Faculdade de Engenharia da Universidade do Porto

Aplicação de técnicas de aprendizagem profunda estruturada para diagnóstico de funcionamento de centrais fotovoltaicas.

David da Silva Moreira Freire

Iniciação à Investigação

Mestrado em Engenharia Eletrotécnica e Computadores

Orientador: Cláudio Domingos Martins Monteiro

January 9, 2023

Aplicação de técnicas de aprendizagem profunda estruturada para diagnóstico de funcionamento de centrais fotovoltaicas.

David da Silva Moreira Freire

MESTRADO EM ENGENHARIA ELETROTÉCNICA E COMPUTADORES

January 9, 2023

Resumo

A proliferação de centrais fotovoltaicas de dimensão industrial leva à necessidade de métodos para detetar e classificar falhas nos seus componentes, sendo que estas que podem ter impactos económicos significativos. Neste trabalho, o estado da arte das ferramentas de deteção de falhas e estimação do estado aplicadas ao campo dos sistemas PV será explorado, com foco na compreensão do seu funcionamento, identificando-se pontos fortes e possíveis limitações. Conclui-se que os métodos estatísticos não são comummente utilizados nas ferramentas modernas. Ainda assim, já foi testada a implementação de diversos domínios para solucionar este tipo de problema, desde teoria dos grafos a processamento de sinal, aprendizagem profunda e aprendizagem quântica. Serão propostas melhorias às abordagens existentes ou desenvolvida uma nova abordagem para abordar esta problemática. Com a inspeção das ferramentas mais bem sucedidas até à data e pela potencial oferta de uma nova abordagem, o objetivo deste trabalho é fornecer aos operadores de instalações fotovoltaicas um aumento na fiabilidade e eficiência dos seus sistemas.

Abstract

The increase in utility-scale photovoltaic power plants has led to the need for effective methods for detecting and classifying component faults, which can have significant economic impacts. This work assesses the current state of fault detection and state estimation tools in the field of PV systems, focusing on understanding how these tools work and identifying their strengths and limitations. It is concluded that statistical methods are not commonly used on modern tools, while machine learning makes up the majority of state-of-the-art fault detection and classification algorithms. Still, many fields have been tested for this problem, from graph theory to signal processing, deep learning, and quantum machine learning. Consequently, this work proposes improvements to existing approaches or a novel technique developed to address this issue. By examining the most successful tools to date and offering new solutions, the intention is to help PV plant operators improve the reliability and efficiency of their systems. Also, it's expected that the developed methodology can become a generalistic data cohesion algorithm, positively impacting other data-driven problems.

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

CNN	Convolutional Neural Network
CXN	Cell Complex Neural Network
DC	Direct current
DL	Deep Learning
DNN	Deep Neural Network
LSTM	Long short-term memory
MCD	Minimum Covariance Determinant
ML	Machine Learning
PV	Photovoltaic
RBFNN	Radial basis function neural network
RMM	Recurrent Neural Network
SRC	Sparse Representation Classifier
STC	Standard Test Conditions
SVM	Support Vector Machine

Chapter 1

Introduction

The XIX century marked a significant shift in the world's perception of energy resources as the desire to invest in renewable energy sources to power modern societies grew. This transition was driven by the need to reduce dependency on fossil fuels, mitigate the effects of global warming, and slow climate change. Renewable energy sources offer a range of benefits, including reduced greenhouse gas emissions, improved air quality, and increased energy security. Solar photovoltaic energy is a desirable renewable energy source due to its abundance, accessibility, and environmental benefits. While solar photovoltaic energy has proven to be both cost-efficient and environmentally friendly, it also comes with unprecedented challenges, such as its intermittent nature, low electrical inertia, complex forecasting, and geographic-dependent operating conditions. Despite these challenges, recent reports [1] show that the economic benefits of investing in renewable energy outweigh the complications, as there is an increasing global investment trend in these sources.

The general construction of PV farms, particularly on the utility-scale, has led to a need for effective maintenance and monitoring to ensure maximum efficiency and operational reliability. Towards this, various algorithms and routines are used to monitor the state of PV farms and identify any potential issues that may arise. Fault detection is crucial to this process, allowing PV farm operators to identify and address problems quickly. Detecting faults and identifying the necessary steps can prevent or minimize downtime and ensure optimal performance. Given the importance of maintaining high levels of operation, knowing if action is needed to restore or fix components from an anomalous scenario is desirable for reducing investment risk and maximizing profits.

Integrating intermittent energy resources into modern electric grids has led to stricter requirements for connecting such power systems to ensure safe grid operating conditions. As a result, companies that own or plan to build photovoltaic farms must comply with these requirements and have adequate power electronics and monitoring/control capabilities. Failure to meet these requirements can result in sanctions or fines for the responsible party, as well as potential impacts on system availability, asset value, and disturbance propagation to the grid. To minimize these risks and maximize the value of their assets, companies may opt to implement fault detection and state estimation tools. These tools allow for the early detection and resolution of potential issues and can prevent or minimize downtime. The need to create or improve existing fault detection and state estimation tools, and the search for the most effective methodologies for addressing these issues, drive research in this field.

Having laid the basis for why there must be system behavior assessment in utility-scale PV plants, it is necessary to understand what business concepts are crucial to this field. In the course of this work, the presented topics will go over the following questions:

- What components mostly fail in photovoltaic power systems?
- What is the average frequency of faults?
- What fault detection/state estimation tools exist for photovoltaic power systems?
- What are the most successful ones?
- What's their structure? Are they mostly centralized or decentralized?
- What are their computational costs/efficiency?
- What is the expected magnitude of precision and confidence?
- Which key performance indicators can evaluate the success of these tools?
- What are their implementation difficulties?

With these questions uncovered, the main objective is to adapt or design a novel algorithm/approach to fault detection based on modern artificial intelligence solutions. However, this can be split into finer goals:

- Identify and study existing fault detection tools for photovoltaic power systems.
- Adapt or develop a new tool.
- Apply and test the new tool in real case study PV assets.
- Validate the developed methodologies by comparison to reference tools.

