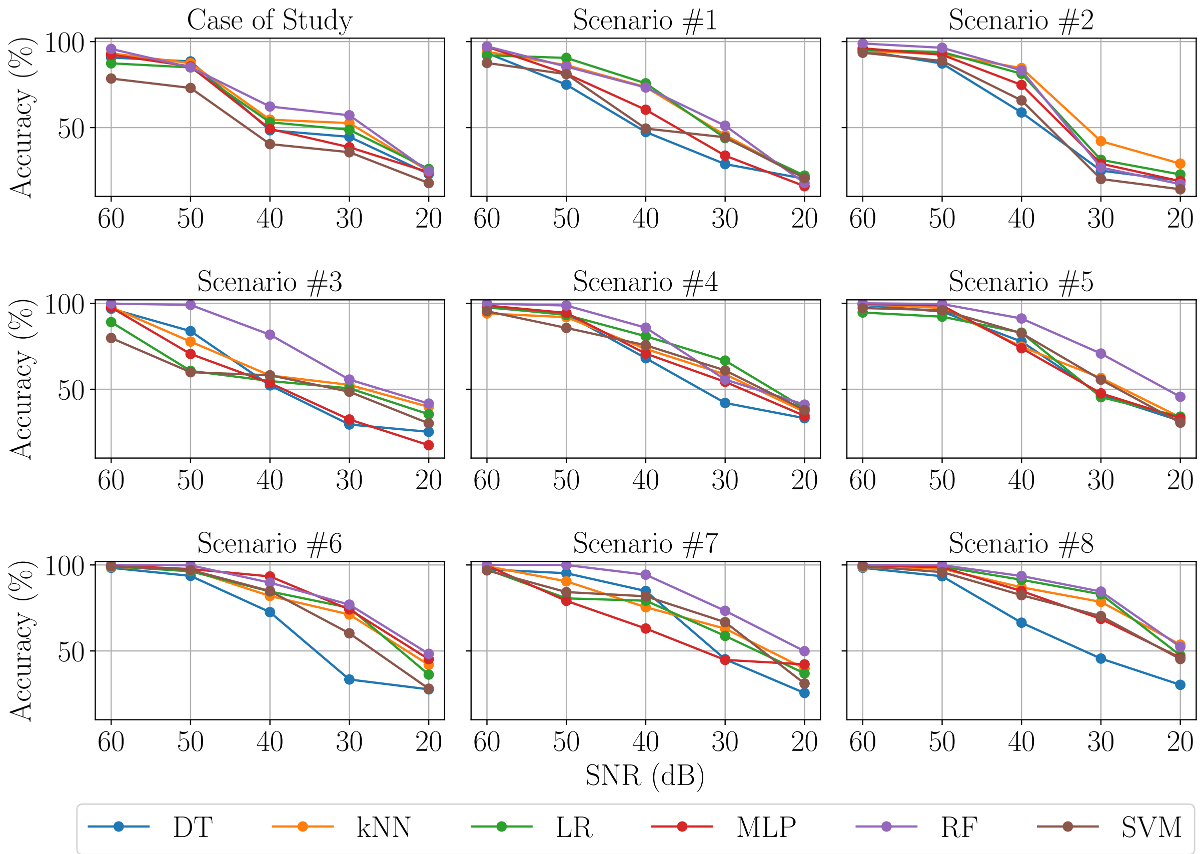
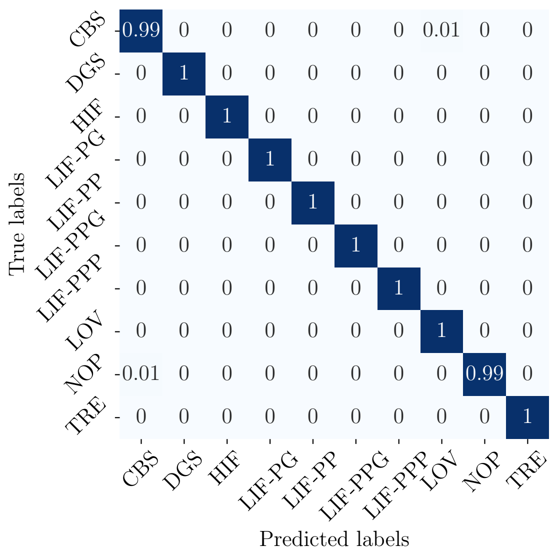
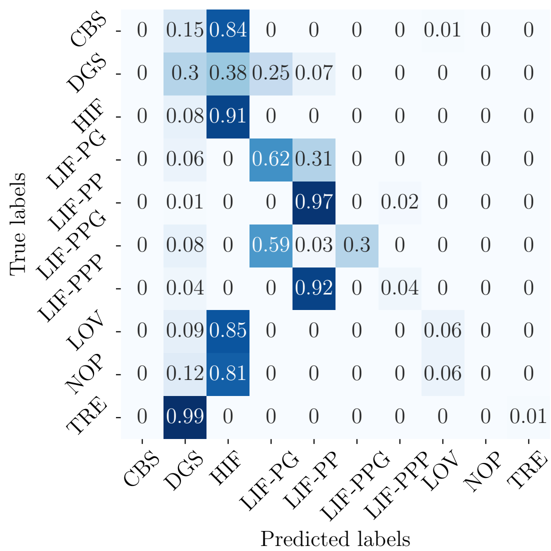


Figure 18 – Accuracy as a function of SNR.

strongly decreases. Nonetheless, the class HIF still being distinguished from the others on most common levels of noise.



(a) SNR = 50 dB.



(b) SNR = 20 dB.

Figure 19 – Confusion matrices for SNR analysis.

The TVE is investigated regarding the harmonic synchrophasors, i.e., it is added to each harmonic ($k > 1$). As long as the IEEE standard, does not address the TVE for harmonics higher than the fundamental, observing how that error can affect the classification process is relevant. Thereby, equal levels of TVE are considered both in current and voltage phasors to observe what level of it degrades the performance of the models. It is considered 5%, 10%, 20%, 30%, and 40% of TVE. All the models are trained and tuned with a 5% of the TVE scenario. The results are shown in Figure 20.

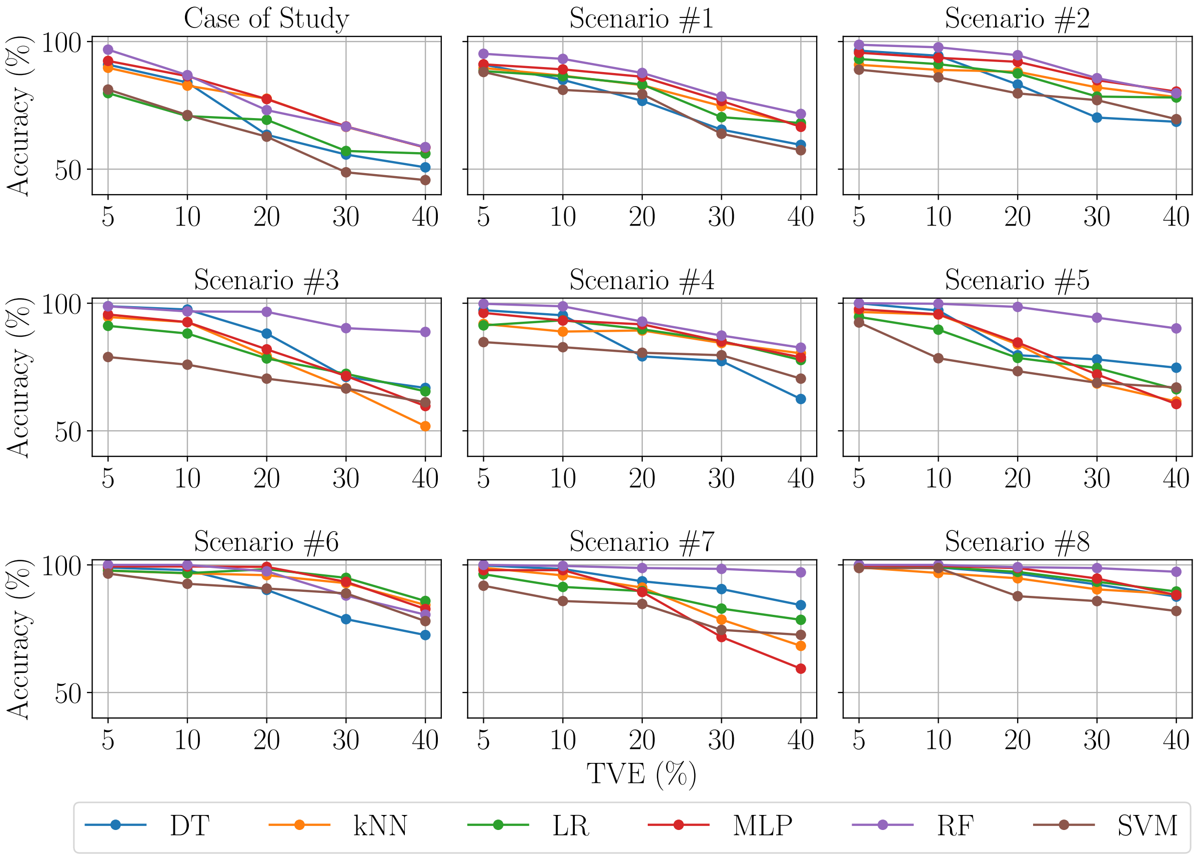
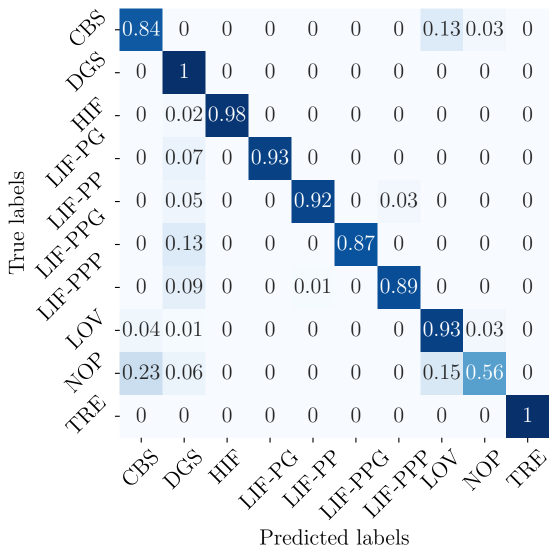


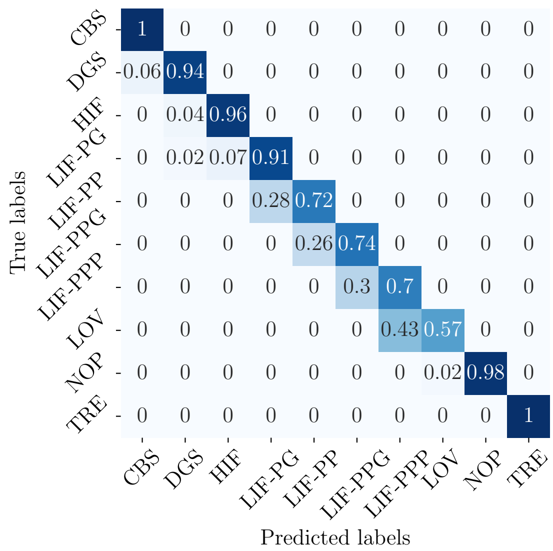
Figure 20 – Accuracy as a function of TVE.

The evaluated TVE scenarios are contained in a 60-40 dB range of SNR. For example, for an SNR of 60 dB, the current TVE only exceeds a rate of 10% at the sixth harmonic. Whereas, at 40 dB of SNR, most of the harmonic synchrophasors of current present a TVE higher than 40%. For the considered controlled TVE scenarios, the added TVE is much lower than the one obtained through SNR analysis in some specific harmonic contents. However, this analysis explores estimation error through the spectrum of interest harmonic contents. A deeper comprehension about the impact of TVE for distinguishing the considered classes of events can be reached by looking at Figures 21b and 21a, where the confusion is presented considering the RF model and the scenario #6 of features. It is possible to see that the class HIF is well separated from the others, even in the scenario that contains the highest level of error.

Once there is no harmonic synchrophasor standard, this analysis's outcomes tend to foster the evaluation of such criteria. Furthermore, a combined SNR-TVE analysis is helpful to evaluate the robustness of the proposed method, where, for common noise



(a) TVE = 10%.



(b) TVE = 40%.

Figure 21 – Confusion matrices for TVE analysis.

scenarios in typical distribution systems, the overall achieved accuracy can be seen as satisfactory for the most common classes of events. This is also observed in distinguishing between high impedance and other non-faulty events.

In order to observe the impact of loss of synchronism, a RF is trained with 1-phasor-per-cycle-data with SNR = 60 dB considering that the measurements are synchronized. In this way, by evaluating scenarios of loss of synchronism (see Figure 22), it is possible to reinforce the robustness of the proposed classification approach. Once the features of the machine learning models depend on the average of segment windows, the number of cycles in which the signal is shifted barely impairs the classification for the best scenarios of features (#5, #6, #7 and #8). Up to 100 ms (5 cycles) of difference between measurements, the classification is barely harmed, however, as far as the measurements drift apart in a time reference, the accuracy of the model suffer a considerable decrease.

