Artificial Intelligence for Operation and Maintenance of PV Plants

Deliverable D1.1

Use cases for O&M of solar power plants

Lead Beneficiary	INESCTEC
Delivery Date	31/10/2021
Dissemination Level	Public
Status	Released
Version	1.0
Keywords	PV systems, Inverters, Artificial Intelligence, O&M, Decision-making

European Union
European Regional Development Fund

Disclaimer

This work is financed by the ERDF - European Regional Development Fund through -the Operational Programme for Competitiveness and Internationalisation COMPETE 2020 under the Portugal 2020 Partnership Agreement within project AI4PV, with reference POCI-01-0247-FEDER-111936 – and Spain's CDTI Multi-regional Operational Programme 2014-2020. International collaborative project EUR 2020058 with the seal of the AI EUREKA CLUSTER.

Document Information

Project Acronym	AI4PV
Work Package	WP 1
Related Task(s)	T1.1
Deliverable	D1.1
Title	Use cases for O&M of solar power plants
Author(s)	Louelson Costa (INESC TEC), Christian Verrecchia (EDP NEW), Ruben Gonzalez Bernal (ISOTROL)

Revision History

Revision	Date	Description	Reviewer
0.1	15/09/2021	First draft	INESCTEC
0.2	08/10/2021	First Version	EDP, INESCTEC
0.3	22/10/2021	Second Version	ISOTROL
1.0	29/10/2021	Final version	EDP

EXECUTIVE SUMMARY

The photovoltaic (PV) plants are large facilities that have multiple sensors monitoring thousands of PV modules, power electronics components, transformers, etc. To develop a solution that will support the operation and maintenance (O&M) technicians of these facilities can reduce risks, reduce downtime, increase power production, increase early fault detection, etc.

This report includes a literature review on the recent development of AI (Artificial Intelligence), ML (Machine Learning), and DT (Digital Twin) solutions for PV plant O&M. Multiple approaches are analysed, mainly for AI, ML solutions which have a wider range of papers and methodologies. Regarding the DT, relatively few papers and solutions are found, showing that this is an area of the large potential for innovative solutions. Besides the scientific papers, various white papers, technical reports, etc., from multiple companies are reviewed to identify the most common issues and practices in a PV plant and how to mitigate them by applying AI, ML, and DT solutions.

The main characteristics that should lead the development of these solutions are:

- Early fault and failure detection and diagnosis;
- Root cause analysis;
- Recommendation of possible causes and solutions of the problem.

The achievement of these characteristics is dependent on multiple features: real-time data streaming, multiple measurement points of electrical data (currents and voltages), historical data analytics in combination with previous maintenance reports. Nowadays, the DT system of PV plants is lacking when compared to other areas of application in the industry, thus, to develop a DT combined with AI, ML algorithms and recommendation systems is a challenging endeavour that will result in many innovative outcomes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

Acronym	Meaning
ABC	Artificial Bee Colony
AC	Alternating Current
AD	Analogic-Digital
AE	Auto Encoders
AEL_ED	Avoided Energy Losses due to Early Detection of failures
AEL_UD	Avoided Energy Losses due to Underperformance and Degradation detection
AI	Artificial Intelligence
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
APM	Asset Performance Management
ATS	Adaptive Tabu Search
CNN	Convolutional Neural Network
DA	Digital-Analogic
DC	Direct Current
DGA	Dissolved Gas Analysis
DL	Deep Learning
DSLF	Daily Soiling Loss Factor
DSP	Digital Signal Processor
DT	Digital Twin
FCNN	Fully Connected Neural Network
FDA	Fault Detection Accuracy
FL	Fuzzy Logic
FPGA	Field Programmable Gate Array
FRA	Frequency Response Analysis
GA	Genetic Algorithm
GHI	Global Horizontal Irradiance
HLUC	High-Level Use Case
HMM	Hidden Markov Model
IC	Integrated Circuit
IFAC	International Federation of Automatic Control
IoT	Internet of Things
IR	Infrared
KPI	Key Performance Indicator
LOF	Local Outlier Factor
LSTM	Long Short Term Memory
MAE	Maximum Absolute Error
MAPE	Mean Absolute Percentage Error
MaxMAE	Maximum Mean Absolute Error
MBE	Maximum Bias Error
ML	Machine Learning
MLP	Multi-Layer Perceptron
MPE	Maximum Percentage Error
MPPT	Maximum Power Point Tracking
MTTR	Mean Time to Repair
NASA	National Aeronautics and Space Administration
NMaxAE	Normalised Maximum Absolute Error

NRMSE	Normalised Root Mean Square Error
O&M	Operation and Maintenance
OC	Open Circuit
PAI	Plant Availability Increase
PD	Partial Discharge
PF	Power Factor
PI	Proportional- Integral
PLL	Phase-Locked-Loop
POCV	Percentage of Open Circuit Voltage
PSCC	Percentage of Short Circuit Current
PSO	Particle Swarm Optimization
PV	Photovoltaic
PVD	Percentage of Voltage Drop
RA	Recommendation Accuracy
RCA	Root Cause Analysis
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RoF	Risk of Failure
RRT	Reduced Response Time
RSL	Reduced Soiling Losses
RUL	Remaining Useful Life
RUO	Reduced Unexpected Outages
SC	Short Circuit
SCADA	Supervisory Control And Data Acquisition
SoTA	State of The Art
STC	Standard Test Condition
SVM	Support Vector Machine
THD	Total Harmonic Distortion
UAV	Unmanned Aerial Vehicles
UC	Use Case
UHF	Ultra-High-Frequency
US	United States
UV	Ultraviolet
VAE	Variational Auto Encoders

COMPETE 2020 PORTUGAL 2020 CDTI @CDTIoficial European Union European Regional Development Fund

GLOSSARY OF KEY TERMS

Artificial Intelligence	Artificial intelligence is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence.
Machine Learning	Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.
Deep Learning	Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviours of the human brain—albeit far from matching its ability—allowing it to "learn" from large amounts of data.
Fault	A fault is an unpermitted deviation of at least one characteristic property (feature) of the system from the acceptable, usual standard condition.
Failure	Permanent interruption of a system's ability to perform a required function under specified operating conditions.
Malfunction	Intermittent irregularity in fulfilment of a systems desired function.
Fault detection	Determination of faults present in a system and time of detection.
Fault diagnosis	Determination of kind, size, location and time of detection of a fault by evaluating symptoms. Follows fault detection. Includes fault detection, isolation and identification

1. INTRODUCTION

The energy generated via photovoltaic (PV) technology is growing more and more every year. China is the leader in cumulative capacity installed (253.4 GW), accounting for almost one-third of the global capacity. China is followed by the European Union (151.3 GW) and the United States (93.2 GW). The cumulative installed capacity stands at 758.9 GW. In Europe, Germany, Netherlands, and Spain are the countries with the largest installed capacity growth [1]. Those numbers show that the PV systems are continuously increasing. The growth of this market brings along with the need for the development of new technologies or the adaptation of existing solutions to improve overall performance, increasing productivity and reducing losses (electrical, financial, or time losses). Thus, automatic fault and failure detection and diagnosis become paramount. The large number of data that a PV plant generates can be very valuable to assist the operator and maintenance team. The use of AI (Artificial Intelligence), ML (Machine Learning), and DT (Digital Twin) tools to supervise multiple variables and parameters can avoid failure, mitigate faults, reduce downtime, and increase the response time for troubleshooting. That is why AI and DTs solutions have been gaining so much attention in the PV-systems area.

Regarding larger PV plants, which can be composed of thousands of PV modules connected in parallel and/or in series, their operation and maintenance are not trivial. Multiple types of equipment are connected to supply the PV power to the grid: PV modules, cables, and connectors, combiner boxes, DC-DC converters, DC-AC converters, transformers, etc. [2]. Each component is subjected to faults and/or failures and having a problem in a single part of the PV plant may cause the downtime of thousands of PV modules, interrupting the power generation.

Thus, the constant monitoring of the PV plant is critical to ensure a safe and continuous power supply during the daytime: this is achieved by the usage of SCADA systems. However, the use of AI, ML tools combined with DT modelling and simulation can greatly improve the work of the O&M (Operation and Maintenance) team. These tools can be a gamechanger, enhancing and fostering predictive maintenance (through fault diagnosis and detection), thus reducing O&M costs and downtime. Beyond that, the development of a recommendation system can supply the O&M team with when, where, and what caused the problem, along with some possible solutions to solve the problems.

It is worth noting that across multiple references (scientific papers and white papers), there are some different concepts of fault and failure. Some refer to faults as anomalies, failures as faults, etc. This report uses the definitions of fault, failure, detection, diagnosis, etc., presented in [3], according to the IFAC Tech Committee Safe process and reported in the glossary of this deliverable.