Before reviewing state-of-the-art fault detection tools, types of failures in photovoltaic systems need to be understood: find which components usually fail, which ones fail more often, and how often. For this, it is necessary to understand such components' physical and electrical properties and the modeling techniques used to characterize them. There will be an assessment of utility-scale power plants architecture through literature, alongside the detection objective of state-of-the-art fault detection tools applied in this field. Then, there shall be an extensive analysis and review of what tools have been designed and used in this field. In this step, critical evaluation of the literature is a must for understanding the tool's scope, ease of implementation, and understanding that the data sets available for this work are compatible. Having selected the most prominent ones, they're to be qualitatively and quantitatively compared to each other in their application context so that the results allow objective evaluations. This process requires implementing these tools, following the guidelines in the respective article/book/report, verifying their metrics, and checking if the achieved results resemble the same as the literature suggests. It will require gathering data sets, which can either be artificially generated through simulation or provided by an enterprise that services photovoltaic plant owners.

There's a desire that, in the end, the developed work helps achieve an improved method for fault detection and state estimation in photovoltaic power systems, resulting in a production-ready software application agile enough to deploy for multiple PV assets. It's intended that the algorithm specializes in data cohesion as a means of anomaly inference, allowing asynchronous and self-healing data transfers between the considered components. Depending on the new algorithm's characteristics, it could result in an approach capable of generalization and application to other engineering systems, benefiting more than just PV systems. No matter the chosen methodology, fault detection will, in most cases, result in an economic benefit, catastrophe prevention, and safety increase.

Chapter 2

Fault detection in Utility Scale Photovoltaic Plants

2.1 Utility-Scale Photovoltaic System's Architecture

Utility-scale photovoltaic (PV) power plants are large-scale systems connected to the electrical grid, having installed capacities ranging from kilowatts peak (kWp) to megawatts peak (MWp). These systems typically consist of many PV panels interconnected through power electronics to aggregate and inject power into the grid. The number and type of components in a PV power plant depend on the plant's scale and topology, with different configurations possible for large-scale applications, including central inverters, string inverters, and multi-string inverters [2]. The physical installation of PV modules can include solar tracking apparatuses, such as single and dual-axis trackers [3], which add to system complexity and change production behavior. Understanding the architecture and components of PV power plants is vital for designing, operating, and maintaining these systems, as it helps optimize their performance and reliability.

Figure 2.1 presents a typical utility-scale PV plant architecture using the central inverter (or possibly multi-string inverter) configuration. It is noticeable that many system components may fail in one or more ways, which is why monitoring and fault detection algorithms are essential to maintain state estimation. The main subsystems considered in this work are the following:

- Solar photovoltaic panels (with or without bypass diodes).
- Tracking mount.
- Electrical cabling.
- Inverter(s) (mostly with Max Power Point Trackers).
- AC Transformer(s).
- Protection components (circuit breakers, fuses, surge protectors, etc.)

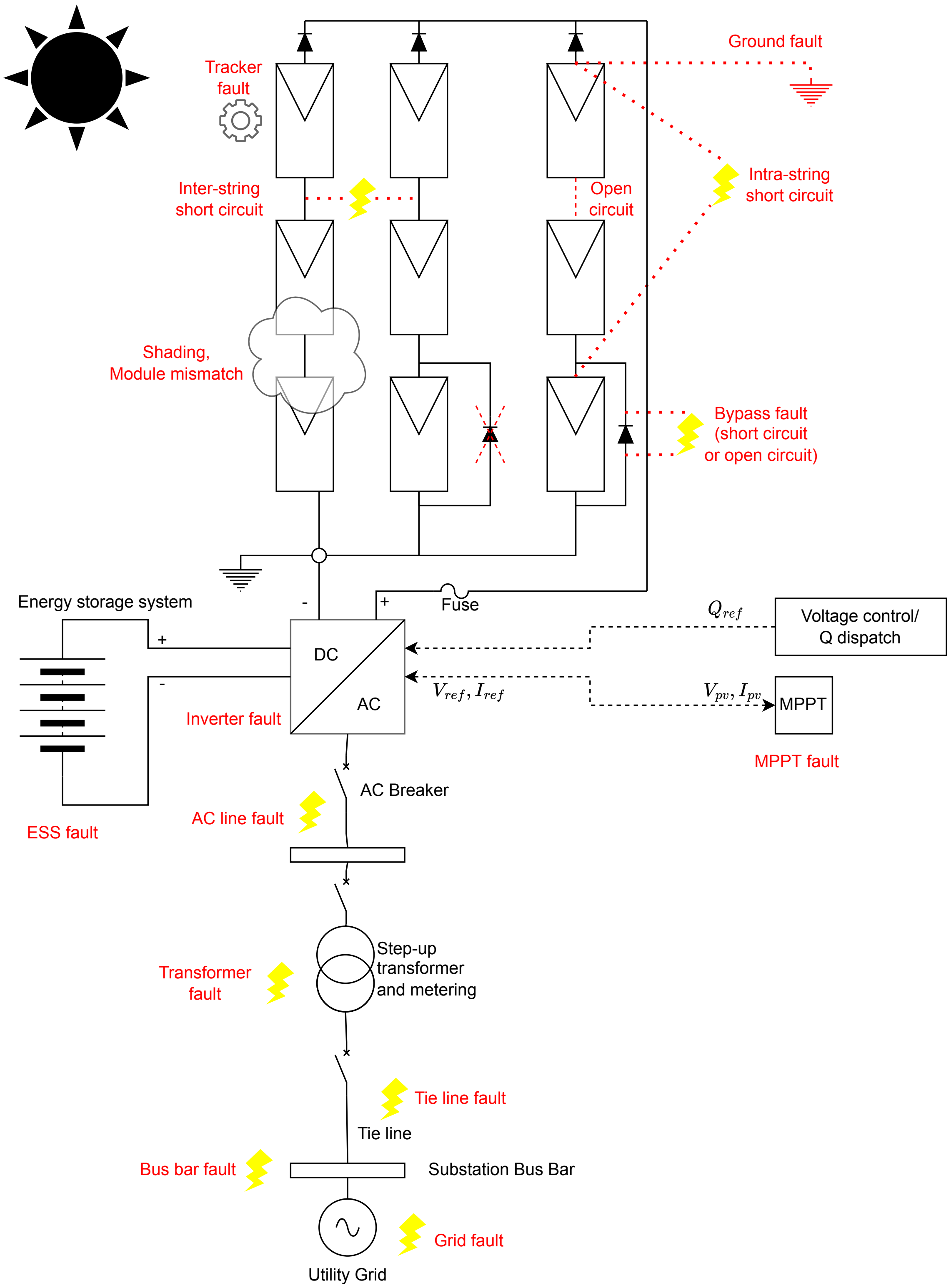

Figure 2.1: Representation of utility-scale PV plant components and some possible faults.