In the same way, in order to observe the impact of missing data in the classification task, a RF model is trained with 1-phasor-per-cycle-data with SNR = 60 dB considering that there is no missing data in the training dataset. The impact of some scenarios of missing data is depicted in Figure 23, where is possible to observe that for most considered scenarios the classification is strongly impacted. The fact that the features are mean-based takes to a problem of adding zeros to the average calculation, i.e, the outcome is dragged to a lower value than the one obtained with in a no missing data scenario, in this way, the most data is lacking, the more zeros appear in the average calculation.

5.2.3 Impact of Number of PMUs

As shown in Figure 5, the presented event classification approach considers 2 PMUs across the grid at buses 1 and 6. In order to observe the impact of the number of PMUs

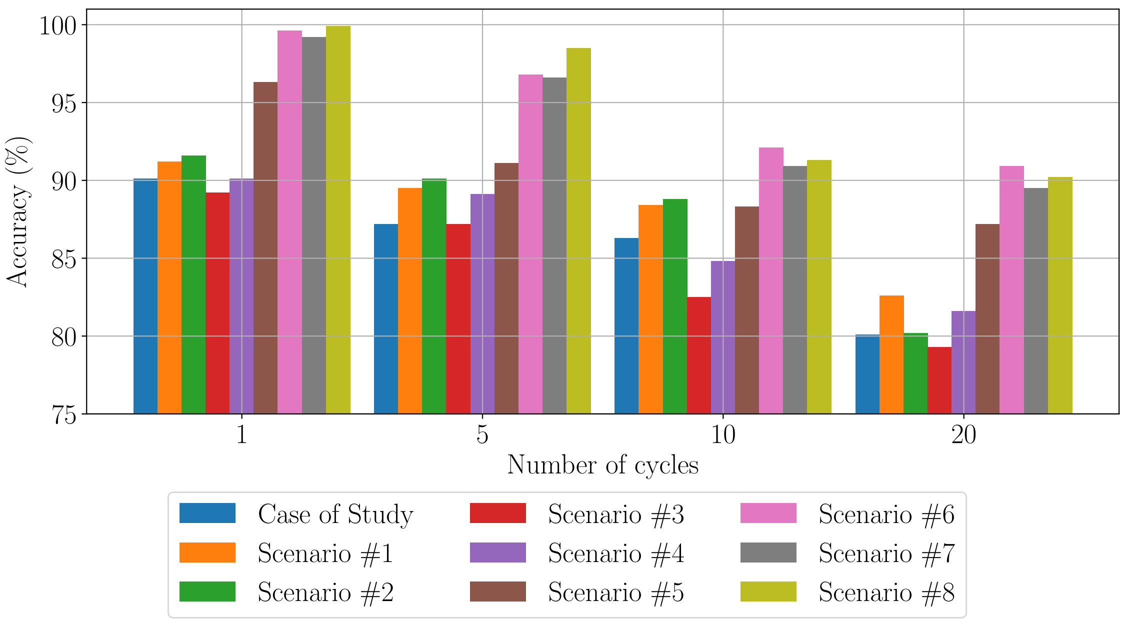


Figure 22 – Accuracy as a function of loss of synchronism.

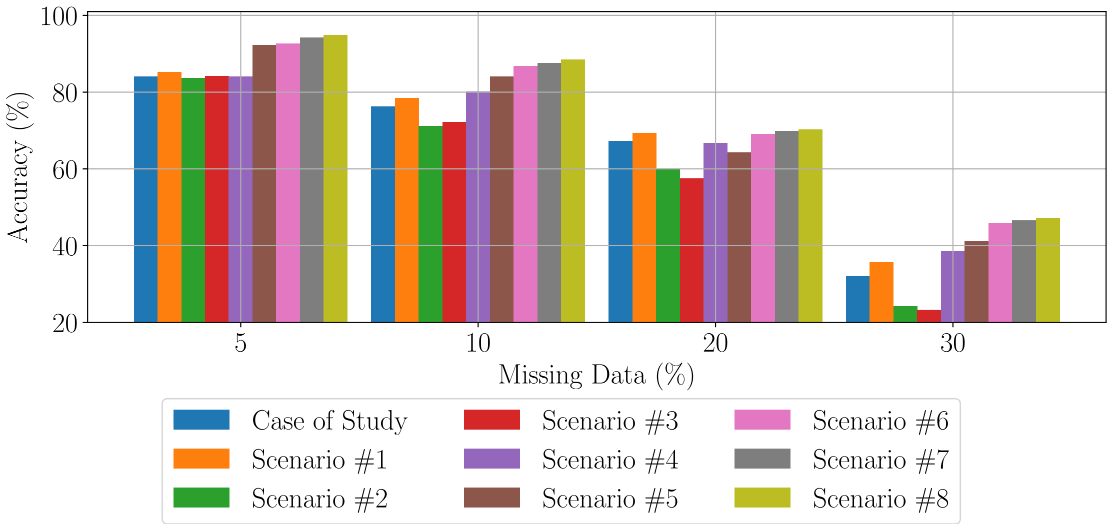


Figure 23 – Accuracy as a function of missing data.

on the performance of the classification task, new arrangements are taken into account, considering that PMUs are also installed at buses 3, 7, and 11. All the methodology above is repeated now, considering an increased amount of processed data. For example, from Table 10, when considering 5 PMUs installed, scenario #8 of features contains 360 variables, i.e., inputs for the machine learning model. The impact of increasing the number of sensors in the system is depicted in Figure 24, considering a RF model trained with SNR = 60 dB in the data.

The increase in model inputs does not significantly increase the performance of

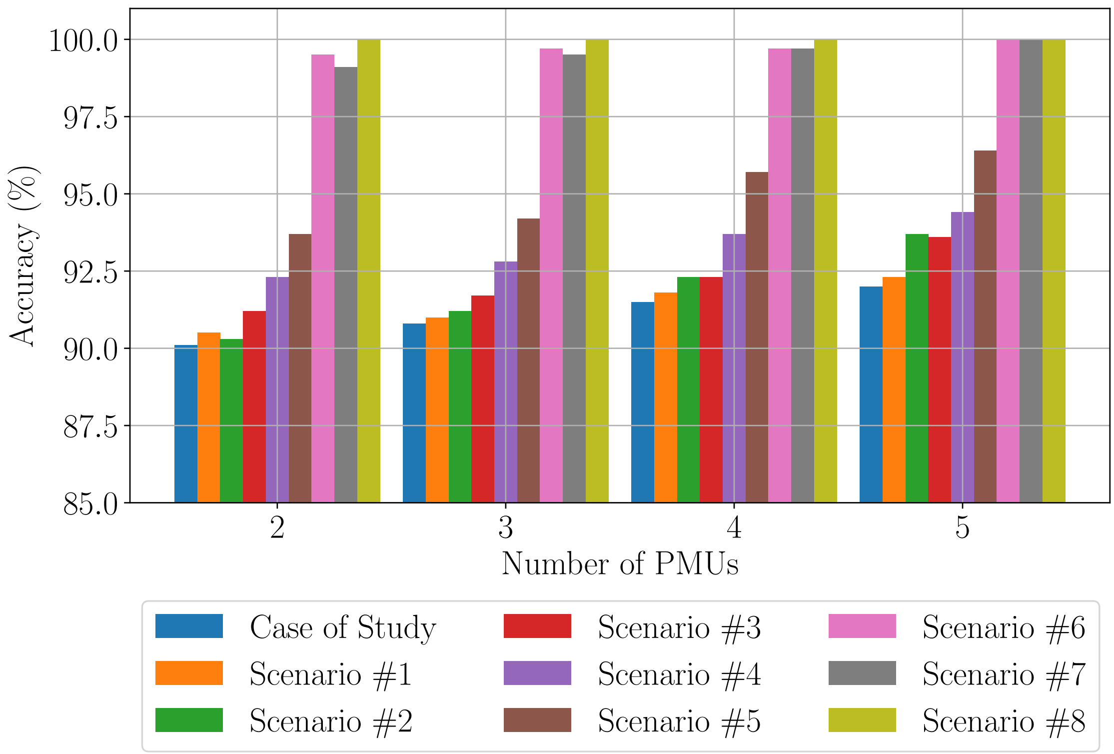


Figure 24 – Accuracy as a function of number of PMUs.

the classification. All the enhancements are marginal and impact around 2% in the total accuracy of the model. This fact corroborates the observation earlier that the increase in information does not necessarily increase the outcomes of the models. In this way, it is notable that only 2 PMUs are suitable for guaranteeing the best trade-off results.

Moreover, by combining the possible arrangements of PMUs in pairs, when fixing one of them at substation level (bus 1), i.e., PMU_{1-3}; PMU_{1-6}; PMU_{1-7} and PMU_{1-11}, the best results are obtained with data from the arrangement PMU_{1-6}, therefore, once the investigation of optimal PMUs allocation is out of the scope of this work, the presented results are considered for establishing the best arrange of sensors.

Finally, considering that the RF model is trained with data from PMUs 1 and 6 (considering a SNR = 60 dB and scenario #6 of features), the impact of event location is also investigated. Considering the classification of the class HIF, the model is tested with data from events occurring in all the system buses, including the ones that do not compound the train/test dataset, i.e., events from unknown locations. The result is presented in Figure 25 and shows that as long as the event goes further from one of the measurement points, the classification slightly degenerates; therefore, for such microgrid, the impact of the distance of the measurement points concerning the HIF event is not a significant concern in terms of correct classification.

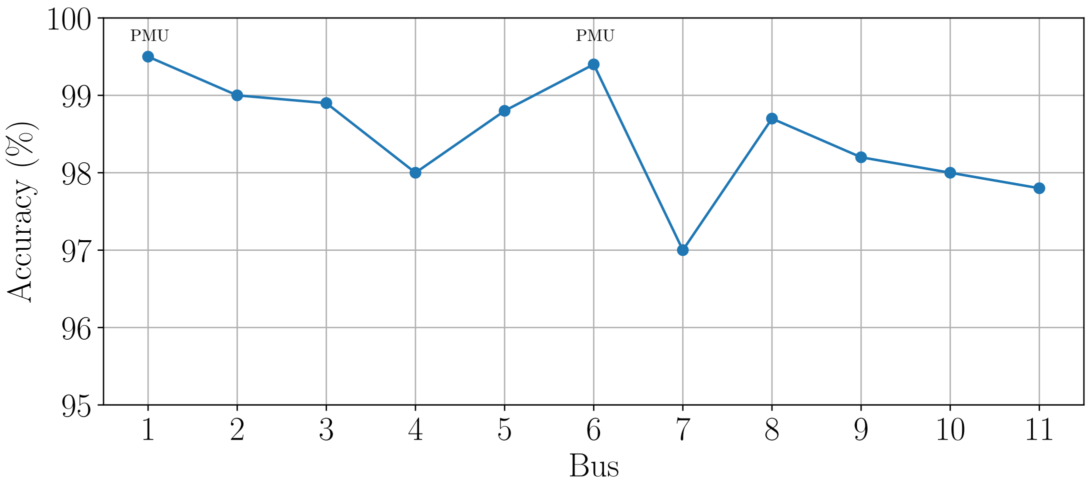


Figure 25 – Accuracy as a function of distance.

5.2.4 Impact of Topology Change

In order to evaluate the generalization of the machine learning models when dealing with data obtained from a different topological scheme than the one known in the training process, several HIF cases are simulated at nodes $1, 3, 6, 9$, and 11, when switches S_2 and S_3 are closed (see Figure 5). The impact of topology changing is depicted in Figure 26, where the branches switching do not substantially impact the measurements, once the magnitude and angle of the phasors for a ring operation vary slightly compared to the radial operation.

Based on the best scores from the results above, this evaluation considers scenario#6 of features to train a RF model. Once the pattern of the event is the same and considering that the mean value takes the features used in the model for each segment, the expected overall of the models is that the change in topology does not depreciate the accuracy of the classification, even considering that the models do not know the ring-data in the training step. HIFs are correctly classified with an accuracy of 96%, with a few confusions with the classes CBS and DGS. This fact suggests that the proposed classification scheme is robust enough to correctly classify HIF even with topological changes in the system (remembering that the models are trained with data from radial operational mode). Moreover, by evaluating the impact of a topological change, the impact of changing the number of branches in the system is also observed. In this way, according to the results, a changing in the number of branches of the system is not necessarily a concern in HIF classification in the considered microgrid.