1.1 SCOPE OF REPORT

This report is a Deliverable of Task 1.1 from Work Package 1 of the AI4PV Project. It contains a literature review on AI, ML algorithms and DT applications for multiple systems, later focusing in power, PV systems regarding O&M. The literature review considers classical and recent scientific papers, as well as white papers. The main advancements and gaps in the technology are identified and discussed on how it can be improved. Finally, an overview of the AI4PV solutions is reported, according to the proposal of the project and updated with the most recent discussion in the area. It is also reported a of set of possible Use Cases (UCs), related to PV plants in which AI4PV solutions can play an important role to detect and diagnose faults and failures.

1.2 OUTLINE OF REPORT

This deliverable articulates over 6 main Sections.

Section 1 provides an overview of the report.

Section 2 provides an overview of the different applications related to PV systems where AI can be employed. A short literature review is included, explaining what the main AI methods are for the different applications, what are their strengths and limitations.

Section 3 provides an overview of the actual AI-based solutions for O&M in PV farms. The review discriminates academic and industrial works and pinpoint the main limitation and gaps that needs to be filled so as to unlock the full potential of AI applications in the solar industry.

In Section 4 AI4PV solutions and modules are described. For each module a short description of their functionalities is included as well as their main objectives, requirements and conditions. Furthermore for each module, a set of UCs that will be addressed within the project is provided.

In Section 5, a list of KPIs is provided. These metrics will be used and monitored during the development phase and test campaign so as to validate the AI4PV solutions. They will serve also as mean of comparison to benchmark AI4PV technologies against the State of the Art.

In Section 6, the main conclusions are drawn according to the work developed over the previous sections.

2. AI APPLICATIONS FOR PV SYSTEMS

The development of AI solutions for the industry has been increasing during the past decades. The use of AI is a large field of research that has multiple applications in different areas: medicine, biologics, engineering, gaming, etc. [4]. Regarding power and energy, PV, and power electronics systems, the use of AI solutions is not new. AI is already being used to help in the design, development, control, and O&M of multiple solutions. For PV plant O&M it is applied through the use of ML and DL (Deep Learning) algorithms for historical data analytics [4] [5].

Even though the application of AIs started in other areas of industry, they quickly started to be studied by power and energy systems and power electronics engineering. The PV systems have features from both of these areas, requiring expertise from both fields. Usually, the AI can be applied in different phases of an asset, such as design, development, operational, and dismissal [6]. In PV systems is not that different, as the AI can be applied for sizing, modelling, control and MPPT (Maximum Power Point Tracking), thermal performance, power production forecasting, and faults and failures detection and diagnosis [7] [8], i.e., O&M. Those applications are summarized in Table 2.1.

TABLE 2.1: AI ALGORITHMS FOR PV APPLICATIONS

Application	Algorithms
Sizing	ANN (Artificial Neural Network) GA (Genetic Algorithm) ANFIS (Adaptive Neuro-Fuzzy Inference System)
Modelling and simulation	ANN GA ANFIS
Control and MPPT	GA-FL (Genetic Algorithm – Fuzzy Logic)
Output power prediction	ANN
Fault and failure (dc side only)	ANN FL

Regarding O&M application of PV AI solutions, the main methodologies use data from current and/or voltage measurements from the PV modules and/or inverters (combined with meteorological data), aerial and IR (Infrared) images or clustering-based detection and diagnosis using unlabelled data [8]. In some cases, multiple methods can be employed to achieve a common goal, besides the combination of those AI techniques with DTs [9]. Those techniques are selected according to the available data for the Root Cause Analysis (RCA). Similar to other AI applications, it is important to pre-process the data fed to the algorithms, aiming to remove inconsistencies and remove fewer valuable data. Besides that, feature engineering is a key factor when developing these solutions [7].

Being large systems, the operation and maintenance (O&M) of photovoltaic plants are not trivial. There are multiple types of equipment that are prompted to present faults or failures. Whilst some of the problems are related to the PV modules soiling, crack, or aging, the inverters, usually, present problems related to overvoltage, overheating, or other electrical issues [9]. There are multiple variables to be monitored and a large area to cover with sensors, monitoring cameras and/or drones, and power electronics equipment [10]. Thus, from the meteorological parameters (solar irradiance and temperature), photovoltaic modules parameters (soiling, temperature, voltage, current, etc.), and the power electronics equipment parameters (operating status, failure detection, voltage, current, total harmonic distortion THD, power factor, temperature, etc.), there could be too much information for a technician to be monitoring. Of course, the use of a SCADA system, which greatly improves the monitoring of the multiple variables of a PV plant, can help a human employee to keep track of what is happening at the plant. However, the tracking capability of a technician is limited and many signals and deviations that may indicate a fault or failure are not detected until a failure occurs. The use of AI systems to do the processing of this large data set collected from PV plants can provide valuable information for the operator and its company [4] [9]. Regarding the energy losses, there are three main causes, detailed in Figure 2-1:

- Internal causes: events that are under control of the O&M team;
- External causes: all events that are not under control of the O&M team; and
- PV plant efficiency.

This project will devout special attention to the inverter, soiling, inverter performance, and transformer efficiency.

Within this section, it is reported a general overview of the applications that employ the use of AI for PV systems.

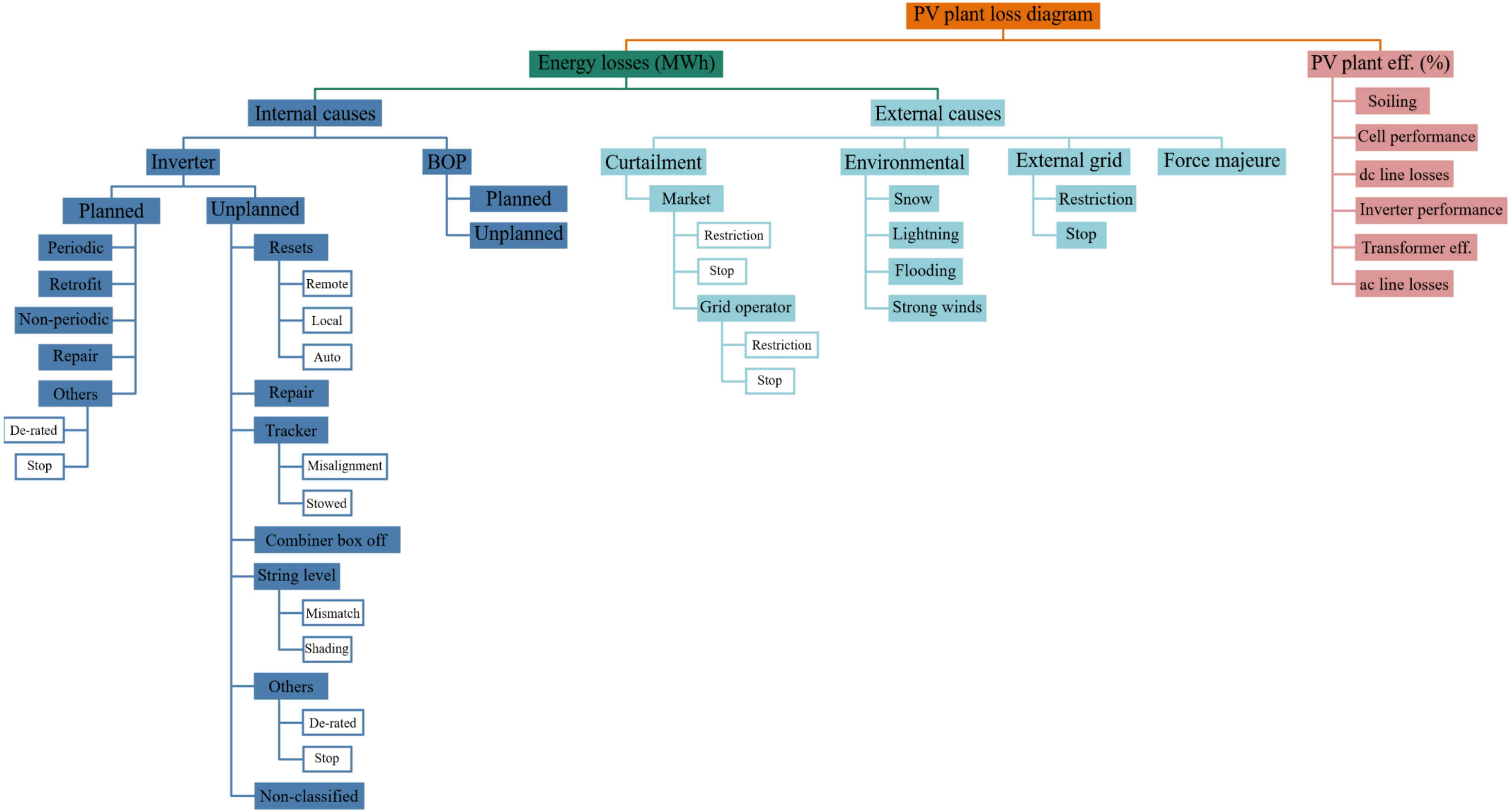


FIGURE 2-1: MAIN CAUSES OF ENERGY LOSSES IN PVPLANTS.