Most of these components have intrinsic variables, such as voltage and current values, that can help determine their operation states. Given that the utility grids (and the associated electricity market) integrate large-scale PV assets, some of the before-mentioned components require constant monitoring and control, achieved with adequate embedded systems and sensor infrastructure [4]. Since monitoring utility-scale PV assets relies on the investment and technologies employed, engineers must consider data availability when developing data-driven algorithms. Thanks to the continuous advancements in communication technologies, namely in IoT (Internet Of Things), data acquisition is becoming faster, more reliable, and more precise. Not only is this fundamental for real-time asset assessment, but it also allows better training of fault detection algorithms. However, on the industrial scale (in the order of MWp production), having sensors embedded in every PV module comes with a high economic cost. Inverters are the components that usually possess monitoring capabilities, though the grid-tie connection should also be equipped with sensors. These can be considered the primary sources of information from utility-scale PV plants, with the most accurate, fast, and reliable data acquisition.

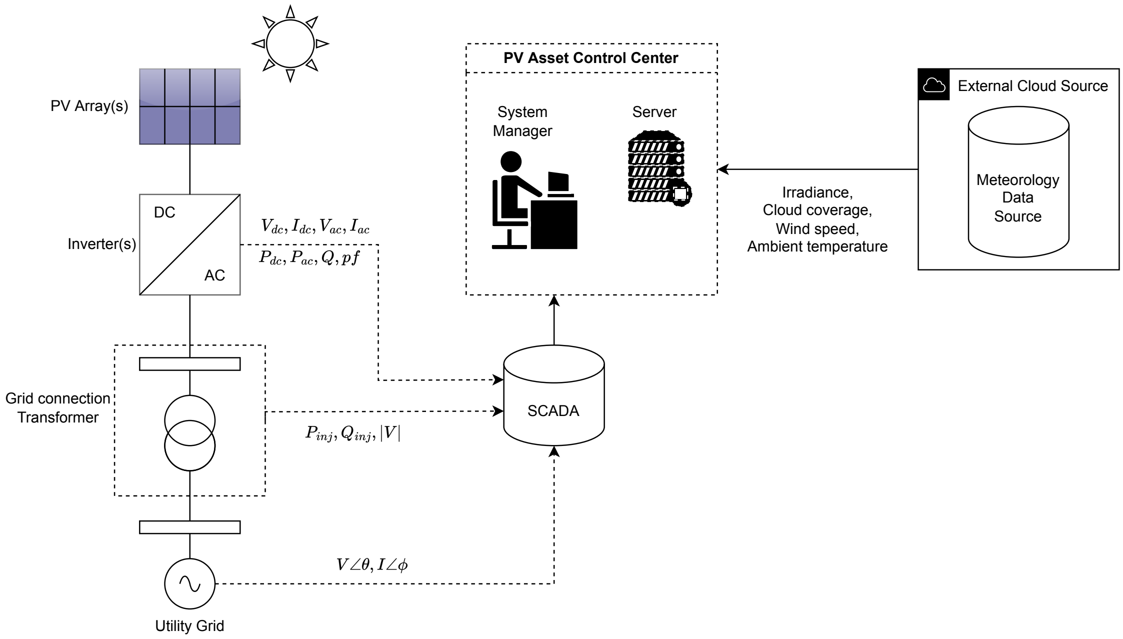

Figure 2.2: Typical data flow of utility-scale PV power plants.

Figure 2.2 represents a simplified data flow representation of a grid-tied PV system's most commonly available state variables, with most of them suggested by the IEC 61724 standard [5]. An external meteorological data source is defined since the PV system manager usually needs climate information for (at least) forecasting purposes.

2.2 Faults in Photovoltaic Systems

Several types of faults can occur in utility-scale photovoltaic (PV) power plants, which impact the performance and reliability of the system negatively. Unfortunately, some are very challenging to detect and protect the electrical installation against, thus requiring sophisticated detection algorithms [6]. Besides the economical price, their occurrence may even cause safety hazards, such as fires [7], thus the urgency in early detecting or preventing such events.

According to [6], these faults can fit into three categories: electrical, mechanical, and environmental. Electrical faults include short circuits, open circuits, and inverter failure, affecting the PV panels' power output and the system's overall efficiency. Mechanical faults include broken panels, damaged cables, and defective inverters, which can lead to system downtime and reduced performance (although not mentioned, solar tracker failures could also belong in this category). Environmental faults include extreme weather events, such as hail or strong winds, which can damage the PV panels and other components [8].

The authors in [9] cover a comprehensive review of most types of faults studied in the ambit of detection and classification algorithms. However, authors in [10] have a more succinct fault categorization that better fits this work's scope. They categorize all the major PV system faults into either DC-side or AC-side. Figure 2.3 represents this detailed categorization with a tree-like structure.

Although also prone to failure, most literature on fault detection and classification for photovoltaic systems does not encompass solar tracking faults: most studies cover fixed PV systems. The supervision and assessment of these subsystems' correct functioning can be sensor-based [11] or image-based. Some authors developed fault detection methods for these apparatuses [12], using image processing on aerial photography to determine modules' slopes. This category of failures should be better supported when developing electrical data-driven algorithms since they can significantly affect the system's efficiency. Hence, this work will attempt to include said fault category in the proposed fault detection methodology.

Throughout the literature [13], some of the most noted faults in the context of fault detection are:

- Shading: partial coverage, temporary or not, of a PV array or module. It might result in a Hot Spot fault.
- Soiling: dirt accumulation, blocking sunlight from reaching PV Cells. It might also result in a Hot Spot fault.
- Short circuit: either line-line or line-ground.
- Open circuit: connection breakage between modules.
- DC arc fault: electricity plasma arc formed on broken connections.

According to a 2017 survey conducted on five utility-scale PV plants in Italy [14], the authors observed failure rates from <1% to 3% in the majority of plants and 81.8% in the worst scenario. The high failure rate of the latter had a demonstrated cause that originated from manufacturing mistakes: snail trails. Besides this phenomenon, hot spot faults and bypass diode faults/disconnections were among the most common.