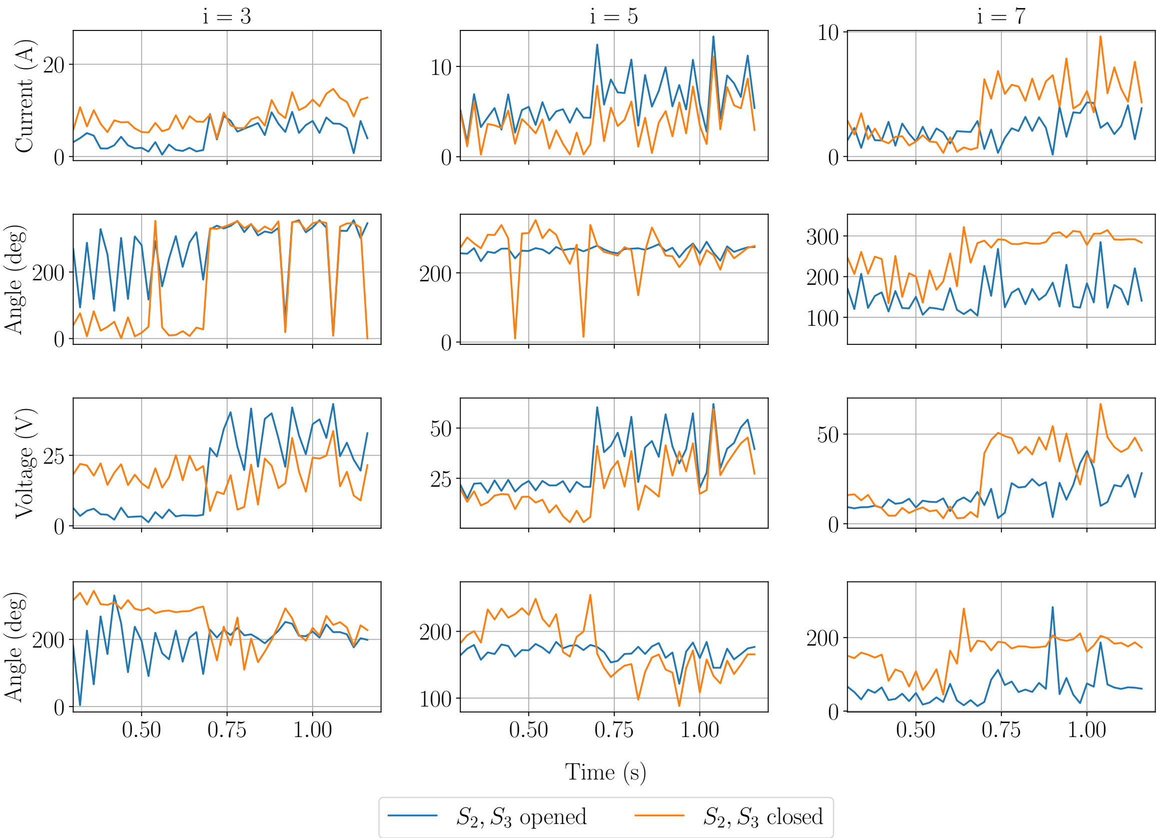


Figure 26 – Impact of topology's change - 50 phasors per second.

5.2.5 Impact of Real Data

The proposed method is validated considering real data of high impedance faults obtained by (MACEDO et al., 2015) in a field test 13.8 kV distribution system where different soil types are tested. The waveforms presented in Fig. 31c are from HIF in different types of soil.

In order to evaluate practical applications of harmonic synchrophasors data, the real data is used to validate the classification. In this way, a RF model is trained with simulated data, considering scenario #6 of features when the data contains 60 dB of SNR and validated with real HIF data. The classification reaches an accuracy of around 99%. Just a few misclassification between classes CBS and DGS showed up. As seen in previous results, the classification gain by using more features (scenarios #7 and #8) is not significant. On the contrary, the number of correct matches decreases when using odd and even features compared to the odd scenarios.

5.2.6 Overall Classification Metrics

Once the reference method approaches a detection technique, a classification comparison is unfeasible. However, using its features methodology, it is possible to observe their potential to distinguish HIF from other events. This way, Table 11 presents the classification performance of the best machine learning model, RF, trained in a noisy scenario of SNR = 60 dB. In order to establish other classification metrics, the multi-class classification problem from Table 9 is simplified to a binary classification problem containing only the classes HIF and non-HIF, where The results reinforce that scenario #6, containing odd features from both current and voltage, magnitude and angle, is the one that reaches the best classification performance when evaluating the best trade-off between the number of features and classification metrics.

Table 11 – Performance of RF model - binary classification.

Scenario	**AC (%)**	**DP (%)**	**SF (%)**	**SC (%)**	**SB (%)**
Case of Study	97	97.5	97.5	97.5	97
#1	97	97.5	97	97	97
#2	97	97.5	97.5	97.5	98
#3	96	96.5	95	95.5	96
#4	96.5	97.5	97	96.5	96.5
#5	98	98.5	98	97.5	98.5
#6	**100**	**99.5**	**100**	**100**	**100**
#7	99.5	99	99.5	99.5	100
#8	99.5	99.5	99.5	100	100

5.2.7 Comparison with Related Works

When compared with the cited works, see Table 12, the novelty of this proposed approach is to combine the usage data from more than one PMU, exploiting the advantages of synchronized measures together with the concept of harmonic phasors, i.e., explore a more extensive set of information in order to enhance the classification of the events based on their harmonic signatures.

Table 12 – Recent approaches for HIF classification presented in literature.

Method	**Signal Considered**	**Performs Error Analysis**	**Uses Phasors**	**Considers Distributed Generation**	**Considers Multiple Measurement Points**	**Uses Real Data**
This work	**I, V**	✓	✓	✓	✓	✓
2022[1]	I	✓	×	×	×	✓
2022[2]	I	✓	×	✓	×	×
2021[3]	I	×	×	×	×	×
2021[4]	I	×	×	✓	×	×
2020[5]	I	×	✓	✓	✓	×
2020[6]	I	✓	✓	×	×	✓
2019[7]	V	×	×	✓	✓	✓
2019[8]	I	×	×	×	×	×
2018[9]	I	✓	✓	×	×	✓
2018[10]	I	×	×	×	×	×

[1] (BHATNAGAR; YADAV; SWETAPADMA, 2022), [2] (LOPES, G. et al., 2022), [3] (SOUSA CARVALHO et al., 2021), [4] (VEERASAMY et al., 2021), [5] (LEDESMA et al., 2020), [6] (WANG, Shiyuan; DE-

Moreover, the microgrid context is also a valency of this work once all the harmonic content from the inverter-based connection is considered in the analysis. The quality of the information is also a significant aspect when the most common noise scenarios and also the impact of the error in the PMU data, i.e., TVE, when looking into harmonics greater than the fundamental, a fact that is relatively new on the field and not investigated by other works. Lastly, the usage of real data in order to validate the method is a relevant aspect, enhancing the reliability of classification methods in real applications.

The objective of the comparison is to prove that the proposed approach presents comparable results with state-of-art methods. Also, by looking again at Table 12 and 10, it is remarkable that the proposed approach contributes to a more complete analysis compared to other related works.

The presented approach uses a benchmark system that allows the investigations' reproducibility and facilitates the comparison with other approaches which are not observed, for instance, in (WANG, Shiyuan; DEHGHANIAN, 2020) and (SOUSA CARVALHO et al., 2021). This argument is reinforced by making codes and datasets publicly available in our work. Also, the kernel of this approach seeks to be as simple as possible through data processing and classification techniques, i.e., the use of PMU data. This fact stands out compared with other related data-based works, once most of the techniques demand higher computational efforts, such as the necessity of a time-frequency information as in (SILVA et al., 2018), (VEERASAMY et al., 2021) and (BHATNAGAR; YADAV; SWETAPADMA, 2022) and more complex machine learning models, as in (VEERASAMY et al., 2021). The authors in (LEDESMA et al., 2020) also use synchronized measures; however, they do not explore the harmonic with an order greater than the fundamental and do not perform any error analysis. Also, in (WANG, Shiyuan; DEHGHANIAN, 2020), only one real measurement was used to validate the system's performance, which does not guarantee detection of high impedance faults under different conditions. Finally, the training of the models with simulated data and the validation with real data proposed here demonstrates the generalization capacity of the proposed solution, being able to be extended to other distribution networks with similar topologies and data acquisition setups, which is also underexplored in the literature.

HGHANIAN, 2020), [7] (CHAKRABORTY; DAS, 2019), [8] (KAVASKAR; MOHANTY, 2019), [9](SILVA et al., 2018), [10](SOHEILI; SADEH; BAKHSHI, 2018)

6 TIME SERIES BASED CLASSIFICATION APPROACH

Methods based on non-processed time series are relatively innovative in the electrical system event classification field. In this way, there is a path to be paved when considering the microgrid context, containing a large infrastructure of monitoring devices. Therefore, approaches considering PMU data combined with modern deep learning architectures can be seen as innovative solutions. Moreover, any event classification approach must be sensitive enough to correctly distinguish most common classes of events on microgrids, majorly the ones that are usually challenging to be observed, as occur with high impedance faults. Another perspective of application is that by using only voltage information, this approach stands out by its simplicity, once such measurement can be obtained even by low-cost devices.

6.1 METHODOLOGY

The results gleaned from the feature-based approach previously delineated have taken to an in-depth investigation into the features that are most pertinent to the classification of events within microgrid systems. In this context, it is noteworthy that the odd harmonic synchrophasors emerge as the most efficacious in terms of classification outcomes. Furthermore, the deployment of two PMUs has been identified as sufficient, striking a balance between the precision of event classification and the monitoring system's size. This is particularly relevant when operating at a phasor estimation rate of 50 phasors per second. Consequently, an evaluative study focusing on an approach that amalgamates time series data from PMUs has been initiated, promising to enhance the understanding and efficiency of event classification in microgrids.

In order to explore the harmonic synchrophasors characteristics and the reason for using multiple measurement points, data from PMU 1 and 6 are combined, creating new time series based on the difference between the magnitude ($|\Delta V|$) and angle ($\Delta\Theta$) of the correspondent harmonic phasor at each time stamp. The odd harmonic contents are considered, resulting in $\mathbf{H}$ that is the matrix of time series:

$$\mathbf{H} = \mathbf{H}^{PMU_1} - \mathbf{H}^{PMU_6} = \begin{bmatrix} |\Delta V_i| \\ \Delta\Theta_i \end{bmatrix}_{i=1,3,5,7} \tag{18}$$

The choice of odd harmonics is corroborated extensively in literature when classifying switching events on electric systems, mainly when considering HIF (KIM; DON RUSSELL, 1988; LIMA; BRITO; SOUZA, 2019; WANG, Shiyuan; DEHGHANIAN, 2020; REZAEIEH; BOLANDI; JALALAT, 2023). The combining PMU data strategy chosen based on the difference of measurements takes into account a certain gap in the literature of approaches that use PMU data but discard angle information when processing its information and also, promising enhancements in outcomes observed by the few approaches that use them (REZAEIEH; BOLANDI; JALALAT, 2023; LEDESMA et al., 2020; CUI; EL-ARROUDI; WENG, 2019; LIMA; BRITO; SOUZA, 2019).

The harmonic behavior of combined voltage measurements for common events is depicted in Figure 27. It is possible to note that each event has its own harmonic manifestation after the event's occurrence. All the events are switched considering a constant time stamp $t = 0.7\,\mathrm{s}$, added to the correspondent inception angle t_{inc}. Some of the information, like the angle of the third harmonic and magnitude of the seventh harmonic, are quite difficult to distinguish, and other patterns are relatively straightforward. These harmonic behaviors in the time series allow the classification models to separate one event from another accurately.

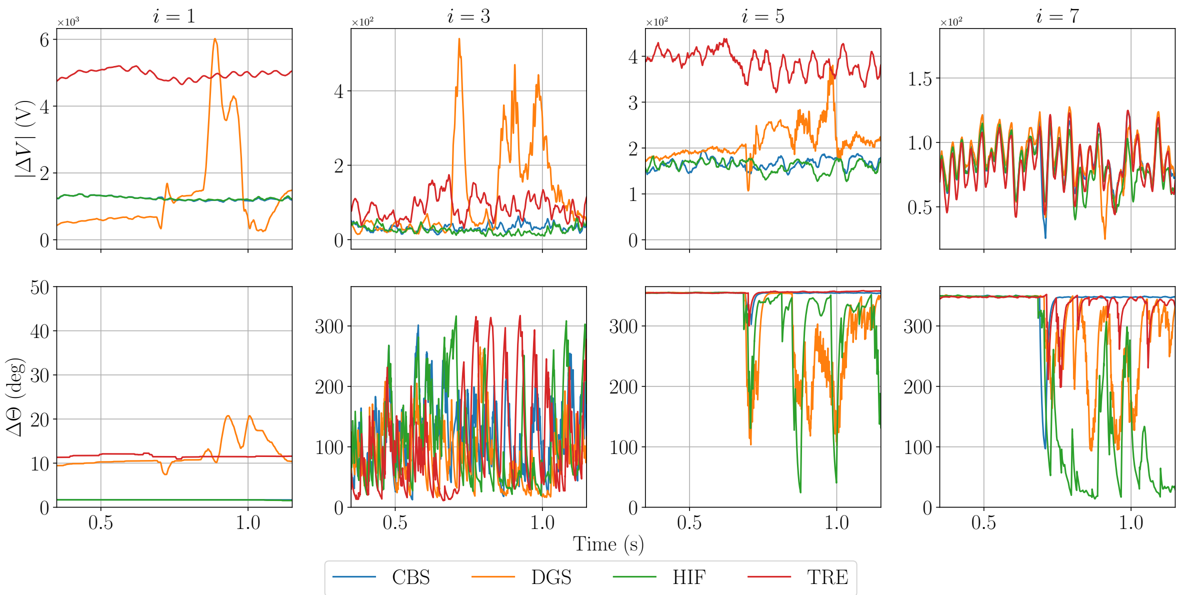


Figure 27 – Behavior of time series in some events - 250 phasors per second.