2.1 SOLAR PANEL PARAMETER IDENTIFICATION

When PV modelling and simulation is employed for both performance prediction and fault diagnosis, parameter identification becomes fundamental to ensure high-quality of the results. Usually, PV systems are modelled through circuit-based representation such as single or double diode models [11]. The error metric used to optimise solar cell parameters is the root mean square error (RMSE) for both models compared with the empirical I-V curve provided in the panel's datasheet. At first hand, equivalent parameters can be extracted by operational characteristics of the PV panels retrievable from the datasheet (i.e., from the I-V curve, power curve, test at STC and/or OC and/or SC, etc.). Nevertheless, in order to take into account ageing processes and deviations due to errors in the manufacturing process as well as to better fit the empirical characteristics there is the need to fine-tune these parameters throughout the lifetime. In this sense, AI can play an important role through numerical methods as pattern search and recognition so as to guarantee high accuracy. Table 2.2 summarises some of the AI method, from the family of the evolutionary algorithms, used for parameter identification.

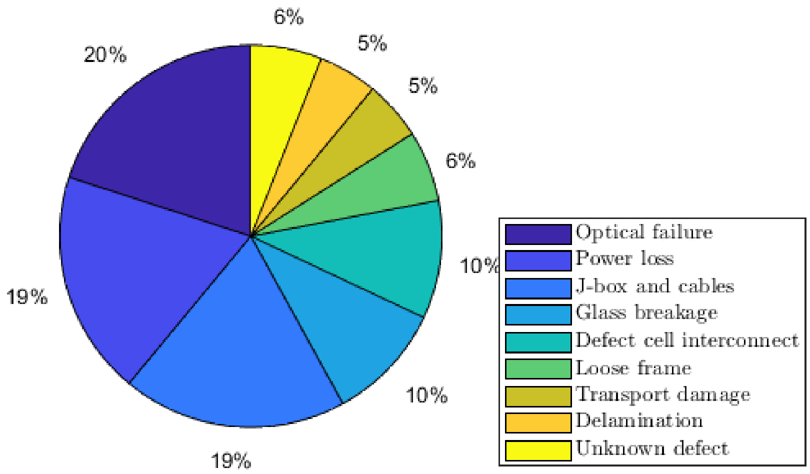
TABLE 2.2: MOST ADOPTED AI METHOD FOR PARAMETER IDENTIFICATION.

Reference	AI technique	Topic covered	Data type	Output
[20]	GA	Double diode solar cell model parameter identification	Starting values used as the diode voltages as a function of their temperature. The currents and shunt resistances where estimated	The best individuals from the final generation closely traces the experimental I-V curve
[21]	Flexible PSO	Single and double diode solar cell parameter identification	The fitness function used was the RMSE, which was dependent on the error function of the single and double diode model as well as the solar panel. Standard data was taken from R.T.C. and the model was	The proposed FPSO algorithm produced lower RMSE than the others. The I-V curves produced from the parameter identification follows the experimental curves under different irradiance and temperature values reasonably well
[22]	ABC	Parameter identification of the singles	The goal was to minimize the RMSE compared to experimental results, gathered from a 57mm	The proposed IABC converged faster and with higher accuracy (lower RMSE) than the

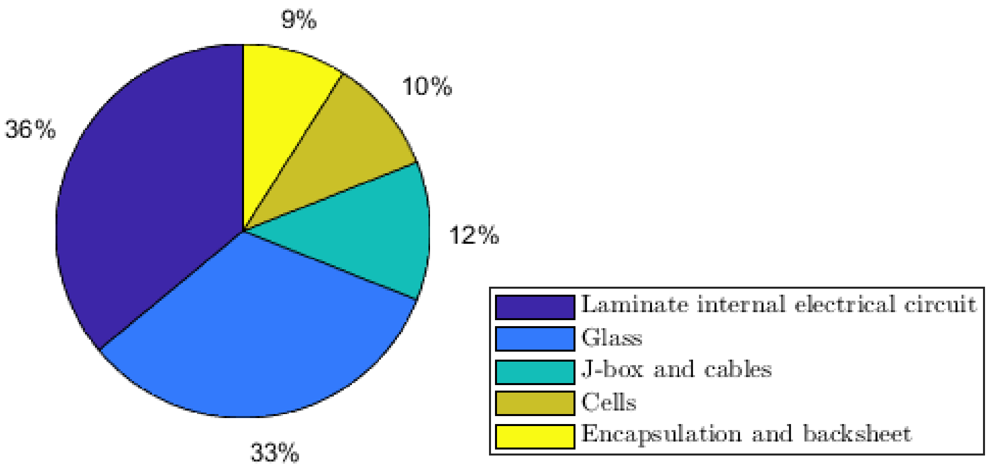
		and double diode mode	diameter commercial silicon solar cell @1000W/m2 and 25,50,75 and 100°C	other algorithms. It was further tested on solar panels with unknown parameters, but given I-V curve, and produced fairly good results
[23]	ANFIS	Parameter identification of the PV model (single, double, three diode model)	Current and Voltage historic data	It identifies the PV model parameters even when characteristics of the PV panels are unknow (i.e. datasheets, etc)

2.2 ANOMALY DETECTION IN PV ARRAYS

PV arrays are one of the most sensitive components in a PV plant and they can be a source of a myriad of different failures and underperformance. In order to maximise the power conversion efficiency, monitoring PV modules at the highest level is imperative. Typical failures of products are grouped in three categories: infant-failures, midlife-failures and wear-out-failures [24]. Infant-mortality failures occur in the beginning of the working life of a PV module. Faulty PV modules fail prematurely, thus impacting on the overall costs. Figure 2-2a shows the distribution of failures registered by a German distributor during the first year of the working life of their PV panels. Besides transport damages, that accounts for 5%, the main failures in the field are: j-box failure, glass breakage, loose frame, delamination and defective cell interconnect. In Figure 2-2b is shown the distribution of midlife failures described in a study of DeGraff [24] on PV panels operating in the field for 8 years. The study shows a high rate of glass breakage (33%), and delamination of internal circuit (36%) followed by significant j-box and cables failures (12%), burn marks on cells (10%) and encapsulant failure (9%). Finally, wear out failures occurs at the end of the PV modules lifetime. The working life of a PV module ends if a safety problem occurs or the PV module power drops under a certain level, which is typically defined between 80% and 70% of the initial power. The predominant failure in this stage is delamination, loss of isolation due to cell cracks and discolouring [24].



(a)



(b)

FIGURE 2-2: A) INFANT MORTALITY FAILURES AND B) MIDLIFE PV FAILURES [24].

In Figure 2-3 are reported type and rate of failures registered by the participants of a survey conducted by IEA and reported in [25]. As it can be seen, a plenty of faults might interest the PV panels during their life. Faults as cell cracks, discolouring of pottant, dust soiling, animal/organic soiling, defective bypass diode, corrosion of the coating, burn marks, are the most frequent failures in PV panels and they also occur throughout the entire lifetime of the panels.

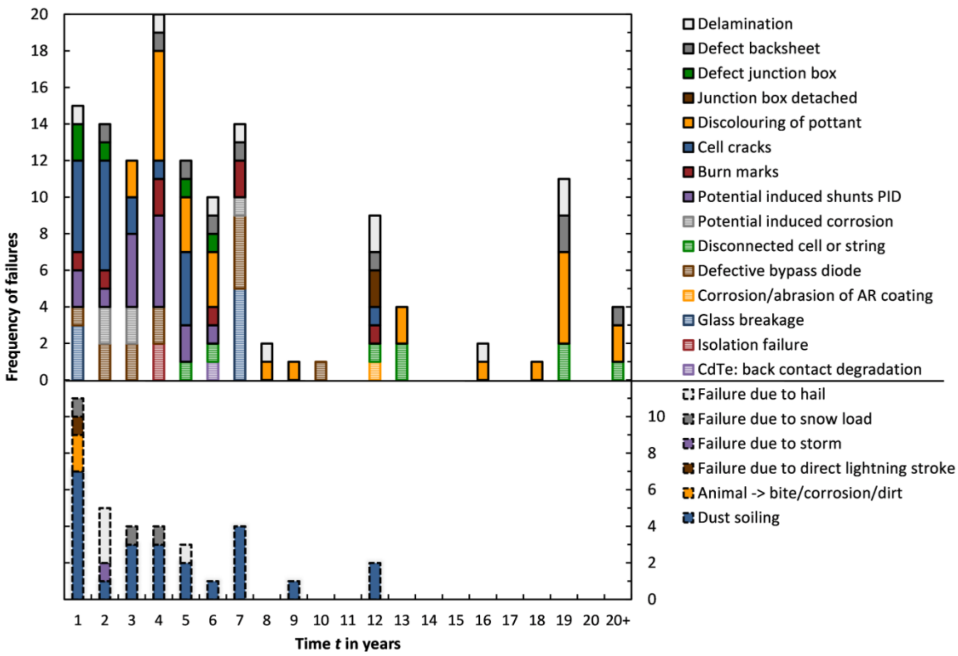


FIGURE 2-3: OCCURRENCE OF THE DETECTED FAILURES [25].