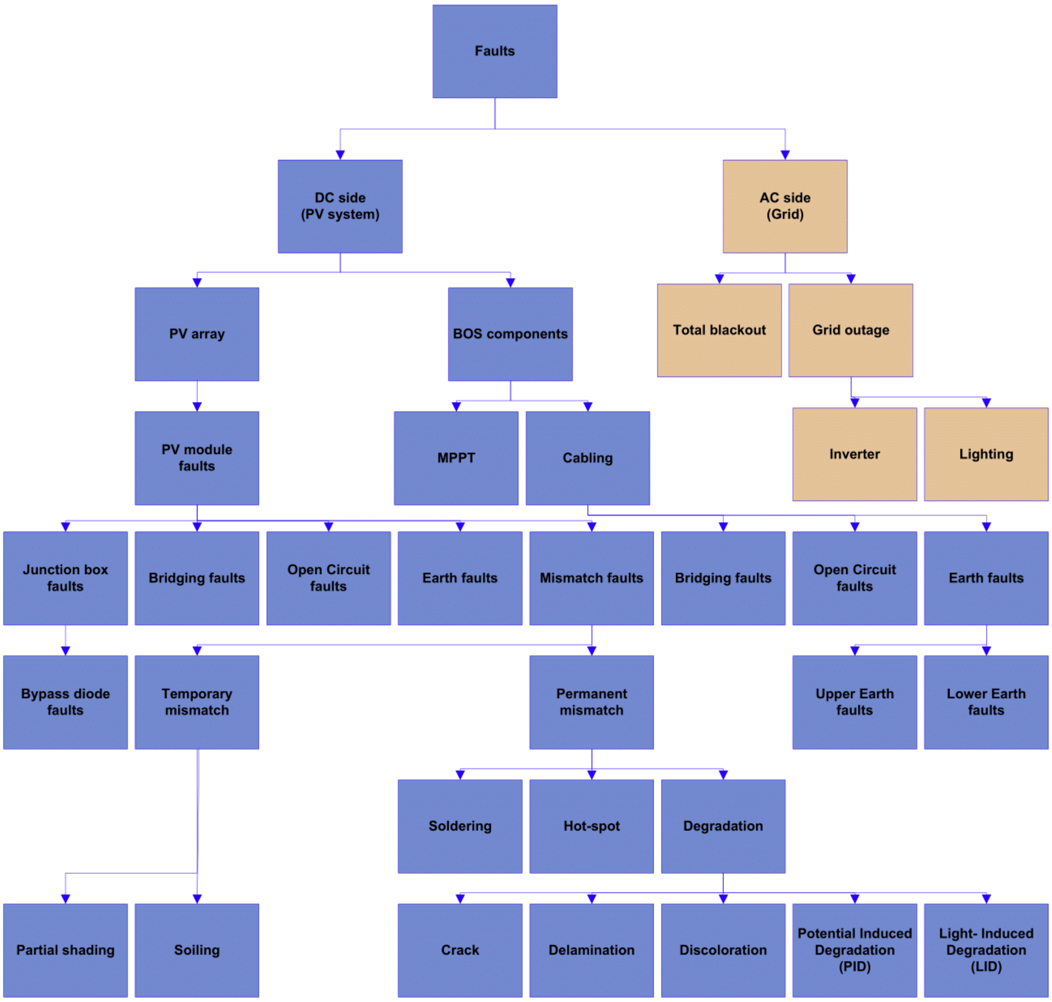

Figure 2.3: "Failures in grid-connected PV systems."
Image source and copyright: [6].

Alongside manufacturing failures, installation, planning, and other external effects can be the root cause for many of the presented faults [15].

Having the distribution of fault types from real-life scenarios is quite helpful for formulating fault detection algorithms. It allows for better generation/selection of training data and decision of classes. In figure 2.4, it is possible to observe the failure type distribution for 24.254 inspected modules. Soiling, shading, and mechanically related failures were not as prominent, with only a group share of around 6%. It is relevant to note that discoloration represents almost a quarter of all faults.

Although the study had a limited geographic scope, with only a few power plants diagnosed, it allows for a more realistic view of the common scenarios encountered in typical operational ground-mounted utility-scale PV power plants.

Due to the difficulty of classifying some of these faults, given their similarity on the consequent effect in the system, it will be seen in further sections that most fault detection algorithms only endeavor to classify between two to five types of reviewed faults.

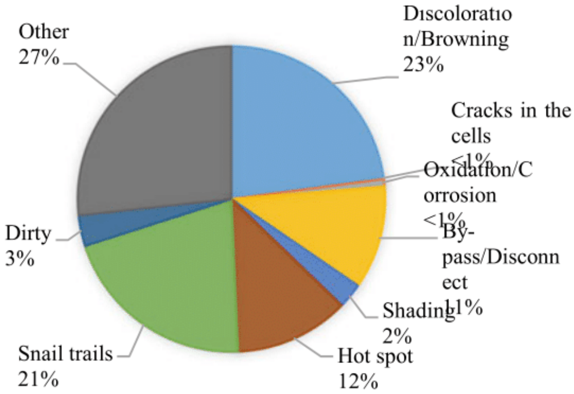

Figure 2.4: "Circle chart related to the module defects in the 5 plants (over the total number of failures)."

Image source and copyright: [14].

2.3 Modeling photovoltaic's physical/electrical behavior

Photovoltaic cells are the fundamental components of photovoltaic panels. They are made from semiconductor materials, such as silicon, and absorb photons that generate an electric current. Their electrical behavior is characterizable using the current-voltage (I-V) equation 2.1. This equation, which represents a fundamental relationship governing the operation of PV cells, can be used to predict their performance under various operating conditions, such as solar irradiance and temperature.

$$I = I_{ph} - I_d \times (e^{\frac{q \times (V_{pv} + I_{pv} \times R_s)}{n \times k \times T}} - 1) - \frac{V_{pv} + I_{pv} \times R_s}{R_p} \tag{2.1}$$

I_{ph} (A) is the light-generated current; I_0 (A) is the reverse saturation current; V_{pv} is the module's terminal voltage; I_{pv} is the module's output current; R_s (Ω) is the series resistance; R_p (Ω) is the shunt resistance; n (adimensional) is the diode ideality factor; k (J/K) is the Boltzman constant; T (K) is the cell temperature; q (C) is the electron charge;

For state estimation, it is crucial to accurately model PV modules' performance from the DC side of power converters. This information is vital for designing and optimizing PV power systems, as it enables predicting PV module performance under different conditions, as mentioned before. Accurate PV module models are also essential for state estimation and fault detection, as they provide critical information about the health and performance of PV modules, allowing for early identification of potential issues. In addition, they can be used to optimize the control and operation of PV power systems, which can improve the overall efficiency and reliability of the system [13].