6.1.1 Data Classification

6.1.1.1 Data Preparation and Machine Learning Models

The synthetic dataset is split with 80% of the data used for training and 20% for the test. In order to validate the methodology, datasets containing HIF data with a topological change in the original system as well as a real dataset of HIF are used. These dataset are not present in the training/testing procedures. In addition, the classification considers a 10-fold cross-validation procedure. All the training data considers a SNR of 60 dB in the data, and there is no lack of information or synchronism error. Since the purpose of this work is to investigate the potential of combined synchrophasor data in event classification and not in developing machine learning models, a classic and a state-of-the-art algorithms are tested to observe the potentiality of PMU time series in the task of multi-class classification. All the data processing steps are developed based on the open-source libraries Scikit Learn and tsai, based on Python.

6.1.1.1.1 Long Short-Term Memory Network

LSTM models are a subtype of RNNs, are characterized by their unique memory cells, LSTMs are adept at maintaining information over extended periods, a critical attribute for analyzing time series data. Integral to their architecture are three types of gates—input, forget, and output—which judiciously regulate information flow, enabling these models to adeptly manage long-term dependencies, a notable challenge in traditional RNNs. This capability is essential in time series contexts where historical data's influence extends across numerous time steps. LSTMs excel in processing data sequentially, a requisite for time series analysis, and their versatility spans various applications, including stock market forecasting and anomaly detection. Their ability to learn and recognize intricate data patterns further cements their standing as a robust tool in time series classification (BURKOV, 2019).

6.1.1.1.2 Transformer Network

The attention mechanism in neural networks is inspired by how human attention works. This technique enhances the model's ability to focus on specific parts of input data. It assigns different weights to different parts of the input, allowing the model to prioritize the most relevant information for the task. These weights, often called attention weights, are learned during training. The mechanism enables the model to consider the context and relationships between elements in the input, resulting in more accurate and contextually relevant predictions. The core of the TST model (ZERVEAS et al., 2021) is the transformer encoder adapted from (VASWANI et al., 2017) and consists of changes that make it compatible with multivariate time series data instead of sequences of discrete word indices.

Considering a training set $\mathbf{X} \in \mathbb{R}^{w \times m}$, with length w and m different variables, constituting a sequence of w feature vectors $\mathbf{x}_t \in \mathbb{R}^m : \mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_w]$. The feature vector $\mathbf{x}_t$ is linearly projected onto a vector $\mathbf{u}_t$ of the same d-dimensional vector space as the internal representation vector of the model, and fed to the first self-attention layer to form the keys ($\mathbf{K}$), queries ($\mathbf{Q}$), and values ($\mathbf{V}$). The parameters $\mathbf{W}_p \in \mathbb{R}^{d \times m}$ and $\mathbf{b}_p \in \mathbb{R}^d$ are fully learnable. Then, one can obtain:

$$\mathbf{u_t} = \mathbf{W_p x_t} + \mathbf{b_p}. \tag{19}$$

Since the transformer is a feed-forward architecture, in order to make it aware of the sequential nature of the time series, positional encoding weights $\mathbf{W}_{pos} \in \mathbb{R}^{w \times m}$ are added to the input vectors $\mathbf{U} \in \mathbb{R}^{w \times m} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_w]$, resulting in $\mathbf{U}^{'} = \mathbf{U} + \mathbf{W}_{pos}$.

This positional encoding is a representation of the position at each index of the time series and can be interpreted as a d-dimensional embedding for these positions in a sense that a linear model can be employed to learn to map the input-output relationship (ZHENG; RAMASINGHE; LUCEY, 2021). Different from the original architecture (VASWANI et al., 2017), which uses sine-cosine characteristics for positioning information, the TST model uses a non-deterministic learnable parameter known as Gaussian embedder Equation 20, where σ is the standard deviation. In short, any embedder $\mathbf{\Psi}$ is defined in terms of shifted basis functions $\Psi(t,x) = \Psi(t-x)$:

$$\Psi(t,x) = \exp\left(-\frac{||t-x||^2}{2\sigma^2}\right). \tag{20}$$

The Gaussian embedder has a higher upper bound for the stable rank that can be controlled by σ. Another model change is that after computing self-attention, TST uses batch normalization instead of layer normalization because it can mitigate the effect of outlier values in time series (ZERVEAS et al., 2021). Such aspects differ in virtue of the statistical nature of the numerical information of the time series when compared with word embeddings from the original encoder-decoder transform model (VASWANI et al., 2017).

Up to here, the input encoding captures the meaning of each value of the time series together with the positional encoding that establishes information about the position of the information inside the time series. Now, the representation $\mathbf{u_t}$ is fed to the transformer encoder, which is liable to perform the self-attention procedure onto features to allow the model to relate values. In other words, such step gives matrices of weights $\mathbf{K}$, $\mathbf{W}$, and $\mathbf{V}$, with $\mathbf{Q}, \mathbf{K}$, and $\mathbf{V} \in \mathbb{R}^{w \times d}$, that are the input transformed input information. The attention $\mathbf{A}$ is obtained with Equation 21, where the softmax function guarantees that the sum of each row's values equals one. The values of each element in $\mathbf{QK}^T$ represent how intense the relationship between the feature at each timestamp is. Hence:

$$\mathbf{A}(\mathbf{K}, \mathbf{Q}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d}}\right)\mathbf{V}. \tag{21}$$

Finally, each row of the matrix $\mathbf{A}$ captures the meaning of the information given by the input embedding or the position in the time series, represented by the positional encoding, and each time-stamp-value interaction with the whole time series. The Equation 21 is called single-head attention. Usually, a TST model uses a multi-head attention encoder. In summary, as in single-head architecture, the input is transformed into three matrices $\mathbf{K}$, $\mathbf{Q}$, and $\mathbf{V}$, and then multiplied by three parameter matrices $\mathbf{W}^K$, $\mathbf{W}^Q$, and $\mathbf{W}^V$, all $\in \mathbb{R}^{d \times d}$. Such multiplications are then split into correspondent smaller matrices with dimension d/n_h, where n_h is the number of parallel attention heads, so every head sees the entire information series with a smaller part of the embedding of every single information. The attention of each head, based on these smaller matrices are:

$$\mathbf{A}_{\text{head}_i}(\mathbf{KW}_i^K, \mathbf{QW}_i^Q, \mathbf{VW}_i^V) = \mathbf{z_t}. \tag{22}$$

In this way, each head monitors the whole time series but with a different perspective of the embedding from each information. The final representation vectors $\mathbf{z}_t \in \mathbb{R}^d$ corresponding to all time steps are concatenated into a single vector $\bar{\mathbf{z}} \in \mathbb{R}^{d \cdot w} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_w]$, which is the input to a linear output layer with parameters $\mathbf{W}_o \in \mathbb{R}^{n \times (d \cdot w)}$ and $\mathbf{b}_o \in \mathbb{R}^n$, where n is the number of classes to be estimated for the classification problem:

$$\hat{\mathbf{y}} = \mathbf{W_o}\bar{\mathbf{z}} + \mathbf{b_o}. \tag{23}$$

In the classification problem, the prediction $\hat{\mathbf{y}}$ will be passed through a softmax function to obtain a distribution over classes. Its cross-entropy with the categorical ground truth label will define the classification loss. The operations described above are depicted in a brief scheme in Figure 28.

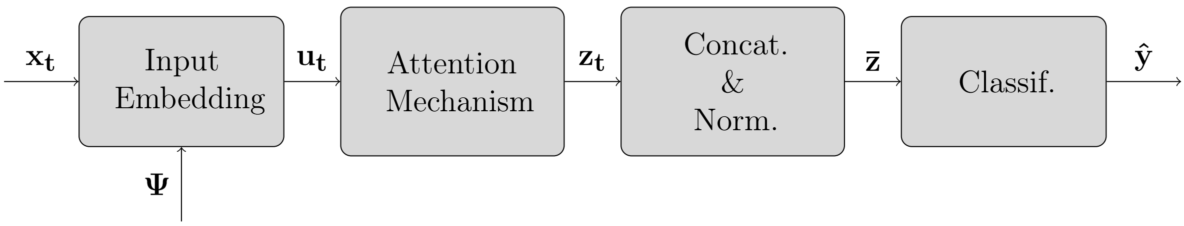


Figure 28 – Generic scheme of TST architecture classification model.

6.2 RESULTS

The state-of-the-art TST model is compared with another time series classification model, LSTM, commonly used in HIF detection (VEERASAMY et al., 2021). The results considering the data quality issues mentioned earlier are presented in Tables 13, 14 and 15 and discussed in the following. Both models consider the same parameters of training/testing, i.e., data with 60 dB of SNR, no lack of data, and synchronism error are used in the training step. The results are presented based on the overall accuracy, i.e, when the dataset contains balanced number of examples in each output class and also, an accuracy considering a binary classification problem, i.e, HIF and non-HIF classes (unbalanced problem).

6.2.1 Impact of Data Quality Issues

The impact of the noise in the classification is presented in Table 13. TST presents a remarkable robustness when dealing with noisy data. The most common scenarios of noise in distribution networks (50-40 dB) do not degenerate the overall accuracy of the classification. High accuracy scores in the multi-class classification are acquired, and, mainly, HIF can be distinguished from other faulty and non-faulty events more than 90% of the time. In more harsh noise scenarios, the classification loses its accuracy significantly; however, even at 20 dB of SNR, more than 60% of the HIFs are correctly classified. This fact is remarkable mainly because HIF is a class of low magnitude voltage profile, and in a 20 dB scenario, most deviations of the phenomena signal are immersed in noise. Moreover, many event classification approaches that investigate the impact of noise affirm that in severe scenarios of SNR, such as 20 dB, the accuracy is strongly depreciated (ZHANG, Y. et al., 2020; GHIGA et al., 2018; ROSCOE et al., 2018).

Table 13 – Impact of SNR in the classification.

Analysis	**Model**							
	TST				LSTM			
SNR (dB)	Overall Acc. (%)	Overall F1-score (%)	HIF Acc. (%)	HIF F1-score (%)	Overall Acc. (%)	Overall F1-score (%)	HIF Acc. (%)	HIF F1-score (%)
50	**100**	**96**	**98**	**97**	83	90	87	86
40	**95**	**89**	**91**	**92**	78	83	80	80
30	**73**	**77**	**75**	**84**	68	77	67	72
20	**67**	**72**	**62**	**70**	55	68	51	65

The impact of missing data in the classification is shown in Table 15. The TST performs a very satisfactory classification considering an extensive range of missing data. Its robustness is immutable up to scenarios of 20% of information loss, and even in the sharpest scenario, 30%, around 85% of the events are correctly classified. Regarding the HIF detection, around 3/4 of the matches are correct, even in the worst-case scenario. TST's superiority over LSTM is significant, mainly when considering an increase in information lost.

Table 14 – Impact of missing data in the classification.

Analysis	**Model**							
	TST				LSTM			
Missing Data (%)	Overall Acc. (%)	Overall F1-score (%)	HIF Acc. (%)	HIF F1-score (%)	Overall Acc. (%)	Overall F1-score (%)	HIF Acc. (%)	HIF F1-score (%)
5	**98**	**95**	**100**	**96**	92	82	93	78
10	**96**	**90**	**98**	**94**	72	77	85	60
20	**93**	**88**	**91**	**90**	56	70	44	51
30	**85**	**86**	**75**	**86**	41	52	20	45

The impact of synchronism error between measurements is presented in Table 15. This data quality issue is the least harmful among the data quality issues. The impact of the number of cycles in which the signal is shifted barely impairs the classification. For most scenarios with TST, over 90% of accuracy in the overall classification is reached, as well as in HIF classification. The LSTM also presents good scores for most scenarios of synchronism error between measurements.

Table 15 – Impact of synchronism error in the classification.