The graph in Figure 2-4 shows the occurrence of the spotted failures in the survey that cause a measurable power loss, thus failures such as delamination, cell cracks, burn marks, discolouring, defective bypass diode, soiling are the ones to which should be paid more attention.

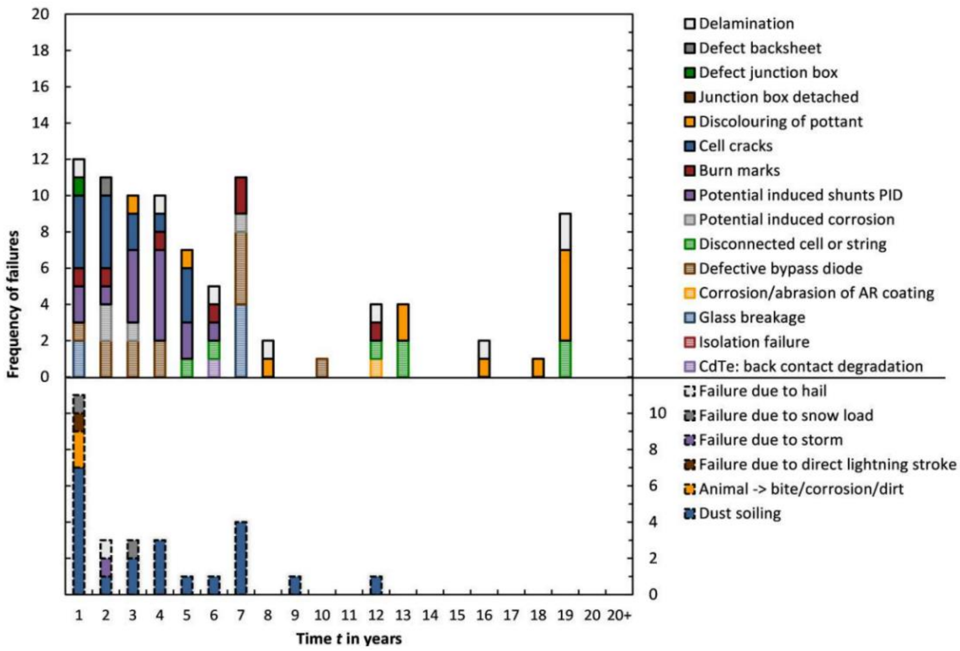


FIGURE 2-4: DETECTED FAILURES THAT REGISTERED MEASURABLE POWER LOSSES [25].

In order to understand how AI algorithms can spot these failures it is important to understand how these impact on the performance of the PV panels.

Delamination (Figure 2-5a) causes peak power losses in PV modules. It is caused by stress forces applied on the module at the interfaces deteriorated by heat, UV and moisture. The delamination can facilitate the moisture ingress and thus accelerate the corrosion of the panel. Due to the optical loss and corrosion-induced the series resistance increase, and the cell delaminated will produce less current. This originates the current mismatch, if it significant it will trigger the bypass diode and will cause further power losses.

Cell cracks (Figure 2-5b) are induced by mechanical and environmental factors that cause stress on the panel, usually as a result of mechanical forces and thermal stress. Cell cracks can origin different result based on the severity of the crack, from soft one such as PV shading to more severe impacts as decrease of the cell efficiency. In electric models with distributed series resistance, its effect can be taken into account by introducing an additional resistance localised in correspondence of the crack [25]. Moreover, in electric models a crack can result also in an increase of the saturation current of the diode: the ratio I_{MP}/I_{SC} decreases while V_{MP}/V_{OC} decreases. Flash testing and two-diode model fitting of the dark I-V curve of modules undergoing cell breakage can be used to show fundamentally the causes of power loss.

(a) (b)

FIGURE 2-5: A) DELAMINATION AND B) CELL CRACKS ON PV PANELS.

Soiling of PV modules is not a typical failure mode as cell cracking and delamination since it does not affect the long-term reliability of the PV panels, but it is rather a reversible effect as it can be removed by cleaning. Soiling can have different origins such as snow, dust accumulation, air pollution, bird droppings etc. All these effects cause PV power losses due to reduced optical transmittance. This loss can account from 5-20% per year, depending on the location and cleaning strategy [25]. *Dust soiling* (see Figure 2-6a) it's one of the major sources of soiling losses in desert-like environments and it can lead to more severe failures such as abrasion of the module surface (either due to frequent cleaning or sand blasting). Abrasion does not affect the reliability of the PV plant, but it rather causes performance losses due to deterioration of optical transmittance. A common practise to assess dust soiling is to use two irradiance sensors (encapsulated PV cells): one cleaned on a daily basis and the other one exposed to the soiling. The transmittance loss is expressed by the Daily Soiling Loss Factor (DSLF), which is the ratio between the sum of daily irradiation measured by both sensors (soiled/cleaned). Biological soiling severely impacts on the PV performance. Biofilms – microbial communities that can grow on the surface of the PV module – can reduce the portion of light that is transmitted through the glass to the PV cells. Besides, they can also contribute to soiling in other ways: they can create a surface for the dust to adhere, thus aggravating the module soiling level. The other main contributor of biological soiling is from bird droppings (Figure 2-6b), which unlike biofilms, cannot be removed by natural cleaning due to rainfall. Bird droppings are opaque and thus can entirely block transmission of light of the PV module until they are "manually" cleaned. Furthermore, in severe cases the occurrence of bird droppings can cause hot spots where the affected cell acts as a load for the remainder of the string.

(a) *(b)*

FIGURE 2-6: A) DUST SOILING AND B) BIRD DROPPINGS ON PV MODULES.

Bypass diodes are used in silicon PV modules to protect against issues that can arise from local defects. If a cell within a string of cells wired in series is shaded or damaged, it will limit the current production of that string, and can cause a local hotspot. A bypass diode mitigates this by allowing for an alternative current path. Bypass diodes can fail in open or short circuit. If it fails in open, it can pass no current and thus it is as the associated cells had not the bypass diode. This failure is insidious as it can affect the safety, reliability and performance of the PV modules: as a result, the affected PV cell can burn or overheat. On the other hand, if the diode fails in short circuit, it acts as a wire and thus the power generated from the affected module is cut off. This failure mode is easily detectable via IR images as the damaged cell overheats and thus appears hot in the image (see Figure 2-7). Besides, there are some studies [26] that developed solutions to detect failures of the bypass diode, by looking at operational data such as voltage drop, open circuit voltage and short circuit current.

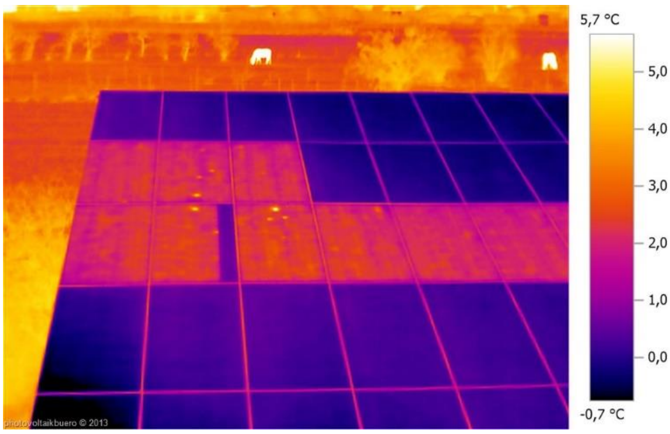


FIGURE 2-7: BYPASS DIODE FAILURE DETECTED VIA IR IMAGE.