Physical and empirical models broadly categorize the several state-of-the-art methods for modeling photovoltaic modules [13]. Physical models lie on the fundamental physical principles governing PV modules' operation. They typically require detailed knowledge of the PV module's electrical and optical properties, such as its current-voltage (I-V) characteristics, spectral response, and temperature dependence. These models can accurately predict the PV module's performance under a wide range of operating conditions, but they may be complex and computationally intensive to implement [16]. On the other hand, empirical models are based on experimental data and are typically more straightforward to implement. However, they may not be as accurate as physical models, especially under conditions that differ significantly from those used to generate the experimental data (usually STC) [13]. Some examples of state-of-the-art physical models for PV modules include the single-diode model (also known as the five-parameter model), and the two-diode model [17]. In contrast, one of the most used state-of-the-art empirical models is the Sandia model [13]. The choice of modeling method will depend on the specific application and the required level of accuracy and complexity; in some cases, there can be a combination of physical and empirical models.

Suppose the need arises to model PV modules in this work. In that case, it is critical to select a simple methodology so that the module's datasheet characteristics are sufficient to model the PV arrays accurately. In the case of utility-scale PV systems, detailed knowledge of the module's electrical and optical properties of empirical data may be limited, and building a model is only possible by using datasheet data. A complex model that requires more detailed information may not be feasible in such cases, and a simpler model that relies on fewer input parameters is more appropriate. The single-diode model seems appropriate for this use case, given the excellent trade-off between complexity and accuracy.

2.3.1 The five-parameter model

Figure 2.5 presents the single-diode model representation of the photovoltaic module. According to the five-parameter model, the unknown parameters are determined by fitting the model to experimental data or using data from the PV module's datasheet. The single-diode model can predict the PV module's performance under a wide range of operating conditions while maintaining reasonable accuracy. However, remembering that the single-diode model is a simplified representation of the PV module, it will have poor accuracy under certain situations compared to the more representative two-diode model [17].

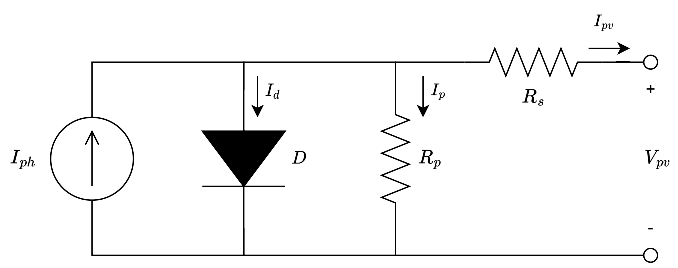

Figure 2.5: Single-diode model for photovoltaic modules.

2.4 Literature on Fault Detection and Classification for Photovoltaic Systems

The parent field of fault detection is anomaly detection (also known as outlier detection), a highly studied subject in the scope of statistics [18], applied in many scientific areas. Classification is also a well-studied subject in this field, with applications in numerous scientific contexts, from medical diagnosis to airport safety [19]. Consequently, usage or adaptations of generic tools and ad hoc methodologies have originated to aid in solving fault detection and classification problems in photovoltaics.

According to [4], the tools dedicated to PV fault detection and state estimation mostly come from mathematical/statistical methodologies, machine learning, and deep learning applications. Regarding the three general problem-solving principles mentioned before, it's known that machine learning and deep learning are the most popular and successful ones for recent applications that ought to solve complex problems. However, this categorization is somewhat limited, with contemporary literature suggesting an abundance of developed methodologies from different backgrounds, thoroughly reviewed in [9] and [10]. In [9], the authors consider two principal fault detection and classification algorithm branches: image-based and electrical-based; while [10] also distinguishes numerical-based techniques. Image-based refers to aerial or visual capture of the PV array by photography and thermal imaging, commonly used along with artificial intelligence algorithms for assessing the photovoltaic module's state. Although the contribution and importance of such methods are appreciable, this work will mainly focus on the electrical-based and numerical-based ones, as the use case of the developed tool is bound to this type of data.

Categorizing methodologies become fuzzy with the often mixed physical behavior models with machine learning, statistics, and signal processing. Figure 2.6 is an attempt to present a structure formulated from the review made by [9], [10], and this work, with a focus on the more relevant techniques (for this work's scope).

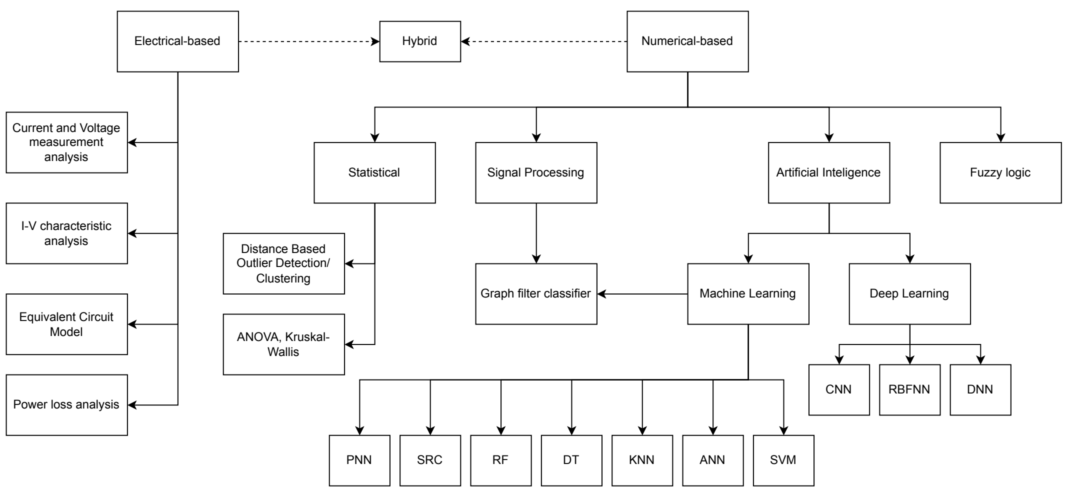

Figure 2.6: Representation of some of the methodologies employed in fault detection for PV systems.

Hybrid models are ubiquitous since combining robust statistical, signal processing, ML, or DL models and PV's electrical characterization can achieve more remarkable results. Hence, a better representation than figure 2.6 would be an incomprehensible mesh of connections representing the permutations between category aggregation.

To not wander in the literature, there must be a decision on which methodologies to revise. The developed tool in this work must meet certain real-life constraints, such as data availability, frequency, accuracy, PV system configuration, and context. Therefore, the (qualitative) potential estimation for each methodology will be based on the capability of adapting the proposed algorithms to the same restrictions. This evaluation process confines the methodology review to emphasize the ones thought to be most capable of implementation in a real scenario. Therefore, the following sections will not cover an extensive literature review, as it is not intended to repeat the works of [9] and [10], only presenting interesting or adequate methodologies related to this work's scope.