Analysis	**Model**							
	TST				LSTM			
Sync. Error (n^o cyc)	Overall Acc. (%)	Overall F1-score (%)	HIF Acc. (%)	HIF F1-score (%)	Overall Acc. (%)	Overall F1-score (%)	HIF Acc. (%)	HIF F1-score (%)
1	**98**	**97**	**99**	**93**	85	88	96	90
5	**95**	**93**	**96**	**90**	82	82	94	88
10	**91**	**90**	**92**	**85**	80	77	88	86
20	**89**	**89**	**90**	**80**	79	75	85	80

6.2.2 Impact of Topology Change

In order to evaluate the generalization of the TST model when dealing with data obtained from a different topological scheme than the one known in the training process, 100 cases of HIF are simulated at nodes $1, 3, 6, 9$, and 11, when switches S_2 and S_3 are closed (see Figure 5), i.e., when the microgrid is in interconnected mode operation. The impact of the change on the measurements can be observed in Figure 29. The classification reached an accuracy of 98%, presenting a subtle confusion with the class CBS. The results elucidate that the proposed classification model based on combined PMU data is robust enough to classify HIF even with topological changes in the system correctly. It is worth pointing out that the model is trained with radial operation data with 60 dB of SNR, and this validation considers that the interconnected data contained 50 dB of SNR, and no lack of data or synchronism error in the measurements.

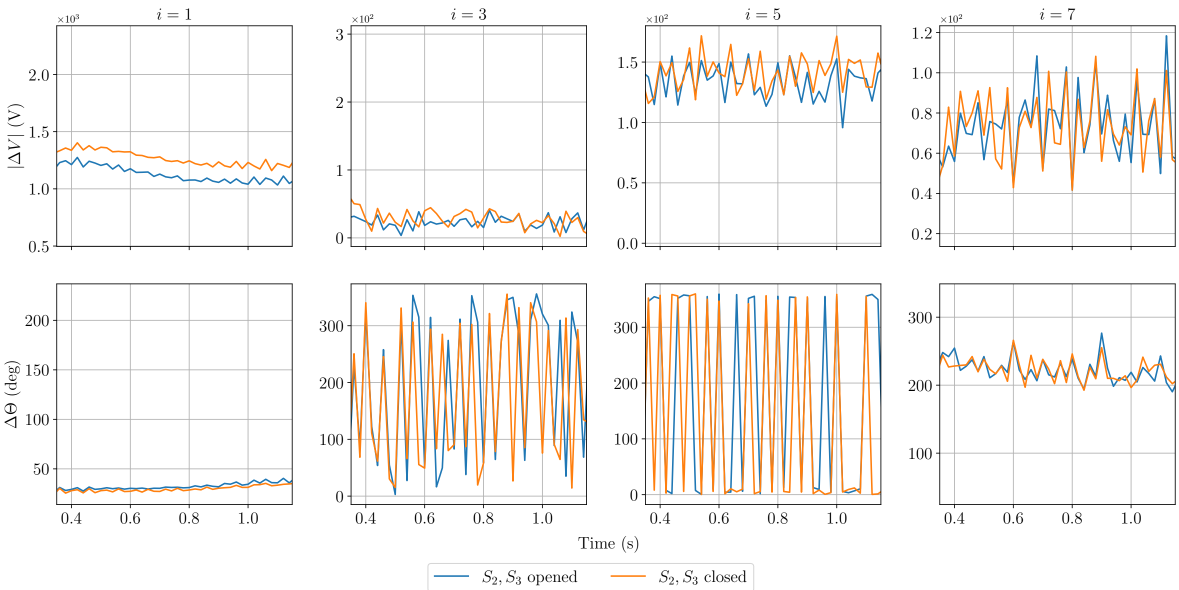


Figure 29 – Impact of changing topology on the measurements - HIF event.

6.2.3 Impact of Real Data

The real measurements are obtained at the substation level and field test level, and combined according to Equation 18. The characteristics of real time series are well accepted by the TST approach, which proves robust when dealing with real data. All the simulated data considers an event's inception based on a fixed time stamp, t, added to a time deviation corresponding to the inception angle, t_{inc}. The real data do not follow this scheme, i.e., the inception of the real HIF is not controlled. TST shows good generality when dealing with a different structure time series. The accuracy reached is around 98%, and the misclassification occurs with class CBS. Combining actual PMU data and obtaining validation from a synthetic-data machine learning model affirms that the approach based on a transformer neural network is robust enough for the most typical applications of microgrid monitoring.

6.2.4 Comparison with Related Works

There are specific gaps in the literature on methods that have adopted PMU measurements as supporters for event classification in microgrid environments. Among the existing approaches, many aspects related to data quality issues are neglected (see Table 16). In this way, a direct comparison with other PMU-data-based approaches is difficult. Therefore, the comparison presented in the sequence takes into account two different regards: the results of our method when compared with other event classification methods in a microgrid context (Table 17(a)) and when compared with others in a time series context (Table 17(b)). Unfortunately, as far as we noticed in recent literature, no method for the microgrid event classification problem employs time series PMU data.

Table 16 – Recent power system event classification approaches and their major characteristics.

Method	**Signal Considered**	**Microgrid Context**	**HIF Context**	**Syncroph. Context**	**Time Series Context**	**Noise Analysis**	**Miss. Data Analysis**	**Sync. Analysis**	**Real Data Analysis**
This work	**V**	✓	✓	✓	✓	✓	✓	✓	✓
2023 [1]	I, V	✓	×	×	×	✓	×	×	×
2023[2]	I, V	×	✓	×	×	×	×	×	×
2023[3]	I, V	✓	✓	×	×	×	×	×	×
2023[4]	I, V	✓	✓	×	×	✓	×	×	×
2022[5]	I, V	×	×	×	✓	✓	✓	×	✓
2021[6]	V	×	×	×	✓	✓	✓	×	✓
2021[7]	I, V	×	×	×	✓	×	✓	×	✓
2020[8]	I, V	✓	✓	×	×	×	×	×	×
2020[9]	I	✓	✓	×	×	✓	×	×	✓
2019[10]	I, V	✓	✓	×	×	✓	×	×	×
2019[11]	I, V	✓	×	×	×	×	×	×	×

The methods presented in Table 17(a) are all feature-based, i.e., the time series are pre-processed to extract meaningful information and then used in the correspondent classification methodology. Most of these methods consider HIF and non-HIF events. However, only some of them clarify binary (HIF and non-HIF) or multi-class (HIF among other events) classification results. In this way, the accuracy presented by the methods is not necessarily related to the HIF classification. So, the best TST results are compared with them. Moreover, most of the compared methods do not perform a significant data quality analysis or consider real data as a validation step of their method. Based on all these aspects, the results of the TST method prove to be relevant enough to a broad context of application in microgrids, although a quantitative comparison here is not the most appropriate, as they are different systems and databases.

[1] (DUA; TYAGI; KUMAR, 2023); [2] (PARAMO; BRETAS; MEYN, 2023), [3] (SOLANKEE; RAI, A.; KIRAR, 2023), [4] (LIU, Yang et al., 2023), [5] (LIU, Yunchuan et al., 2022), [6] (LI, Z. et al., 2021), [7] (YUAN et al., 2021), [8] (ZHANG, Y. et al., 2020), [9] (WEI et al., 2020), [10] (CUI; EL-ARROUDI; WENG, 2019), [11] (SHARMA; SAMANTARAY, 2019)

17(a): Comparison with recent researches on microgrid context.

Method	Event Classification Acc. (%)	Real Data Acc. (%)
This work	**100**	**98**
(DUA; TYAGI; KUMAR, 2023)	100	-
(SOLANKEE; RAI, A.; KIRAR, 2023)	99	-
(LIU, Yang et al., 2023)	99	-
(WEI et al., 2020)	97	91
(ZHANG, Y. et al., 2020)	96	-
(CUI; EL-ARROUDI; WENG, 2019)	99	-

17(b): Comparison with recent researches on time series context.

Method	Event Classification Acc. (%)	Real Data Acc. (%)
This work	**100**	**98**
(LIU, Yunchuan et al., 2022)	94	94
(LI, Z. et al., 2021)	98	98
(YUAN et al., 2021)	99	99

When comparing the TST results with other time series-based methods (Table 17(b)), the context of microgrid is lost once all the compared works are applied at the transmission level and use real data to train the models (equal accuracy in both columns). These methods are aware of data quality issues and perform at least one of the analysis aforementioned. Following the same criteria, the best TST results are compared with them in each approach. Again, it is important to point out that a direct quantitative comparison is not the most appropriate for these cases. Notwithstanding, the TST approach is entirely consistent when considering other state-of-the-art methods to classify events based on PMU data.

7 CONCLUSIONS

This work presented a deep investigation about the impact of using harmonic synchrophasors in event classification, focusing on high impedance faults on microgrids, aiming at demonstrating the usage of PMU data on such applications.

7.1 FEATURES BASED APPROACH

First, it was observed that one phasor per second is sufficient to achieve a satisfactory trade-off between the amount of data and overall accuracy for most tested feature-based machine learning models. Based on the evaluated classic machine learning models, it can be seen that the tree-based architecture presents the best results, in this sense, the choice for the RF model is suitable for improving the classification of the most common microgrids events. An interest aspect about this approach is that all the models do not take into account the usage of the fundamental content of the evaluated signals. This is relevant once the most interest class, HIF, do not manifest any significant variation in the fundamental content. Moreover, it was investigated the impact of both current and voltage measures, magnitude and angle. Using all the synchrophasor information increases the dimensionality of the machine learning models, and not necessarily improves the accuracy of the classification task. Indeed, the best trade-off between the classification's accuracy and the model's complexity is achieved using the odd harmonics contents.

A large set of common events in active distribution networks were considered and the best evaluated arrange of features-model proved to be robust enough in distinguishing HIF from the other events due to the approach based on harmonic synchrophasors and their particular patterns. In addition, the penetration of renewable energy sources on the distribution network and a noise analysis focused on exploring the robustness for two specific noisy conditions, SNR and TVE, are relevant hypotheses and contributions. Moreover, the most subtle evaluated event, HIF, reached a good classification performance when considering data from different system's topologies. The robustness of the approach is also reinforced by evaluating how many PMUs are sufficient to guarantee good results in the classification. When dealing with data quality issues, like loss of synchronism between measures, the classification model still performs well for mild scenarios, however, the feature-based approach suffers when data is lost, severally impacting the classification task. The inclusion of real data in the validation of the proposed analysis also reinforces that the approach presented here is promising and feasible from a practical point of view.

7.2 TIME SERIES APPROACH

Another perspective of solution was investigated throughout the usage of time-series classification models by combining voltage harmonic synchrophasor time series. The strategy proved to be accurate enough to distinguish HIF of other faulty and non-faulty events by using the behavior of odd harmonics. The method also uses low sampling rate PMU data, and by needing only voltage measurements that are easier to acquire, its applicability can be extended to low-cost devices.

Using a state-of-the-art time series classification technique, i.e., the transformer neural network, is quite innovative considering the context of event classification based on PMU data applied to microgrids. Once the attention mechanism within transformer neural networks enables the model to selectively attend to distinct segments of the input, assigning heightened significance to salient features. This attribute proves advantageous in scenarios characterized by missing data, as the model can adeptly navigate through temporal voids, discerning underlying patterns despite temporal omissions. In this sense, this presented method is notably suitable and robust in real-world applications, mainly because it reached relevant outcomes in many scenarios of data quality issues, such as noise, lack of data, and lack of synchronism between measures, mitigating the problems observed in the feature-based approach, which are also underexplored in the literature for distribution systems. Moreover, by validating the approach with real HIF data, an interesting level of generalization can be observed in the solution.

The efforts of this work seek to contribute and enhance the usage of PMU data in the operation of modern active distribution networks. Moreover, some specific investigations can be considered from here, fostering the field of event classification in distribution networks via PMU data.

7.2.1 Future Works

The following topics ascend as promising in the field of event classification using PMU data in distribution networks context.

- To explore the concept of multi-measuring points, creating new features that could represent more details about events;
- To use the concepts of feature engineering to extract the most meaningful information from the large amount of data generated from the multi-measure architectures;
- To extend the approach to more complex systems in order to enhance its generality over scenarios;
- To increase the number of events of interest, like non-linear load switching, as well as different operational conditions of the microgrid, like island operation.

REFERENCES

AALAM, Mir Khadim; SHUBHANGA, KN. Power system event detection and localization—A new approach. **Electric Power Systems Research**, Elsevier, v. 223, p. 109553, 2023.

ALI, Mohd Syukri; ABU BAKAR, Ab Halim; MOKHLIS, Hazlie; AROF, Hamzah; AZIL ILLIAS, Hazlee. High-impedance fault location using matching technique and wavelet transform for underground cable distribution network. **IEEJ Transactions on Electrical and Electronic Engineering**, v. 9, n. 2, p. 176–182, 2014. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/tee.21953`.

ALTAF, Muhammad Waseem; ARIF, Mohammad Taufiqul; ISLAM, Shama Naz; HAQUE, Md. Enamul. Microgrid Protection Challenges and Mitigation Approaches–A Comprehensive Review. **IEEE Access**, v. 10, p. 38895–38922, 2022.