Discolouration is one of the most over degradation mechanisms for PV modules. Discolouration of the PV module is a type of degradation generally described as a change of colour of the ethylene-vinyl acetate (EVA) encapsulant that turns to brown or yellow (see Figure 2-8a). Discolouration can become visible to an observer before module performance (current and therefore power production) start decreasing. Discolouration does not present any safety issues, unless it is very severe and localized at a single cell, where it could turn on the bypass diode resulting in all the negative effects described before [24]. EVA discolouration thus, can significantly reduce the module power output due to the reduction of sunlight reaching the module cells. Another form of discolouration is the "snail track" (Figure 2-8b). A snail track is a grey/black discolouration of the silver past of the front metallisation of screen-printed solar cells [24]. The discolouring speed depends on the season and environmental conditions. PV modules affected by snail tracks present high leakage currents and despite discolouration does not lead to significant power losses, snail tracks can make cell cracks visible and thus reduce PV module production.

(a) *(b)*

FIGURE 2-8: A) EVA BROWNING (OR DISCOLOURATION) AND B) SNAIL TRACKS ON PV MODULES.

Table 2.3 summarises the main failures for PV panels, their causes and effects as well as some detection method that can be adopted to spot these undesirable conditions.

TABLE 2.3: SUMMARY OF THE MAIN PV ARRAYS' FAILURES.

Failure/Defect	Cause	Effect	Detection method
Delamination	Stress forces on the module surface deteriorated by heat, UV and moisture.	Peak Power losses – increase of the series resistance.	[27] uses Convolutional Neural Network (CNN) for feature extraction from aerial images and SVM for defect classification, including delamination.
Cell crack	Mechanical forces and thermal stress.	From PV shading to drop of the cell efficiency. Increase of the saturation current.	[28] uses SVM, RF (random forest) and k-NN (k-Nearest Neighbours) to process EL images and detect and classify PV cracks, micro-cracks and finger failures.

			[29] processes EL with CNN to detect PV cracks.
Soiling	i) Dust accumulation and ii) bird droppings.	i) PV power losses and ii) hot spots.	[30] used ANN to estimate soiling losses, considering different parameters (panel temperature, total irradiance, solar altitude, relative humidity and short circuit current of the panel). [31] uses as inputs the PM10 and PM2.5 concentrations, the rainfall data, and the tilt angle to estimate the soiling losses. [32] uses image processing to quantify the soiling level of the PV panels.
Bypass Diode failure	Open or short circuit of the bypass diode, it can be caused by discolouration.	In open circuit it causes overheat of the affected cell. In short circuit it causes a drop of the power output.	[26] uses FL algorithm with three inputs, Percentage of Voltage Drop (PVD), Percentage of Open Circuit Voltage (POCV), and the Percentage of Short Circuit Current (PSCC). to detect up to 13 different faults associated with defective and non-defective bypass diodes.
EVA discolouration and snail tracks	Environmental conditions, thermal stress	It causes visible change in the colour of the PV modules, drop of the power output and increase of leakage currents.	[27] uses CNN for feature extraction from aerial images and SVM for defect classification, including discolouration and snail tracks.

2.3 ANOMALY DETECTION IN THE INVERTERS

The publications of AI in power electronics systems have been generally increasing since 1990, having some ups and downs along the way. However, this continuous increase shows that this is a promising field [4] [33]. Basically, the AI application is present in the three life cycles of a power converter, i.e., design, control, and maintenance. Most of the publications study the control of the power converters, followed by the maintenance, and design. The maintenance has been the object of study of 12.4% of the publications, showing that there is a lot more to be investigated [4]. Some application of AI technology to the maintenance of power electronics converters is directly connected to photovoltaic systems, as the inverter is one of the main sources of failures and maintenance operation in a solar plant [34] [35] [36].

According to a study that considered around 100 PV systems, accounting for around 500 MW_{dc}, from the year 2003 to 2017, the majority of the issues of the PV plants are related to the inverter (including faults on the dc side that led to inverter problem) [36]. The percentage of each problem across several PV plants is presented in Figure 2-9 (others refers to trackers, transformers, ac meters, and weather station). Still, according to [36], trackers may also be a large problem, being responsible for up to 58% of the faults and failures of a PV system where they are employed. As the power converter is the gateway for the power from the PV modules to the grid (or any load), they are key equipment to be investigated when studying how to improve PV systems with the assistance of AI and DTs. Thus, the fault detection and diagnosis of an inverter, both by the application of digital twin or the processing of historical data, is a key factor regarding the study of AI technology applied to solar plant O&M [34]. A system that models the inverter as a black box is not capable of diagnosing a fault inside it and will be missing valuable data, as the inverter is responsible for more than 50% of production loss followed by the PV modules (30%) [2].

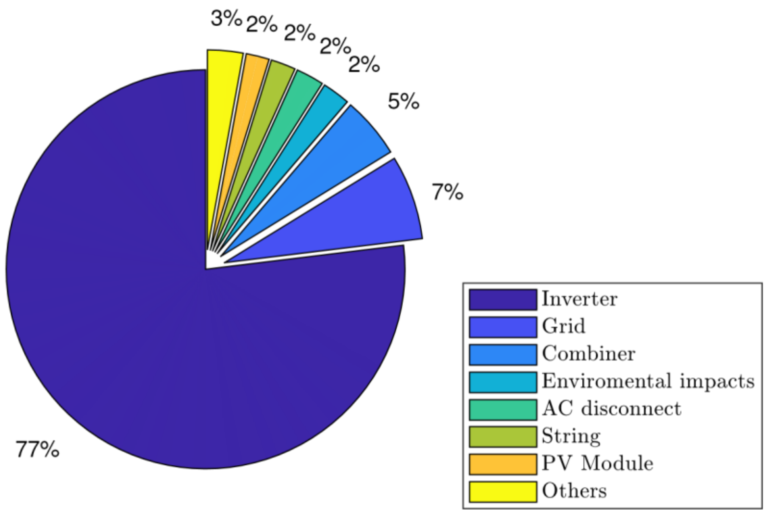


FIGURE 2-9: SUMMARY OF FAULTS AND FAILURES ACROSS MULTIPLE PV PLANTS [36].

Some commercial solutions do not model the inverter, relying on input irradiance vs. output power analysis, for instance, thus addressing only the PV plant efficiency. This is a valid approach for fault and failure detection, but they lack the diagnosis and recommendation, which would be a great addition to the O&M of solar plants. Some commercial solutions present inverter-fault diagnosis, but are limited to voltage, current, and/or temperature measurements of the inverter [35]. The detailed modelling of the inverter allows a richer diagnosis [34], however, they are very specific and may lack the interchangeability that would be desired for such a non-uniform configuration of multiple PV plants. Depending on the scope of an application, it must be high-level enough to apply to different solar plants. This is a challenging trade-off, as it is necessary to define how much low-level a model must be to present a good fault diagnosis vs. how much high-level a model must be to present good interchangeability.

However, an AI-related technology has been surfacing in the PV systems area: the DT. The digital twin is a digital doppelganger of a physical entity (which can be a motor, a car, an airplane, a wind turbine, etc.), which is fed with real data from their physical twin to detect faults, failures, performance gaps, possible improvements, etc. [6]. This technology, used in multiples areas of industry, is being applied for PV systems as well, not only for the design of PV modules and power converters, or to support control algorithms like MPPTs, but it is being applied as a powerful tool for O&M of PV plants (solar farms) as well [35].

Table 2.4 summarises the main failures for inverters [37] [38].

TABLE 2.4: SUMMARY OF THE PV INVERTERS' MAIN FAILURES.

Failure/Defect	Cause	Effect	Detection method
IGBT, MOSFETs or diodes malfunction	Thermal cycling due to inverter power on, power off, and power level change.	Thermal runaway, ceramic substrate to base plate solder fatigue, and emitter wire bond fatigue, etc., leading to short circuits.	Prediction of thermal cycle based on data provided by the semiconductors' suppliers and continuous measurements
Reactive components degradation, mainly the capacitors	Stress such as overvoltages, overheating, pollution, humidity, radiation, etc.	Irreversible change to its properties. Reliability reduction.	Capacitor useful life calculation method (it doesn't have a reliable method still) and continuous measurements
Fans malfunction	Operating rating not considered during the operation of the inverter.	Electrical components overheat, such as IGBTs, MOSFETs, diodes, capacitors, inductors, etc., and resulting failure.	Temperature monitoring to indirectly identify a fan malfunction.