2.4.1 Statistical and Signal Processing Algorithms

Statistical methodologies look into historical data to find the characteristics of how samples relate to the population (interpolation). These methodologies yield good results in case studies of PV farms that have been logging data for a considerable time, suffering in the cases that do not. Therefore, they are limited in that it is required to have curated data sets of historical significance for relevant features of the studied systems.

The literature on statistical and signal processing fault detection algorithms for PV is mostly quite dated ([20], [21], [22]), given that more recent machine learning methods have become increasingly attractive in this matter. Nonetheless, anomaly (or outlier) detection statistical algorithms can be used for fault detection in PV systems by identifying unusual patterns or deviations from normal behavior in the data collected from the PV system. Distance-based methods, such as Euclidean, Mahalanobis, and MCD-based distances [13], may be adequate. Although simple, these techniques might only work for detecting outliers in the context of PV systems if they are scale-invariant (due to the different magnitude in the system's state variables) and resilient to outlier contamination (only with MCD-based distance). In [22], the authors applied Analysis of Variance (ANOVA) and Kruskal-Wallis test for inverter failure detection, with the downside of only being able to identify outliers in a sub-array resolution, i.e., not for specific string or module failures.

Some algorithms consider incoming data from PV systems as signals, allowing the adaptation of signal processing theory to develop ad hoc algorithms. Coming up with a relatively simple algorithm, the authors in [23] propose a power-based fault detection method that only requires delayed samples of the PV array's power output and a threshold. Its logic is that since the power output of PV systems can't vary beyond a given point, considering a very short-term period (milliseconds), significant perturbations in this variable can be associated with faults. Although the simplicity and ease of implementation, it's clear that the success of this method requires feeding the algorithm with relatively high-frequency data, which would only be feasible on-site (and with specialized equipment).

In [24], the authors successfully formulated a graph signal processing algorithm for fault classification that yields increasingly better results when there is a considerable amount of labeled data, although its training is only semi-supervised. The results outperformed other standard machine learning methods for the same training data, given 30% or more of labeled data. On another note, the data utilized came from the PVWatts [25] dataset, and the PV system is on a small scale (ASU testing facility [26]) possessing a monitoring density and capability that can be considered irrealistic for utility-scale.

The authors in [27] displayed another excellent use for graph theory, although not specifically for fault detection: they implemented a consensus-based distributed approach to minimize the impact of noise in acquired data from the PV array. By formulating a data propagation algorithm that resulted in measurement convergence, they achieved higher accuracy for state estimation. With both graph theory-based algorithm proposals, this field sparks interest in its usage for the upcoming formulated methodology, given that it is desired to achieve an algorithm that achieves fault detection alongside data consensus.

2.4.2 Machine Learning Algorithms

Machine learning came to solve some of the complications referred to in the two past subjects, as neural networks (or other learning structures) are easily capable of modeling complex, non-trivial, and nonlinear relations between data. Still, they are as good as the training data, with many structures requiring many representative learning examples to achieve good results. Their output can also be very obfuscated, meaning that many methods do not allow a direct interpretation of the relationship between inputs and outputs. This "black-box" characteristic, specifically of neural networks, is considered a disadvantage. Besides, extrapolating data remains a challenge when classically using these structures. Still, they have immense applications for PV systems, from MPP (Max Power Point) estimation to power forecasting, soiling, and fault prediction.

In [28], an ANN is utilized to classify short circuit and hot spot faults. This algorithm achieved an outstanding 98.4% classification accuracy, yet the data was simulated in *MatLab/Simulink* and only considered two classes of faults. Because the inputs were the variation of voltage and current ($\frac{dV}{dt}$, $\frac{dI}{dt}$), the algorithm required data sampling with relatively high frequency (>5Hz). The present work will not regard such methodologies as background for the upcoming tool since requiring high-frequency simulated data while covering only two fault types is quite far from a real utility-scale PV system scenario.

The trend of utilizing simulated data (sometimes without even added noise) has been a target of criticism in [29]. Accordingly, this work also emphasizes that the literature shows many proposed ML (and other types of) techniques that fall into this concept, which makes selecting appropriate methodologies to base future work on a challenging task.

The proposed ANN solution in [30] is remarkable by the diversity of fault classification achieved. It presents one of the most fault class coverage with high accuracy, considering the literature that utilizes synthetic noiseless data.

In [31], there is a captivating proposal of utilizing an autoencoder and pruned neural network to separate the tasks of detecting and classifying faults, which resulted in one of the most performant ML approaches in the literature. The algorithm classifies five states: degraded, shaded, soiled, short circuit, and STC, utilizing nine inputs representing voltage, current, power, and irradiance available from the MPPT, datasheet, or meteorological sources. While the neural network pruning adds complexity, it resulted in a better generalized and lighter-weight trained model suitable for faster detection times. Even though using data from a small-scale PV system, the presented algorithm could be adapted and implemented in an industrial scenario.

On the note of performance, the work in [32] proposes a sparse representation classifier (SRC) that evaluates if the system has line-to-line or line-to-ground faults for varying operating conditions. Although a drop in accuracy occurred for extreme circumstances, it is impressive that the algorithm identifies faults in such varied operating conditions: 10 to 50 degrees ambient temperature, 200 to 1000 w/m^2 irradiance, 10 to 60 % of mismatch, and 0 to 25 Ω of fault resistance. The feature extraction step was also very impressive, which could be a determining factor in the method's performance. Unfortunately, this work does not validate results with experimental data and only uses simulation as a source. However, the demonstrated computational performance, both in terms of training cost and utilization speed, its usage without the need for training for parameter tuning, the straightforward implementation, and consistent convergence, suggests the potential for this alternative in the face of other ML methodologies. The authors also emphasize that sparse representation might be utilized alongside different learning algorithms for classification, opening the door to many possible future implementations.

Although many other classical ML methods have been tested in this field, using SVM, KNN, RF, and many more commonly used structures, the proposals of [31]-[32] were the most inspiring for this work.

2.4.3 Deep Learning Algorithms

The field of deep learning is a branch of machine learning, with the term "deep" referring to amplified machine learning structures that ought to understand data patterns through more complex and intertwined artificial neuron connections. A simple example of a deep learning model would be the design of an artificial neural network with multiple hidden layers (DNN), with the intuition that each of these "extra" layers achieves feature/pattern recognition in a cascade. Other DL structures include the LSTM, CNN, and RBFNN. They have been explored alongside classical machine learning techniques for PV fault detection, although the known disadvantage is a usually high computational cost and relatively tricky implementation.