ATENCIA-DE LA OSSA, J.; OROZCO-HENAO, C.; MARÍN-QUINTERO, J. Master-slave strategy based in artificial intelligence for the fault section estimation in active distribution networks and microgrids. **International Journal of Electrical Power & Energy Systems**, v. 148, p. 108923, 2023. ISSN 0142-0615.

AZEROUAL, Mohamed et al. Fault location and detection techniques in power distribution systems with distributed generation: Kenitra City (Morocco) as a case study. **Electric Power Systems Research**, Elsevier, v. 209, p. 108026, 2022.

BAHDANAU, Dzmitry; CHO, Kyunghyun; BENGIO, Yoshua. Neural machine translation by jointly learning to align and translate. **arXiv preprint arXiv:1409.0473**, 2014.

BEVRANI, Hassan; WATANABE, Masayuki; MITANI, Yasunori. **Power system monitoring and control**. [S.l.]: John Wiley & Sons, 2014.

BHATNAGAR, Maanvi; YADAV, Anamika; SWETAPADMA, Aleena. Fuzzy based relaying scheme for high impedance faults in DG integrated distribution system. **Electric Power Systems Research**, v. 202, p. 107602, 2022. ISSN 0378-7796.

BHATTARAI, Bishnu P et al. Big data analytics in smart grids : state-of-the- art , challenges , opportunities , and future directions. v. 2, p. 141–154, 2019.

BISWAL, Monalisa; MISHRA, Manohar; SOOD, Vijay K.; BANSAL, R.C.; ABDELAZIZ, Almoataz Y. Savitzky-Golay Filter integrated matrix pencil method to identify high impedance fault in a renewable penetrated distribution system. **Electric Power Systems Research**, v. 210, p. 108056, Sept. 2022. ISSN 03787796.

BONANOMI, P. Phase Angle measurements with sinchronized clock. **IEEE Transactions on Power Aparatus and Systems**, n. 12, p. 5036–5043, 1981.

BURKOV, A. **The Hundred-page Machine Learning Book**. [S.l.]: Andriy Burkov, 2019. ISBN 9781999579500.

CHAKRABORTY, Soham; DAS, Sarasij. Application of Smart Meters in High Impedance Fault Detection on Distribution Systems. **IEEE Transactions on Smart Grid**, v. 10, n. 3, p. 3465–3473, 2019.

CHANG, Nana; SONG, Guobing; HOU, Junjie; CHANG, Zhongxue. Fault identification method based on unified inverse-time characteristic equation for distribution network. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 146, p. 108734, 2023.

CHOWDHURY, S.; CROSSLEY, P. **Microgrids and Active Distribution Networks**. [S.l.]: Institution of Engineering and Technology, 2009. (Energy Engineering). ISBN 9781849190145.

CIGRE. [S.l.]: Task Force C6.04.02 - Benchmark Systems for Network Integration of Renewable and Distributed Energye Resources, 2014.

CISNEROS-SALDANA, Jorge Ignacio D.; SAMAL, Smrutirekha; BEGOVIC, Miroslav M.; SAMANTARAY, Subhransu Ranjan. On Protection Schemes for AC Microgrids: Challenges and Opportunities. **IEEE Transactions on Industry Applications**, p. 1–12, 2024.

CUI, Qiushi; EL-ARROUDI, Khalil; WENG, Yang. A Feature Selection Method for High Impedance Fault Detection. **IEEE Transactions on Power Delivery**, IEEE, v. 34, n. 3, p. 1203–1215, 2019. ISSN 19374208. eprint: `1808.04454`.

DENG, Xianda; BIAN, Desong; SHI, Di; YAO, Wenxuan; WU, Ling; LIU, Yilu. Impact of low data quality on disturbance triangulation application using high-density PMU measurements. **IEEE Access**, IEEE, v. 7, p. 105054–105061, 2019.

DUA, Gagandeep Singh; TYAGI, Barjeev; KUMAR, Vishal. Fault Detection Technique for Distribution Networks and Microgrids Using Synchrophasor Data. **IEEE Transactions on Industry Applications**, p. 1–14, 2023.

EHSANI, Narges; AMINIFAR, Farrokh; MOHSENIAN-RAD, Hamed. Convolutional autoencoder anomaly detection and classification based on distribution PMU measurements. **IET Generation, Transmission & Distribution**, Wiley Online Library, v. 16, n. 14, p. 2816–2828, 2022.

ELMETWALY, Ahmed Hussain; ELDESOUKY, Azza Ahmed; SALLAM, Abdelhay Ahmed. An Adaptive D-FACTS for Power Quality Enhancement in an Isolated Microgrid. **IEEE Access**, v. 8, p. 57923–57942, 2020.

EMANUEL, A. E.; CYGANSKI, D.; ORR, J. A.; SHILLER, S.; GULACHENSKI, E. M. High impedance fault arcing on sandy soil in 15kV distribution feeders: Contributions to the evaluation of the low frequency spectrum. **IEEE Transactions on Power Delivery**, v. 5, n. 2, p. 676–686, 1990. ISSN 19374208.

FAZAL, Sana; HAQUE, Md Enamul; ARIF, Mohammad Taufiqul; GARGOOM, Ameen; OO, Aman Maung Than. Grid integration impacts and control strategies for renewable based microgrid. **Sustainable Energy Technologies and Assessments**, v. 56, p. 103069, 2023.

FUCHS, Ewald F; MASOUM, Mohammad A S. **Power Quality in Power Systems and Electrical Machines**. [S.l.]: Academic Press, 2008.

FUENTES-VELAZQUEZ, Jose; BELTRAN, Ernesto; BAROCIO, Emilio; ANGELES-CAMACHO, Cesar. A fast automatic detection and classification of voltage magnitude anomalies in distribution network systems using pmu data. **Measurement**, Elsevier, v. 192, p. 110816, 2022.

GALASSI, Andrea; LIPPI, Marco; TORRONI, Paolo. Attention in natural language processing. **IEEE transactions on neural networks and learning systems**, IEEE, v. 32, n. 10, p. 4291–4308, 2020.

GAO, Jian-Hong; GUO, Mou-Fa; LIN, Shuyue; CHEN, Duan-Yu. Application of semantic segmentation in High-Impedance fault diagnosis combined signal envelope and Hilbert marginal spectrum for resonant distribution networks. **Expert Systems with Applications**, v. 231, p. 120631, 2023. ISSN 0957-4174.

GAO, Jie; WANG, Xiaohua; WANG, Xiaowei; YANG, Aijun; YUAN, Huan; WEI, Xiangxiang. A High-Impedance Fault Detection Method for Distribution Systems Based on Empirical Wavelet Transform and Differential Faulty Energy. **IEEE Transactions on Smart Grid**, v. 13, n. 2, p. 900–912, 2022.

GHADERI, Amin; GINN, Herbert L; MOHAMMADPOUR, Hossein Ali. High impedance fault detection: A review. eng. **Electric power systems research**, Elsevier B.V, v. 143, p. 376–388, 2017. ISSN 0378-7796.

GHIGA, Radu; MARTIN, Kenneth; WU, Qiuwei; NIELSEN, Arne Hejde. Phasor Measurement Unit Test Under Interference Conditions. **IEEE Transactions on Power Delivery**, v. 33, n. 2, p. 630–639, 2018.

GHORBANIAN, Maedeh; DOLATABADI, Sarineh Hacopian; SIANO, Pierluigi; MEMBER, Senior. Big Data Issues in Smart Grids : A Survey. IEEE, v. 13, n. 4, p. 4158–4168, 2019.

GOMES, Douglas P.S.; OZANSOY, Cagil; ULHAQ, Anwaar; DE MELO VIEIRA JÚNIOR, José Carlos. The effectiveness of different sampling rates in vegetation high-impedance fault classification. **Electric Power Systems Research**, v. 174, April 2018, 2019. ISSN 03787796.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: MIT Press, 2016. `http://www.deeplearningbook.org`.

GRANDO, Flavio Lori; LAZZARETTI, André Eugênio; MORETO, Miguel. The impact of PMU data precision and accuracy on event classification in distribution systems. **IEEE Transactions on Smart Grid**, IEEE, v. 13, n. 2, p. 1372–1382, 2021.

GUO, Yuanjun; YANG, Zhile; FENG, Shengzhong; HU, Jinxing. Review Article Complex Power System Status Monitoring and Evaluation Using Big Data Platform and Machine Learning Algorithms : A Review and a Case Study. v. 2018, 2018.

HAMANAH, Waleed M; HOSSAIN, Md Ismail; SHAFIULLAH, Md; ABIDO, MA. AC microgrid protection schemes: A comprehensive review. **IEEE Access**, 2023.

HANNAH RITCHIE, Max Roser; ROSADO, Pablo. Energy. **Our World in Data**, 2020. https://ourworldindata.org/energy.

HASSANI, Hossein; HALLAJI, Ehsan; RAZAVI-FAR, Roozbeh; SAIF, Mehrdad. Unsupervised concrete feature selection based on mutual information for diagnosing faults and cyber-attacks in power systems. **Engineering Applications of Artificial Intelligence**, Elsevier Ltd, v. 100, May 2020, p. 104150, 2021. ISSN 09521976.

HOJABRI, Mojgan; DERSCH, Ulrich; PAPAEMMANOUIL, Antonios; BOSSHART, Peter. A comprehensive survey on phasor measurement unit applications in distribution systems. **Energies**, v. 12, n. 23, p. 1–23, 2019. ISSN 19961073.

IEEE Standard for Synchrophasor Measurements for Power Systems. **IEEE Std C37.118.1-2011**, p. 1–61, 2011.

JOSHI, Prachi Mafidar; VERMA, H. K. Synchrophasor measurement applications and optimal PMU placement: A review. **Electric Power Systems Research**, Elsevier B.V., v. 199, June, p. 107428, 2021. ISSN 03787796.

KAVASKAR, Sekar; MOHANTY, Nalin Kant. Detection of High Impedance Fault in Distribution Networks. **Ain Shams Engineering Journal**, v. 10, n. 1, p. 5–13, 2019. ISSN 2090-4479.

KEZUNOVIC, Mladen; PINSON, Pierre; OBRADOVIC, Zoran; GRIJALVA, Santiago; HONG, Tao; BESSA, Ricardo. Big data analytics for future electricity grids. **Electric Power Systems Research**, v. 189, p. 106788, 2020. ISSN 0378-7796.

KHAN, Salman; NASEER, Muzammal; HAYAT, Munawar; ZAMIR, Syed Waqas; KHAN, Fahad Shahbaz; SHAH, Mubarak. Transformers in vision: A survey. **ACM computing surveys (CSUR)**, ACM New York, NY, v. 54, 10s, p. 1–41, 2022.

KHETARPAL, Poras; TRIPATHI, Madan Mohan. A critical and comprehensive review on power quality disturbance detection and classification. **Sustainable Computing: Informatics and Systems**, Elsevier Inc., v. 28, p. 100417, 2020. ISSN 22105379.

KIM, C. J.; DON RUSSELL, B. Harmonic behavior during arcing faults on power distribution feeders. **Electric Power Systems Research**, v. 14, n. 3, p. 219–225, 1988. ISSN 03787796.

KUMMEROW, Andre; MONSALVE, Cristian; BROSINSKY, Christoph; NICOLAI, Steffen; WESTERMANN, Dirk. A novel framework for synchrophasor based online recognition and efficient post-mortem analysis of disturbances in power systems. **Applied Sciences (Switzerland)**, v. 10, n. 15, 2020. ISSN 20763417.

LEDESMA, Jorge Javier Gimenez; DO NASCIMENTO, Kelvin Bryan; DE ARAUJO, Leandro Ramos; PENIDO, Débora Rosana Ribeiro. A two-level ANN-based method using synchronized measurements to locate high-impedance fault in distribution systems. **Electric Power Systems Research**, v. 188, p. 106576, 2020. ISSN 0378-7796.

LI, Zikang; LIU, Hao; ZHAO, Junbo; BI, Tianshu; YANG, Qixun. A power system disturbance classification method robust to PMU data quality issues. **IEEE Transactions on Industrial Informatics**, IEEE, v. 18, n. 1, p. 130–142, 2021.

LIAO, Anna L.; STEWART, Emma M.; KARA, Emre C. Micro-synchrophasor data for diagnosis of transmission and distribution level events. **Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference**, 2016-July, p. 0–4, 2016. ISSN 21608563.

LIMA, Érica Mangueira; BRITO, Núbia Silva Dantas; SOUZA, Benemar Alencar de. High impedance fault detection based on Stockwell transform and third harmonic current phase angle. **Electric Power Systems Research**, v. 175, p. 105931, 2019. ISSN 0378-7796.