2.4 ANOMALY DETECTION OF THE TRANSFORMERS

Transformers are a key component in PV plants as they connect the system to the electric grid, and they provide electric insulation. Nevertheless, failures on the transformer stations such as overheating, and electrical protection failure can lead to drastic consequences such as unplanned outages, loss of assets and profit losses (undelivered energy). In [39] failure rates and impact of transformers have been studied for 15 real operating PV plants in Spain and Italy, monitored for 15 months. This study broke down the transformer station failures into two different categories: failures due to operations causes and due to extreme weather conditions. Despite transformer station registered a quite low rate, weighting for almost 4% of the total failures, their impact was really important as they account for almost 34% of the energy losses [39] . [39] estimates that the energy losses associated to transformer operational failures were around 3000 kWh/failure while extreme weather conditions lead to a loss of 2500 kWh/failure. It is thus paramount, to detect failures in the transformer station (either associated to protection systems or to the transformer itself) at early stage so as to allow prompt intervention and contain the associated losses and downtime. [40] have studied learning models of transformer behaviour for anomaly detection and condition monitoring using Hidden Markov Model (HMM) of healthy transformer behaviour and unexpected operation by processing Ultra-High-Frequency (UHF) data collected through UHF sensors mounted on the transformer. Currently, different AI-based methods have been studied for transformer faults detection and diagnosis. [41] found out that the combination between ANN with evolutionary PSO yields better performance in the transformer oil fault prediction than the commonly adopted Dissolved Gas Analysis (DGA). [42] on the other hand, focused on transformer fault diagnosis. It was noticed that SVM approaches have higher accuracy than ANN due to their better generalization ability. However, online detection of failures in the insulation of the transformer is a very rare process, as transformer are hardly ever equipped with sensor measuring the status of the insulation, but these measurements/tests are rather performed offline, with the transformer disconnected form the plant and deployed in a lab. Transformer monitoring and fault detection are mainly focused on electric faults. Lightening, over-excitation, switching surges, winding resonance, turn to turn short circuit, layer to layer short circuit, partial discharges, insulation tracking, static electrification of oil and flashovers are all forms of electrical failure modes [43]. Table 2.5 summarises some of the AI-based method studied by the scientific community to detect and spot transformer's failures.

TABLE 2.5: SUMMARY OF THE TRANSFORMERS' MAIN FAILURES.

Failure/Defect	Cause	Effect	Detection method
Saturated core	Transformer's primary winding overloaded.	Distorted secondary waveshape. As a result, the protection system will intervene, and the system will be disconnected from the grid.	[44] uses feedforward backpropagation ANN classifier to detect saturated core, based on transformers' Frequency Response Analysis (FRA).
Shorted turn faults	Mechanical damage of insulation, breakdown of electric insulation due to overvoltage from wrong operations or lightning strike.	Massive current flows through the windings, causing overheating and thus power losses. Moreover, as a result electromagnetic forces are applied on	[45] uses Fully Connected Neural Network (FCNN) and Decision Tree, fed with FRA measurements to spot different fault (short circuit between

		the windings threatening and damaging the insulation structure.	turns, broken coil, axial displacement, etc)
Open circuit faults	Broken winding.	Disconnection from the grid. Thus, the power generated is not supplied.	[44] uses feedforward backpropagation ANN classifier to detect open circuit faults, based on transformers' frequency response analysis (FRA).
Partial discharge (PD)	Loss of insulation due to soiling (i.e., dust falling in the transformer winding/core), defects in the structure.	Degradation of the dielectric/ insulating material performance.	[46] uses Adaptive Tabu Search (ATS) based on acoustic measurements to localise PD in power transformers.

3. PV O&M SOLUTIONS

Not only academia has been investigating the DT concept and its applications, as some private companies have been developing AI solutions to improve the performance of PV plants: production forecasting, control, and MPPTs, and PV plants O&M. Both academia and industry have been presenting a good advancement in this area, however, whilst the academia lack the experimental verification in larger PV plants, the industry, and its commercial solutions do not have a DT model as detailed as other applications (like wind power, manufacturing, medical, etc.).

The most common issues of PV plants are listed in Table 3.1 [2] [34]. It can be noted that multiple types of equipment can present multiple faults and failures. However, the inverter is a very important piece of equipment as it is the gateway between the power source and the load (i.e., grid). Some commercial solutions lack the detailing of the inverter, modelling it as a black box. Such important equipment, responsible for most of the faults, failures, and power losses [2] [36] [47], should have an improved DT modelling to better address its problems.

The main goal of O&M technicians and techniques is to increase the availability (uptime divided by the total time) of the PV plant, as the availability depends on the reliability (probability that equipment will perform as expected) [2]. Thus, it is interesting to increase the reliability, and to achieve that some proper monitoring tools should be employed. Generally, most of the energy lost due to an operation or maintenance problem is related to the inverter (up to 50%), followed by the PV modules (up to 30%). Nowadays, some manufactures are greatly improving the data collection of their inverters [48], whilst the PV modules could have the data collection improved by retrofitting [49], even though this is a very expensive employment [2]. Related to non-electrical detection and diagnosis, the main technique is the use of aerial and infrared imaging, captured by drones or UAV [2] [8], using DL algorithms for the image processing. These image DL processing can be combined with the electrical-based detection and diagnosis technique to improve the faults and failures detection and diagnosis.

TABLE 3.1: PV PLANT EQUIPMENT MOST COMMON ISSUES.

Equipment	Faults and Failures
Combiner boxes	Overcurrent (string fuses tripping)
Connectors	Low dc insulation
PV modules	Soiling Partial shading Cracks Encapsulation discolouring Hotspots
Electrical sensor	Open circuit Calibration drift
Power converters	Switches open or short circuit Capacitor or inductor degradation Overheating
Transformers	Overvoltage and overheating

As these techniques are time-consuming, it is advisable to focus at least on those failures that have major impact on the PV plant performance, thus the ones that reduce the availability the most. Along with the data analytics strategies that can be applied for PV plants, the digital twin concept is also a powerful method that can help the fault detection and identification through a root cause analysis.

3.1 THE DIGITAL TWIN CONCEPT

The development of AI and DTs technologies are still in the initial stage for power and energy systems O&M [5] [50] [51]. Similar to other areas of industry, for instance, the DT can be applied to power equipment health state evaluation. Other applications are digital twin modelling of substations, power plant intelligent management, and power equipment failure prediction [52] [53] [54]. It is desired that the digital twin of power systems be data-driven, to work in closed-loop with the physical, and to present a real-time interaction with its real counterpart [54] [55].

The application of AI and DTs solutions is very interesting for PV plants. For instance, if an O&M team is called for a repair without knowing what the problem is, they may not be fully prepared to fix the problem promptly. In some cases, a single tool or component/part that is missing can result in a downtime of days of a solar plant, harming its production. If the O&M team has some initial lead on what caused the problem and some suggestions supplied by an AI, they can be more prepared to tackle the problem and reduce the downtime of the solar plant, saving costs.

Even though the DTs are presented as AI solutions, which is relatively true, in this report they are treated as two distinct technologies. When referring to AI, in this report, it is willing to mention ML or deep learning techniques, whilst when mentioning DT, it is willing to mention modelling of the physical twin that is fed by real data in real-time.

According to the Digital Twin Consortium, a DT is a "virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity." The DT concept was first introduced in 2002 by Michael Grieves, and later formalized in a white paper [56] [57]. A DT must have three basic elements (illustrated in Figure 3-1):

I. A real space containing a physical object;
II. A virtual space containing a virtual object;
III. A link for data flow from virtual to real space and vice-versa.

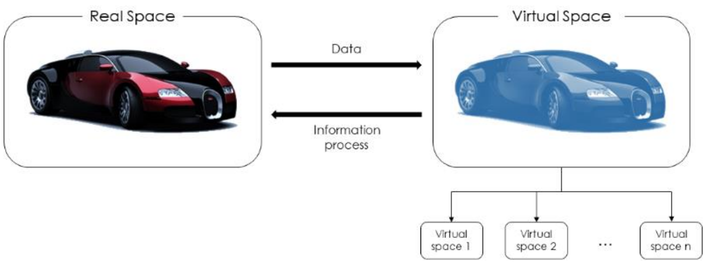


FIGURE 3-1: DIGITAL TWIN MODEL AS SEEN IN [6], AS PROPOSED BY GRIEVES [56].

Following that, it was used by NASA, twining their spacecraft. Then, the US Air Force started to use it as well, using it for forecasting maintenance needs for aircraft. Since then, the DT concept has been evolving ever since [58] [59]. Throughout the literature, the DT had multiple definitions, which are derivations of the one stated by Grieves, adapting it to the application of each study. The definitions have some keywords, such as integrated system; clone, counterpart; ties, links; description, construct, information; simulation, test, prediction; and virtual, mirror, replica. All of these classifications have a similar objective, as the correlation between the words is showing [6].