In [29], not only is a DL technique presented for fault detection and classification, but there is also the best attempt at comparative evaluation against other methodologies. As already mentioned, this work exposes that much of the literature presents results solely based on particular datasets comprising simulated noiseless data, which invalidates any significant quantitative comparison. Consequently, a CNN model based on the pre-trained AlexNet [33] is used both for classification or feature extraction, allied with a classical ML model for classification in the latter. The faults considered were arc fault, partial shading, open circuit, and short circuit. While the experiments utilized simulation data, adding noise and an abundance of heterogenous operating conditions better represented a real scenario. Considering noisy data, tested methodologies present 22-70% average accuracies, with the proposed fine-tuned AlexNet CNN reaching a maximum of 70.45%. Hence, this work presents one of the best benchmarks in the literature, with decent coverage of other ML and DL algorithms, demonstrating the most realistic accuracy results.

2.4.4 Proposed method's scope

While classical fault detection lies in the synchronous and direct evaluation of state estimation variables, realistic industrial scenarios can have data from various types, sources, and acquisition rates. It's also important to realize that monitoring equipment can register erroneous information, and current communication technology is also susceptible to delays and data loss. With this in mind, recent developments in the intelligent composition of deep learning structures aligned with graph theory spark some interest in their application to this field, such as the new deep learning technique named Cell Complex Neural Networks [34]. The motivation for choosing such a structure comes from its data propagation and consensus capability. The propagation techniques utilized in a CXN appeal to graph theory, dividing a system into other subsystems and components (nodes, also called cells in [34]) that share information. Even if the direct application of this structure might not be feasible or grant better results in the context of fault detection, its modification to meet the scope's needs could result in a robust and efficient solution. Further investigation of this state-of-the-art tool will unroll throughout the development of this work in an attempt to adapt this knowledge to the PV fault detection field.

The proposed method should pertain to the DL or the hybrid category since, while having a central component of DL, it may also require modeling the PV system's components.

Chapter 3

Preliminary Work Plan

Será recolher dados, implementar os modelos mais interessantes e relevantes encontrados na literatura e estudar as aplicações de cell neural networks para adaptar a deteção de falhas/coesão de dados. Capítulo por desenvolver...

Chapter 4

Conclusions

With the contemporary state of investment in grid-connected big-scale photovoltaic systems, we can affirm that resource allocation to fault recovery is essential both to meet grid code requirements, safety and health standards, reduce investment risk and maximize the asset's throughput. The literature confirms such claim, by presenting a plenitude of research work regarding this problematic. Throughout its review, it is noted that there is a sparse variety of different methodologies all applied for both fault detection and classification on PV systems, from image-based to electrical-based, using signal processing, graph theory, statistics, machine learning, deep learning and even quantum machine learning. Different authors utilize the same dataset and research facility in various fault detection research works, which has the benefit of data consistency across multiple pieces of literature. However, there is a significant downside, as this work's scope encompasses PV systems with differing characteristics: utility-scale instead of small-scale, with fewer monitoring capabilities.

Bibliography

[1] *Renewable Capacity Statistics 2021*, 2021. [Online]. Available: www.irena.org

[2] A. Cabrera-Tobar, E. Bullich-Massagué, M. Aragüés-Peñalba, and O. Gomis-Bellmunt, “Topologies for large scale photovoltaic power plants,” pp. 309–319, 6 2016.

[3] A. H. I. Mourad, H. Shareef, N. Ameen, A. H. Alhammadi, M. Iratni, and A. S. Alkaabi, “A state-of-the-art review: Solar trackers,” *2022 Advances in Science and Engineering Technology International Conferences, ASET 2022*, 2022.

[4] A. Mellit and S. Kalogirou, *"Handbook of Artificial Intelligence Techniques in Photovoltaic Systems"*. Academic Press, 2022, https://www.sciencedirect.com/book/9780128206416/handbook-of-artificial-intelligence-techniques-in-photovoltaic-systems.

[5] “Iec 61724-1:2021 rlv | iec webstore,” https://webstore.iec.ch/publication/70170, accessed: 2023-01-07.

[6] D. S. Pillai and N. Rajasekar, “A comprehensive review on protection challenges and fault diagnosis in pv systems,” pp. 18–40, 8 2018.

[7] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, “A comprehensive review of catastrophic faults in pv arrays: Types, detection, and mitigation techniques,” pp. 982–997, 5 2015.

[8] H.-Y. Li, *Assessment of Photovoltaic Module Failures in the Field*, 05 2017.

[9] Y. Y. Hong and R. A. Pula, “Methods of photovoltaic fault detection and classification: A review,” pp. 5898–5929, 11 2022.

[10] A. Livera, M. Theristis, G. Makrides, and G. E. Georghiou, “Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems,” pp. 126–143, 4 2019.

[11] A. Stepanov, A. Sokolovs, and L. Dzelzkaleja, “Solar tracker supervisory system,” *2014 55th International Scientific Conference on Power and Electrical Engineering of Riga Technical University, RTUCON 2014*, pp. 79–83, 12 2014.

[12] T. G. Amaral, V. F. Pires, and A. J. Pires, “Fault detection in pv tracking systems using an image processing algorithm based on pca,” *Energies 2021, Vol. 14, Page 7278*,

vol. 14, p. 7278, 11 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/21/7278/htmhttps://www.mdpi.com/1996-1073/14/21/7278

[13] H. Braun, S. T. Buddha, V. Krishnan, C. Tepedelenlioglu, A. Spanias, T. Takehara, T. Yeider, M. Banavar, and S. Takada, “Signal processing for solar array monitoring, fault detection, and optimization,” 2011, summary:
chapter 2 describes physics and electrical behaviour of the PV cell
chapter 3 has description of the most common faults
chapter 4 has statistical signal processing-based techniques for determining the presence
of a fault.

chapter 2:
“The most common metric comparing measured module performance to predicted behavior
is the performance ratio”
Modelling performance:
Sandia Model: very accurate (1utilizes only data provided by the manufacturer at standard test conditions and does not require additional measurements to derive parameters. [Online]. Available: https://link.springer.com/10.1007/978-3-031-02497-9

[14] F. Grimaccia, S. Leva, A. Dolara, and M. Aghaei, “"survey on pv modules’ common faults after an o&m flight extensive campaign over different plants in italy",” *IEEE Journal of Photovoltaics*, vol. 7, pp. 810–816, 5 2017.