LIU, Yang; ZHAO, Yanlei; WANG, Lei; FANG, Chen; XIE, Bangpeng; CUI, Laixi. High-impedance Fault Detection Method Based on Feature Extraction and Synchronous Data Divergence Discrimination in Distribution Networks. **Journal of Modern Power Systems and Clean Energy**, v. 11, n. 4, p. 1235–1246, 2023.

LIU, Yikui; WU, Lei; LI, Jie. D-PMU based applications for emerging active distribution systems: A review. **Electric Power Systems Research**, Elsevier, v. 179, November 2019, p. 106063, 2020. ISSN 03787796.

LIU, Yunchuan; YANG, Lei; GHASEMKHANI, Amir; LIVANI, Hanif; CENTENO, Virgilio A; CHEN, Pin-Yu; ZHANG, Junshan. Robust event classification using imperfect real-world PMU data. **IEEE Internet of Things Journal**, IEEE, v. 10, n. 9, p. 7429–7438, 2022.

LOPES, G.N.; MENEZES, T.S.; SANTOS, G.G.; TRONDOLI, L.H.P.C.; VIEIRA, J.C.M. High Impedance Fault detection based on harmonic energy variation via

S-transform. **International Journal of Electrical Power & Energy Systems**, v. 136, p. 107681, 2022. ISSN 0142-0615.

LOPES, Gabriela N; MENEZES, Thiago S; GOMES, Douglas PS; VIEIRA, Jose Carlos M. High Impedance Fault Location Methods: Review and Harmonic Selection-based Analysis. **IEEE Open Access Journal of Power and Energy**, IEEE, 2023.

MACEDO, José Rubens; RESENDE, José Wilson; BISSOCHI JR, Carlos Augusto; CARVALHO, Daniel; CASTRO, Fernando C. Proposition of an interharmonic-based methodology for high-impedance fault detection in distribution systems. **IET Generation, Transmission & Distribution**, v. 9, n. 16, p. 2593–2601, 2015.

MIRANDA, Vladimiro; A. CARDOSO, Pedro; BESSA, Ricardo J.; DECKER, Ildemar. Through the looking glass: Seeing events in power systems dynamics. **International Journal of Electrical Power and Energy Systems**, Elsevier, v. 106, July 2018, p. 411–419, 2019. ISSN 01420615.

MISSOUT, G. Dynamic measurement of the absolute voltage angle on long transmission lines. **IEEE Transactions on Power Aparatus and Systems**, n. 11, p. 4428–4434, 1981.

MOHAMED, N. Detection of High Impedance Faults in Medium Voltage Distribution Networks Using Discrete Wavelet Transform. **IET Conference Proceedings**, Institution of Engineering and Technology, 0202–0202(1), 2013.

NGUYEN, Bang LH; VU, Tuyen V; NGUYEN, Thai-Thanh; PANWAR, Mayank; HOVSAPIAN, Rob. Spatial-Temporal Recurrent Graph Neural Networks for Fault Diagnostics in Power Distribution Systems. **IEEE Access**, IEEE, 2023.

OLIVEIRA, Roger Alves de; BOLLEN, Math H.J. Deep learning for power quality. **Electric Power Systems Research**, v. 214, p. 108887, 2023. ISSN 0378-7796.

OUBRAHIM, Zakarya; AMIRAT, Yassine; BENBOUZID, Mohamed; OUASSAID, Mohammed. Power Quality Disturbances Characterization Using Signal Processing and Pattern Recognition Techniques: A Comprehensive Review. **Energies**, v. 16, n. 6, p. 2685, 2023.

PARAMO, Gian; BRETAS, Arturo; MEYN, Sean. High-Impedance Non-Linear Fault Detection via Eigenvalue Analysis with low PMU Sampling Rates. In: 2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). [S.l.: s.n.], 2023. P. 1–5.

PAVLOVSKI, Martin; ALQUDAH, Mohammad; DOKIC, Tatjana; HAI, Ameen Abdel; KEZUNOVIC, Mladen; OBRADOVIC, Zoran. Hierarchical convolutional neural networks for event classification on PMU measurements. **IEEE Transactions on Instrumentation and Measurement**, IEEE, v. 70, p. 1–13, 2021.

PHADKE, A.G.; THORP, J.S. **Synchronized Phasor Measurements and their Applications**. [S.l.: s.n.], 2008. P. 246. ISBN 978-0-387-76535-8.

RAHMAN FAHIM, Shahriar; K. SARKER, Subrata; MUYEEN, S. M.; SHEIKH, Md. Rafiqul Islam; DAS, Sajal K. Microgrid Fault Detection and Classification: Machine Learning Based Approach, Comparison, and Reviews. **Energies**, v. 13, n. 13, 2020. ISSN 1996-1073.

RAI, Khushwant; HOJATPANAH, Farnam; BADRKHANI AJAEI, Firouz; GROLINGER, Katarina. Deep Learning for High-Impedance Fault Detection: Convolutional Autoencoders. **Energies**, v. 14, n. 12, p. 3623, 2021.

RAI, Praveen; LONDHE, Narendra D; RAJ, Ritesh. Fault classification in power system distribution network integrated with distributed generators using CNN. **Electric Power Systems Research**, Elsevier, v. 192, p. 106914, 2021.

RASCHKA, Sebastian; LIU, Yuxi Hayden; MIRJALILI, Vahid. **Machine Learning with PyTorch and Scikit-Learn Develop machine learning and deep learning models with Python**. [S.l.: s.n.], 2022. P. 771. ISBN 9781801819312.

REZAEIEH, Mohammad-Reza Hosseinkhani; BOLANDI, Tohid Ghanizadeh; JALALAT, Sajjad Miralizadeh. A novel approach for resilient protection of AC microgrid based on differential phase angle of superimposed complex power. **Sustainable Energy, Grids and Networks**, v. 34, p. 101024, 2023.

ROSCOE, Andrew J.; BLAIR, Steven M.; DICKERSON, Bill; RIETVELD, Gert. Dealing With Front-End White Noise on Differentiated Measurements Such as Frequency and ROCOF in Power Systems. **IEEE Transactions on Instrumentation and Measurement**, v. 67, n. 11, p. 2579–2591, 2018.

ROUTRAY, P.; MISHRA, M.; ROUT, P.K. High Impedance Fault detection in radial distribution system using S-Transform and neural network. In: 2015 IEEE Power, Communication and Information Technology Conference (PCITC). [S.l.: s.n.], 2015. P. 545–551.

SARWAR, Muhammad; MEHMOOD, Faisal; ABID, Muhammad; KHAN, Abdul Qayyum; GUL, Sufi Tabassum; KHAN, Adil Sarwar. High impedance fault detection and isolation in power distribution networks using support vector machines. **Journal of King Saud University - Engineering Sciences**, v. 32, n. 8, p. 524–535, 2020. ISSN 1018-3639.

SEKAR, Kavaskar; MOHANTY, Nalin Kant. Combined Mathematical Morphology and Data Mining Based High Impedance Fault Detection. **Energy Procedia**, v. 117, p. 417–423, 2017. ISSN 1876-6102.

SEN, Sachidananda; KUMAR, Vishal. Microgrid control: A comprehensive survey. **Annual Reviews in Control**, v. 45, p. 118–151, 2018. ISSN 13675788.

SHADI, Mohammad Reza; AMELI, Mohammad-Taghi; AZAD, Sasan. A real-time hierarchical framework for fault detection, classification, and location in power systems using PMUs data and deep learning. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 134, p. 107399, 2022.

SHAHSAVARI, Alireza; FARAJOLLAHI, Mohammad; STEWART, Emma M.; CORTEZ, Ed; MOHSENIAN-RAD, Hamed. Situational Awareness in Distribution Grid Using Micro-PMU Data: A Machine Learning Approach. **IEEE Transactions on Smart Grid**, v. 10, n. 6, p. 6167–6177, 2019.

SHAHZAD, Sulman; ABBASI, Muhammad Abbas; ALI, Hassan; IQBAL, Muhammad; MUNIR, Rania; KILIC, Heybet. Possibilities, Challenges, and Future Opportunities of Microgrids: A Review. **Sustainability**, MDPI, v. 15, n. 8, p. 6366, 2023.

SHARMA, N; SAMANTARAY, S. Assessment of PMU-based wide-area angle criterion for fault detection in microgrid. **IET Generation, Transmission and Distribution**, v. 13, n. 19, p. 4301–4310, 2019. ISSN 17518687.

SHAW, Priyabrata; JENA, Manas Kumar. A Novel Event Detection and Classification Scheme Using Wide Area Frequency Measurements. **IEEE Transactions on Smart Grid**, v. 3053, n. 100, p. 1–11, 2020. ISSN 19493061.

SILVA, Sergio; COSTA, Pyramo; GOUVEA, Maury; LACERDA, Alcyr; ALVES, Franciele; LEITE, Daniel. High impedance fault detection in power distribution systems using wavelet transform and evolving neural network. **Electric Power Systems Research**, v. 154, p. 474–483, 2018. ISSN 0378-7796.

SIROJAN, Tharmakulasingam; LU, Shibo; PHUNG, BT; ZHANG, Daming; AMBIKAIRAJAH, Eliathamby. Sustainable deep learning at grid edge for real-time high impedance fault detection. **IEEE Transactions on Sustainable Computing**, IEEE, v. 7, n. 2, p. 346–357, 2018.

SOHEILI, Adel; SADEH, Javad; BAKHSHI, Reza. Modified FFT based high impedance fault detection technique considering distribution non-linear loads: Simulation and experimental data analysis. **International Journal of Electrical Power & Energy Systems**, v. 94, p. 124–140, 2018. ISSN 0142-0615.

SOHEILI, Adel; SADEH, Javad; LOMEI, Hadi; MUTTAQI, Kashem. A new high impedance fault detection scheme: Fourier based approach. In: 2016 IEEE International Conference on Power System Technology (POWERCON). [S.l.: s.n.], 2016. P. 1–6.

SOLANKEE, Laxman; RAI, Avinash; KIRAR, Mukesh. High Impedance Fault Detection in Microgrid to Enhance Resiliency Against PMU Outage. **International Journal of Computing and Digital Systems**, University of Bahrain, v. 14, n. 1, p. 1–xx, 2023.

SOUSA CARVALHO, José Genilson; ROCHA ALMEIDA, Aryfrance; FERREIRA, Danton Diego; DOS SANTOS, Bartolomeu Ferreira; PEREIRA VASCONCELOS, Luis Henrique; DE OLIVEIRA SOBREIRA, Danilo. High-impedance fault modeling and classification in power distribution networks. **Electric Power Systems Research**, v. 204, p. 107676, 2021. ISSN 0378-7796.

STEFANIDOU-VOZIKI, P; SAPOUNTZOGLOU, N; RAISON, B; DOMINGUEZ-GARCIA, J L. A review of fault location and classification methods in distribution grids. **Electric Power Systems Research**, Elsevier B.V., v. 209, April, p. 108031, 2022. ISSN 0378-7796.

SUSTO, Gian Antonio; CENEDESE, Angelo; TERZI, Matteo. Chapter 9 - Time-Series Classification Methods: Review and Applications to Power Systems Data. In: BIG Data Application in Power Systems. [S.l.]: Elsevier, 2018. P. 179–220. ISBN 978-0-12-811968-6.

THERON, J. C.Jacobus; PAL, Amit; VARGHESE, Abraham. Tutorial on high impedance fault detection. **71st Annual Conference for Protective Relay Engineers, CPRE 2018**, 2018-Janua, p. 1–23, 2018.

THOMAS, Jibin B; CHAUDHARI, Saurabh G; SHIHABUDHEEN, KV; VERMA, Nishchal K. CNN-based transformer model for fault detection in power system networks. **IEEE Transactions on Instrumentation and Measurement**, IEEE, v. 72, p. 1–10, 2023.

THOMAS, Jibin B; SHIHABUDHEEN, KV. Neural architecture search algorithm to optimize deep transformer model for fault detection in electrical power distribution systems. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 120, p. 105890, 2023.

TU, Chunming; HE, Xi; SHUAI, Zhikang; JIANG, Fei. Big data issues in smart grid – A review. **Renewable and Sustainable Energy Reviews**, v. 79, May, p. 1099–1107, 2017. ISSN 18790690.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N; KAISER, Łukasz; POLOSUKHIN, Illia. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

VEERASAMY, Veerapandiyan; ABDUL WAHAB, Noor Izzri; RAMACHANDRAN, Rajeswari; MANSOOR, Muhammad; THIRUMENI, Mariammal; OTHMAN, Mohammad Lutfi. High impedance fault detection in medium voltage distribution network using discrete wavelet transform and adaptive neuro-fuzzy inference system. **Energies**, v. 11, n. 12, 2018. ISSN 19961073.