The difference between a digital twin and a simulation is that a DT is fairly more complex than a simulation, as it is an evolving virtual counterpart of an entity or process. It follows the lifecycle of the physical twin, monitoring, controlling, and optimizing the process [6]. The DT allows the prediction of failures and allows testing novel scenarios to anticipate the maintenance needs. It can be seen as a closed-loop optimization process, where the DT receives the data, process it, and return strategic information to the physical twin for its improvement. It is very important to develop a modular and highly parameterized architecture, to allow the evolution of the DT as a whole or its modules with the physical twin. Nowadays, the DT is applied to manufacturing, medical, transportation, education, business, and other industries [60] [61]. Being an engineering/industrial tool, eventually, the DT began to be applied to power systems [62].

The usage of DT for power electronics systems, which is a core element of a PV plant, has not been widely investigated. Some papers are studying the usage of DT for power systems as a whole, such as grid control, power grids, power consumption, etc. Besides that, a trend can be noticed for the application of DTs for PV systems [34] [63], whilst some have general power electronics application [33] [64] [65]. Of course, detecting a fault or failure is important, but the identification of the problem is very important too. This methodology is commonly named the RCA (Root Cause Analysis). Not all solutions (from academia or industry) have RCA. Some of the commercial solutions, for instance, have only detection algorithms, needing an O&M team to survey the PV plant and do the diagnosis. On the other hand, some academic solutions have a higher degree of detection and diagnosis but have been only validated using simulation of large PV plants or experimental analysis of small PV plants (less than ten PV modules residential strings, for instance) [66].

Even though the digital twin technology has its origins in spacecraft engineering, it expanded to other areas such as manufacturing, aviation (expanding the idea of the first DTs), and healthcare [6]. Naturally, the DT also started to be applied to power systems engineering, creating digital twins power grids, smart grids, wind turbines, substations, power plant management, power equipment, etc. [50] [55] [62].

3.2 STATE OF THE ART FOR THE ACADEMIA

The usage of historical data analytics algorithms combined with electrical modelling of the solar plant for fault and failure detection, diagnosis, and recommendation for O&M reports and scheduling is a powerful tool for technicians. Whilst some commercial solutions already have developed the engineering for such process, academia is improving it trying to develop faster algorithms and/or more detailed digital twins. Some of those researches are focusing on power electronics modelling.

The electrical modelling can be developed into a DT, as some solutions for O&M usually combine DTs with historical data analytics [9]. In that way, some trending on the historical data may be detected and diagnosed as a fault or failure, which would halt the solar plant production. By using forecasting algorithms, the O&M team can schedule a routine before the failure happens, reducing the downtime of the solar plant or performing the maintenance overnight, for instance. Besides that, some underperformance can also be detected and diagnosed by the algorithm, informing that, even though a behaviour that has been observed in the solar plant will not lead to a failure, it will reduce the power plant production (i.e., soiling or inverter overheating).

Some of the data that can be used for detection and diagnosis is already displayed in a SCADA system (which most of the O&M companies provide along with the DTs). However, with the help of AI algorithms and electrical modelling/DT, the problems can be diagnosed earlier, in some cases weeks ahead of a failure [35] [67]. Thus, the solutions that are being proposed are tools that will help the supervisor of a solar plant to identify faults shortly after their trending/deviation start or to detect and diagnose a failure based on the input data from a SCADA system and/or based on the test scenarios developed using a DT. As a solar plant has multiple equipment interconnected, a lot of data is daily recorded. This data can be more useful than simply being displayed on a screen, rather it can be used to help the supervisor on monitoring its solar plant. There are interesting academic solutions that tackle those problems using different approaches.

In [67], a inverter-focused solution is presented. The predictive maintenance system consists of the learning of some behaviours of the inverter (based on data analytics) that will lead to failure. By learning these behaviours, the algorithm can generate alarms, suggesting predictive maintenance of the inverter. This strategy must be fed with historical data of operation and maintenance logs.

In [68], a PV module-centred strategy is developed, looking for problems specifically linked to the modules: hot spots, cracks, soiling, disconnection, etc. This solution may be supported by aerial, thermal images, and power production data. This is an example that the segmentation of the solar plant can be an interesting strategy: the more detailed is the model and the larger the number of collected data is, the more precise the diagnosis will be.

In [66], it is proposed a data-based solution that allows the detection of faults with a prediction rate greater than 90% while being able to recognize "degradation patterns" that would lead to failures up to one weak ahead of time. Even though the solution is feasible and remarkable, it has a single model for the various situations, not considering the seasonality of the solar irradiance for instance. However, this strategy has a great potential for development.

In [69], it is proposed a DTs development based on the analysis of time series and application of DL. They achieved a DT with a reconstruction error of 0.1, which is supported by the use of data processed by characterization techniques in both the time and frequency domain. However, their solution was tested only in specific situations, remembering the gap between academia and industry, and their study did not present experimental results using meteorological data.

In [34], it is proposed a DT development in a FPGA directly connected to a custom-made DC-DC converter interfacing the PV module with a DC-bus with other PV modules. It is presented a very fast-response solution, detecting and diagnosis multiple faults and failures. However, it was tested in a two-PV-modules DC-bus, thus, it still needs to be validated in a large PV plant. Also, the employment of an FPGA for each PV module and its DC-DC converter may be very expensive, similar to the problem of employing retrofitting for data-collection of the PV modules [2]. Besides that, the use of a custom-made power converter and controller (FPGA) may not be feasible in a larger PV plant. Usually, the inverter firmware is not accessible nor the control algorithms. A feasible solution for practical application must consider that some valuable data is not accessible in a real scenario. This is one of the concerns that should be addressed in this area: filling the gap between academia and industry by combining DTs and AI solutions that will be validated using real operational data and tested in real PV farms. This paper has a very interesting solution, however, the gap between academy and industry is its weakest point.

In [70], is presented a 1SVM solution for fault and failure detection of problems on the DC-side (PV modules) of a solar plant. They proved that the 1SVM is a better solution than other clustering algorithms (such as k-means, mean-shift, etc.) for this application. Even though their solution requires a high level of expert knowledge on PV systems (which is not a major problem), they do not make use of SCADA information. The SCADA data is the available information that a commercial PV plant could provide and should be considered when addressing PV plant O&M solutions. Similar to other works, they have not tested its solution using real data of a real system, they have only used simulation data.

In [71], it is proposed an unsupervised fault detection algorithm, featuring an unsupervised reconstruction-based model using variation autoencoder using a generic framework. They achieved a solution that can be applied not only for PV systems but to text and videos, for instance, due to its generic framework. Thus, this is a generalized approach having PV data as a dataset for training and validation. Their study is limited to uni-variable series, which is a limitation due to the multiple correlations between the available data in PV plants. Nevertheless, this work presented a promising strategy.

The use of data analytics to achieve a digital twin PV plant is a relatively new idea, having some publication such as [66] [69], whilst the idea itself has been applied in other fields of industry [72]. This novel idea is to combine the strategy of the digital twin and the data analytics-based forecasting algorithms into a single solution. In that way, the DT that will be achieved is not only developed based on design, modelling, and control of a power converter, nor the predictions are only based on historical data sets. The proposed solution applies the AI algorithms to develop a DT base on the data set, allowing detection, diagnosis, and some level of recommendations. This is an interesting solution since the combination of AI algorithms with DT has been applied to multiple areas of industry, thus, it should be a future improvement for PV plants O&M.

3.3 COMMERCIAL SOLUTIONS

The industry has already been developing AI and/or DTs solutions for O&M of solar farms. However, based on their white papers, blog posts, reviews, etc., it still is a challenge to develop a highly detailed fault and/or failure diagnosis solution. The most promising solutions seem to be the ones that combine historical data analytics with DTs.

GE Renewable Energy [73] have been using the Predix combined with APM (Asset Performance Management) software, their industrial internet platform, to process collected data from solar plants and to provide insightful information about patterns, trends, and behaviours. They exemplify that by having a "crystal ball", the operator of a solar plant can schedule maintenances during the night-time, days ahead of a failure coming to occur. Also, it allows the maintenance team to move to the solar plant site with most of (or all) the tools and spare parts needed for the specific maintenance, avoiding the solar plant to be inoperative during days while the problem is identified and fixed. Their strategy consists of the DT being a "perfectly healthy" doppelganger of solar farms and to use it as a benchmark to detect faults that are resulting in lower efficiency.