[15] “Fire safety for pv systems - sunny. sma corporate blog.” [Online]. Available: https://www.sma-sunny.com/en/fire-safety-for-pv-systems/

[16] M. Kumar and D. V. S. K. Rao, “Modelling and parameter estimation of solar cell using genetic algorithm,” *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, pp. 383–387, 5 2019.

[17] R. Godina, E. M. Rodrigues, E. Pouresmaeil, and J. P. Catalão, “Simulation study of a photovoltaic cell with increasing levels of model complexity,” *Conference Proceedings - 2017 17th IEEE International Conference on Environment and Electrical Engineering and 2017 1st IEEE Industrial and Commercial Power Systems Europe, EEEIC / I and CPS Europe 2017*, 7 2017.

[18] N. R. Prasad, S. Almanza-Garcia, and T. T. Lu, “Anomaly detection,” *ACM Computing Surveys (CSUR)*, vol. 14, pp. 1–22, 7 2009. [Online]. Available: https://dl.acm.org/doi/10.1145/1541880.1541882

[19] M. Ilas and C. Ilas, “Towards real-time and real-life image classification and detection using cnn: a review of practical applications requirements, algorithms, hardware and current trends,” 10 2020, pp. 225–233.

[20] S. Buddha, H. Braun, V. Krishnan, C. Tepedelenlioglu, A. Spanias, T. Yeider, and T. Takehara, “Signal processing for photovoltaic applications,” *2012 IEEE International Conference on Emerging Signal Processing Applications, ESPA 2012 - Proceedings*, pp. 115–118, 2012.

[21] Y. Zhao, F. Balboni, T. Arnaud, J. Mosesian, R. Ball, and B. Lehman, "Fault experiments in a commercial-scale pv laboratory and fault detection using local outlier factor," *2014 IEEE 40th Photovoltaic Specialist Conference, PVSC 2014*, pp. 3398–3403, 10 2014.

[22] S. Vergura, G. Acciani, V. Amoruso, and G. Patrono, "Inferential statistics for monitoring and fault forecasting of pv plants," *IEEE International Symposium on Industrial Electronics*, pp. 2414–2419, 2008.

[23] H. Iles and Y. Mahmoud, "Power based fault detection method for pv arrays," *IECON Proceedings (Industrial Electronics Conference)*, vol. 2021-October, 10 2021.

[24] J. Fan, S. Rao, G. Muniraju, C. Tepedelenlioglu, and A. Spanias, "Fault classification in photovoltaic arrays using graph signal processing," *Proceedings - 2020 IEEE Conference on Industrial Cyberphysical Systems, ICPS 2020*, pp. 315–319, 6 2020, "We propose here a graph signal processing based semi-supervised learning technique, which achieves good performance in fault classification with relatively limited data."
Classified faults: standard test conditions (STC), shaded modules, degraded modules, soiled modules, and short circuit conditions.
Also uses PVWatts dataset
MOST SIMILAR WITH THE THESIS.

[25] A. P. Dobos, "Pvwatts version 1 technical reference," 2013. [Online]. Available: www.nrel.gov/publications.

[26] S. Rao, D. Ramirez, H. Braun, J. Lee, C. Tepedelenlioglu, E. Kyriakides, D. Srinivasan, J. Frye, S. Koizumi, Y. Morimoto, and A. Spanias, "An 18 kw solar array research facility for fault detection experiments," *Proceedings of the 18th Mediterranean Electrotechnical Conference: Intelligent and Efficient Technologies and Services for the Citizen, MELECON 2016*, 6 2016.

[27] S. Katoch, G. Muniraju, S. Rao, A. Spanias, P. Turaga, C. Tepedelenlioglu, M. Banavar, and D. Srinivasan, "Shading prediction, fault detection, and consensus estimation for solar array control," *Proceedings - 2018 IEEE Industrial Cyber-Physical Systems, ICPS 2018*, pp. 217–222, 6 2018, they have sensors per each panel, not realistic for utility scale pv plants
good information about consensus in graphs
consensus is used for determining average values along the entire array, based on each sensor's measurements.

[28] A. Kumari, A. Shekhar, and M. S. Kumar, "An artificial neural network-based fault detection technique for pv array," *2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies, ICEFEET 2022*, 2022, uses ANN for classification Classified Faults: short circuit faults and hot spot faults Inputs: dI/dt and dV/dt Results: 98.4classification accuracy Pros: is that it can use the dI/dt and dV/dt as network inputs to assess faults, yields high accuracy. Cons: uses high frequency data from current and voltage of the panels, sampled every few milliseconds.

[29] F. Aziz, A. U. Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*, vol. 8, pp. 41 889–41 904, 2020, "utilizes deep two-dimensional (2-D) Convolutional Neural Networks (CNN) to extract features from 2-D scalograms generated from PV system data in order to effectively detect and classify PV system faults."

"A survey study conducted in 2010 showed that such faults can reduce the generated power of photovoltaic systems annually by about 18.9to noice. Heavy deep learning model.

[30] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar array fault detection using neural networks," *Proceedings - 2019 IEEE International Conference on Industrial Cyber Physical Systems, ICPS 2019*, pp. 196–200, 5 2019, covers a lot of faults: 8 in total.

[31] S. Rao, G. Muniraju, C. Tepedelenlioglu, D. Srinivasan, G. Tamizhmani, and A. Spanias, "Dropout and pruned neural networks for fault classification in photovoltaic arrays," *IEEE Access*, vol. 9, pp. 120 034–120 042, 2021, uses an autoencoder machine learning framework
Emphasizes the performance
Autoencoder detects the fault, and neural network classifies it
"We consider the approach of fault detection and classification by monitoring the electrical signals such as maximum power point tracking (MPPT) parameters"
They use the THE PVWatts DATASET
BEST ARTICLE SO FAR!

[32] H. Kilic, B. Khaki, B. Gumus, M. Yilmaz, and P. Palensky, "Fault detection in photovoltaic arrays via sparse representation classifier," *IEEE International Symposium on Industrial Electronics*, vol. 2020-June, pp. 1015–1021, 6 2020, detect faults: DC short circuit faults of PV array. line to line and line to ground
Has a lot of sources and compares a lot of them!
"The aim of the proposed method is to detect the faults under low-mismatch, low-irradiance and high-impedance conditions."
"The common faults in PV systems are short and open circuit faults as well as panel mismatch and module failures"
Relatively light algorithm that doesn't require tuning.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84–90, 6 2012.

[34] M. Hajij, K. Istvan, and G. Zamzmi, "Cell complex neural networks," 2020.