VEERASAMY, Veerapandiyan; WAHAB, Noor Izzri Abdul; OTHMAN, Mohammad Lutfi; PADMANABAN, Sanjeevikumar; SEKAR, Kavaskar; RAMACHANDRAN, Rajeswari; HIZAM, Hashim; VINAYAGAM, Arangarajan; ISLAM, Mohammad Zohrul. LSTM Recurrent Neural Network Classifier for High Impedance Fault Detection in Solar PV Integrated Power System. **IEEE Access**, v. 9, p. 32672–32687, 2021.

VINAYAGAM, Arangarajan; VEERASAMY, Veerapandiyan; TARIQ, Mohd; AZIZ, Asma. Heterogeneous learning method of ensemble classifiers for identification and classification of power quality events and fault transients in wind power integrated microgrid. **Sustainable Energy, Grids and Networks**, v. 31, p. 100752, 2022. ISSN 2352-4677.

VON MEIER, Alexandra; STEWART, Emma; MCEACHERN, Alex; ANDERSEN, Michael; MEHRMANESH, Laura. Precision Micro-Synchrophasors for Distribution Systems: A Summary of Applications. **IEEE Transactions on Smart Grid**, v. 8, n. 6, p. 2926–2936, 2017. ISSN 19493053.

WANG, Shaorui; LU, Tianguang; HAO, Ran; WANG, Fei; DING, Tao; LI, Jing; HE, Xueqian; GUO, Yu; HAN, Xueshan. An Identification Method for Anomaly Types of Active Distribution Network Based on Data Mining. **IEEE Transactions on Power Systems**, p. 1–11, 2023.

WANG, Shiyuan; DEHGHANIAN, Payman. On the Use of Artificial Intelligence for High Impedance Fault Detection and Electrical Safety. **IEEE Transactions on Industry Applications**, v. 56, n. 6, p. 7208–7216, 2020.

WEI, Mingjie; SHI, Fang; ZHANG, Hengxu; JIN, Zongshuai; TERZIJA, Vladimir; ZHOU, Jian; BAO, Hailong. High Impedance Arc Fault Detection Based on the Harmonic Randomness and Waveform Distortion in the Distribution System. **IEEE Transactions on Power Delivery**, v. 35, n. 2, p. 837–850, 2020.

WU, Jia; CHEN, Xiu Yun; ZHANG, Hao; XIONG, Li Dong; LEI, Hang; DENG, Si Hao. Hyperparameter optimization for machine learning models based on Bayesian optimization. **Journal of Electronic Science and Technology**, Elsevier B.V., v. 17, n. 1, p. 26–40, 2019. ISSN 1674862X.

YANG, Zhiwei; LIU, Hao; BI, Tianshu; LI, Zikang; YANG, Qixun. An adaptive PMU missing data recovery method. **International Journal of Electrical Power & Energy Systems**, v. 116, p. 105577, 2020.

YAO, Wenxuan et al. Impact of GPS signal loss and its mitigation in power system synchronized measurement devices. **IEEE Transactions on Smart Grid**, IEEE, v. 9, n. 2, p. 1141–1149, 2016.

YE, Zong-Jhen; FARAJOLLAHI, Mohammad; MOHSENIAN-RAD, Hamed. Impact Analysis and Mitigation of Losing Time Synchronization at Micro-PMUs in Event Location Identification. In: 2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). [S.l.: s.n.], 2022. P. 1–5.

YUAN, Yuxuan; GUO, Yifei; DEHGHANPOUR, Kaveh; WANG, Zhaoyu; WANG, Yanchao. Learning-based real-time event identification using rich real PMU data. **IEEE Transactions on Power Systems**, IEEE, v. 36, n. 6, p. 5044–5055, 2021.

YUAN, Yuxuan; WANG, Zhaoyu; WANG, Yanchao. Learning latent interactions for event classification via graph neural networks and PMU data. **IEEE Transactions on Power Systems**, IEEE, v. 38, n. 1, p. 617–629, 2022.

ZERVEAS, George; JAYARAMAN, Srideepika; PATEL, Dhaval; BHAMIDIPATY, Anuradha; EICKHOFF, Carsten. A Transformer-Based Framework for Multivariate Time Series Representation Learning. In: PROCEEDINGS of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York, NY, USA: Association for Computing Machinery, 2021.

ZHANG, Junhao; TANG, Lu; MINGOTTI, Alessandro; PERETTO, Lorenzo; WEN, He. Analysis of White Noise on Power Frequency Estimation by DFT-Based Frequency Shifting and Filtering Algorithm. **IEEE Transactions on Instrumentation and Measurement**, v. 69, n. 7, p. 4125–4133, 2020.

ZHANG, Yongjie; WANG, Xiaojun; LUO, Yiping; XU, Yin; HE, Jinghan; WU, Guohong. A CNN Based Transfer Learning Method for High Impedance Fault Detection. In: 2020 IEEE Power & Energy Society General Meeting (PESGM). [S.l.: s.n.], 2020. P. 1–5.

ZHENG, Jianqiao; RAMASINGHE, Sameera; LUCEY, Simon. Rethinking positional encoding. **arXiv preprint arXiv:2107.02561**, 2021.

APPENDIX A – CONSIDERATIONS ABOUT ANGLE ESTIMATION

The estimated angle waveforms have been corrected in order to set theirs values to the interval [0° – 360°], once the angular outcome of the estimation is contained in the interval of ±180°. The phasor of a i_{th} harmonic obtained from the estimation process is defined by two components: X_{ic} and X_{is}, whenever at least one of these components are close to the edge of the quadrants, the next estimated phasor can be dragged to the adjacent quadrant. This fact is observed as a discontinuity in the angular time series signal and an example of the neighborhood where it happens is depicted by the red region in Figure 30.

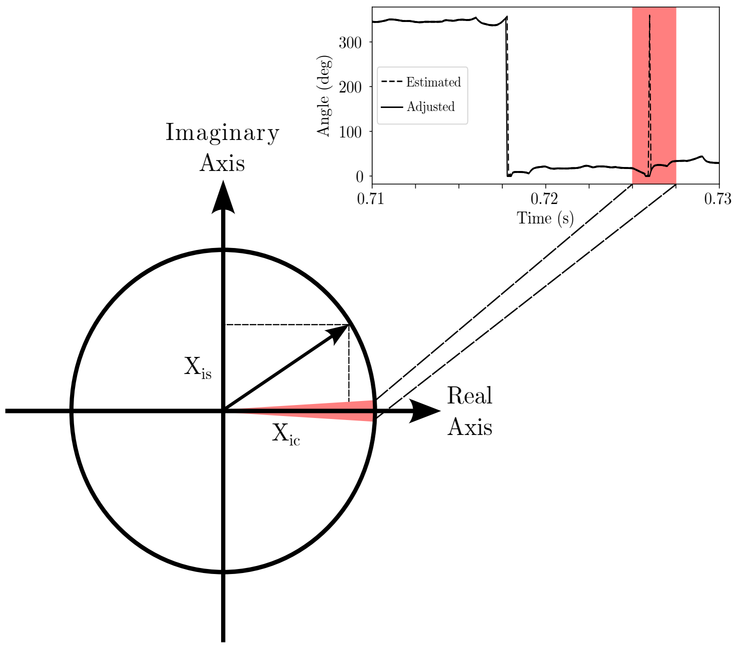


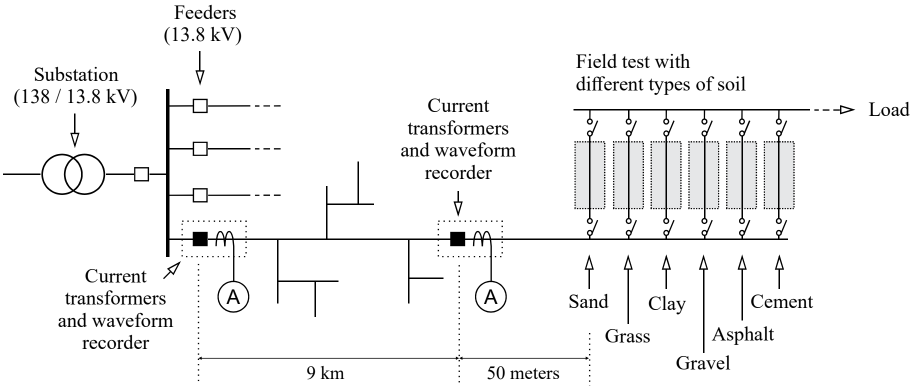
Figure 30 – Angle discontinuity visualization.

Once these discontinuities may harm the application of the data, some corrections can be performed. As example, by successive experiments, it is observed that small angles may regularly cause this discontinuity and then, a threshold could be considered in order to mitigate this behavior. Moreover, the discontinuity problem can be smoothed by applying some signal processing technique, like moving average techniques.

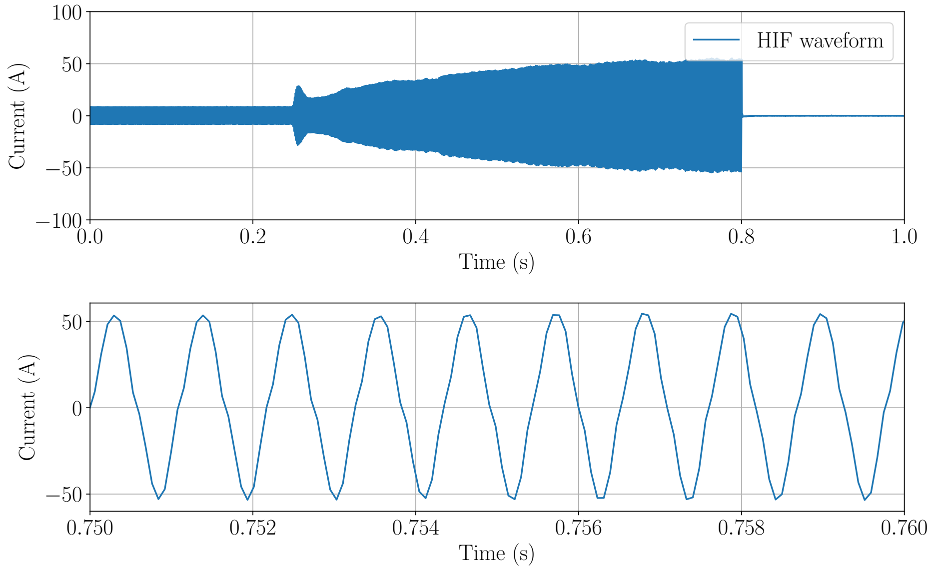
APPENDIX B – REAL HIGH IMPEDANCE FAULT DATA

Commonly, real data of electric networks are private and managed by utilities, i.e, their obtaining is not always possible. Fortunately, through a collaboration, all the results achieved in this thesis are validated with real data from high impedance faults obtained by (MACEDO et al., 2015) in a field test 13.8 kV distribution system where different soil types are tested, as depicted in Figure 31a. A detail of a HIF waveform is presented in Figure 31b, and the waveforms presented in Figure 31c are a comparison from HIFs of the used simulated model and real data from different soil types. These soil scenarios are specially developed for HIF identification purposes. The system counts with two-meter devices with GPS synchronization, one at the substation and one at the field test site, measuring with a sampling rate of 1024 samples per cycle. Two types of HIFs are considered on tests: faults that do not consider the rupture of the conductor and faults that do.

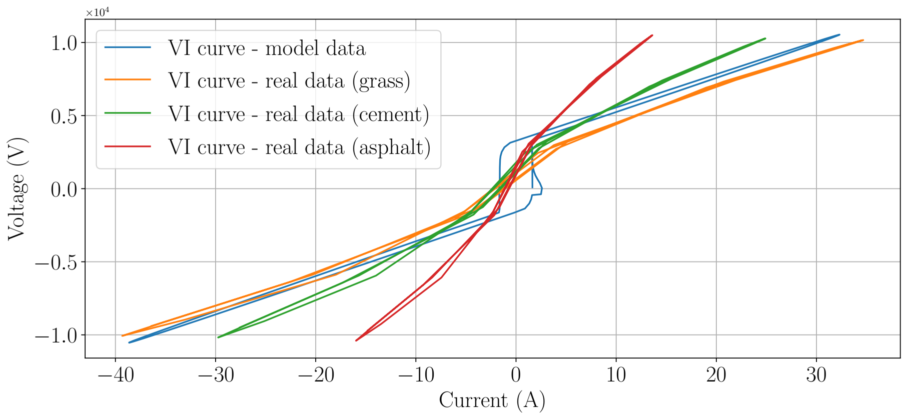
These real data are used in this work as a validation procedure, i.e., the classification training occurs with simulated data, as described before, and is validated with real data from a different electric network. All the real data is processed according to the strategies used with simulated data, i.e., harmonic phasor computing. So, considering that the real data contemplates many soil scenarios similar to the simulated data, it is possible to confirm that the real data is suitable for validating the classification model. —



(a) Diagram of the distribution system and the field tests.



(b) Behavior of a real HIF.



(c) VI characteristics: data of HIF model (CUI; EL-ARROUDI; WENG, 2019), and real HIF (MACEDO et al., 2015).

Figure 31 – Test system and real HIF waveforms.