They have solar plants powered by GE's inverters, measuring 200 different pieces of data, such as critical components temperatures and voltages. However, in some cases, this may cause a "data-overflow", thus, they advocate that rather than having big data at their disposal, it is better to have intelligent data (this is a recurring worry in this area, as only the useful/valuable data should be stored and shared with the DT). They install small, secure data-collection devices that will feed Predix, which will clean the data, analyse it and turn it into a "morning newspaper" of the solar plant for the operator. Similar to other solutions, at a high level, their Solar Plant software compares the real outputs with the outputs of the DT, tracking the key performance indicators (KPIs) and searching for deviations. Their software is capable of pointing out what/where the solar plant could be modified for a performance improvement and it generates alerts before failures, fostering predictive maintenance.

Their solution has a level of detail that is capable of giving insight for O&M teams of which resources they are going to need to repair, recommending actions to prevent and repair components and subcomponents failures. They summarize their solutions with the following highlights: *understand performance gaps*, *identify areas for improved performance*, and *adopt predictive maintenance strategies*. They can achieve a 40% reduction in power production losses, 30% increase in plant team productivity, and 20% reduction in O&M cost. They present a strong impact solution, using previous tools developed by the company for other areas of industry, but that can be adapted for PV plants.

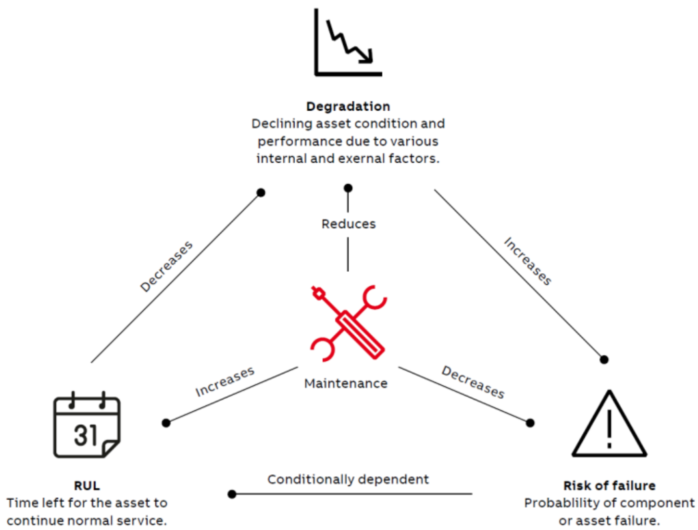
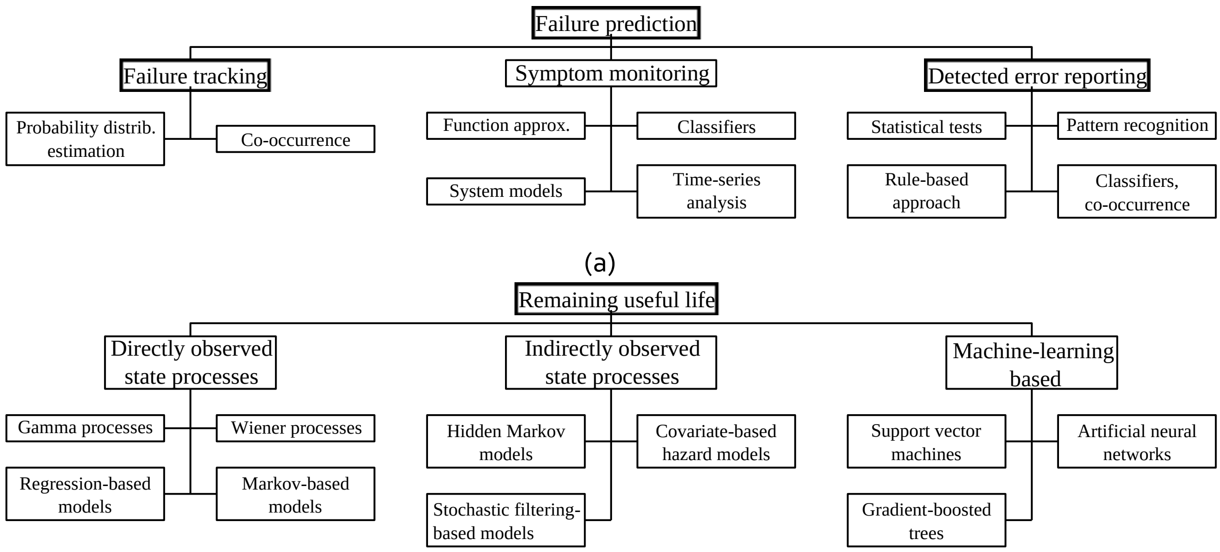


FIGURE 3-2: ABB'S KEY CONCEPTS FOR DT MODELING: DEGRADATION, ROF, AND RUL [48].

ABB [48] advocates that the use of simulations and digital twins is a necessary technology for further development of the industry. The DTs simplify the accessibility of data and verification of scenarios and properties. However, the models do not need to be exact clones of the physical twin, as they should merely reflect the most important asset behaviours to be explored. They have developed the ABB Ability Aurora Vision Plant Management Platform, which has an input of ABBs inverters data and multiple ABB's powered solar plants throughout the world. Their modelling features PV modules, inverters, sensors, meteorological units, energy storage systems, and power grids. They focus on three key interrelated concepts (illustrated in Figure 3-2): *degradation*, *RoF* (Risk of Failure), and *RUL* (Remaining Useful Life).

In their concept, the degradation of a physical asset decreases the RUL and increases the RoF. However, the degradation can be mitigated by maintenance actions to enhance production and performance, before failure or end of life. They also specify the used algorithm for certain types of data, as shown in Figure 3-3. The usage of different algorithms for different faults or failures is a recurrent solution in this area. The data classification is a very important step in their development, as the DT will generate results as good as the data that is fed to it. They have developed a strong solution based on various historical data from weather, production, inverter-metrics, etc., providing a strong competitor for solar power DTs solutions.



(B)

FIGURE 3-3: ABB'S CLASSIFYING ANALYTICS ALGORITHMS [34]: (A) ROFS; AND (B) RULS.

DNV GL [35] calls for a "ramp up" from industry to improve their measurements and data collection, aiming the use of AI technologies for solar plant improvement, embracing predictive maintenance. They also mention that there is a gap between academia and industry and that both should work hand in hand to fill this gap. They call for an approach more inverter-centred, as it is the equipment that fails more often. However, they say that their approach can be extended to any part of the solar plant. They advocate that the predictive maintenance should include:

- Upcoming failure;

- Anomaly detection;
- Performance degradation;
- Planned maintenance;
- Remaining useful life.

They have developed a predictive maintenance system that can detect up to 83% of inverter failures a week ahead of the problem. But this rate is only achievable if the ML algorithm is fed with detailed and high-quality maintenance logs. Later, they developed a more generalized algorithm, which relies on LoF (Local outliner Failure), VAE (Variational Auto Encoders) and AE (Auto Encoders) for baseline modelling. They demonstrated two case studies for their algorithms: predictive maintenance and anomaly detection. They mention that the development of the IoT can allow the evaluation of data that are not usually available in SCADA systems, such as vibration and THD. DNV GL presents a very interesting solution, targeting some key parameters of the inverter, which seems to be an intelligent approach.

Sunsniffer [49] developed a set of hardware and software that help the scheduling of replacement, cleaning, or maintenance of photovoltaic modules. They use additional hardware to be connected to the PV modules (retrofit), improving the data collection and availability: they are focused on the PV module side. The more sensors added to the solar plant, the more data collected, resulting in precise fault detection and diagnosis. They do not seem, however, to have a large development investigating the power electronics equipment.

Reuniwatt [74] has an irradiance-to-production digital twin that combines the estimated irradiance that uses satellite data to calculate the global irradiance, then compares it with the output power of the solar plant, which is provided by the grid manager. Their product allows to spot production reduction over time, which is valuable data for the supervisors, but they lack detailing of the fault detection and diagnosis. They have a solution that sees the whole PV plant as a "black-box". After a fault or failure detection, the diagnosis stage must be done on-site. They present a high-level solution that can detect some production issues but cannot provide root cause analysis. To provide fault diagnosis (or identification) it is required multiple data from multiple points of the solar plant: the more "segmented" the data collected is, the more precise the model is, thus allowing to point where the failure is. This, at first, is a simpler analysis, but it can be greatly improved depending on how much data is available and how interesting is to create such a detailed model of the solar plant.

Other companies also have been developing digital twin solutions for the industry as a whole (such as IBM) or with PV-focused products (such as Pratiti tech), however, it can be noticed that the commercial solutions are mainly focused on: *production evaluation*, if the production is lower than the expected/simulated by the digital twin, the system will alert the supervisor; and *image evaluation*, by visually inspecting the modules for hotspots or crackers, which will lead to failures. These parameters are important, but there is more to explore. The more detailed information extracted from the measured data is generated, a faster and higher precision failure detection and diagnosis system can be implemented.

The detailed digital twinning of the inverter is an interesting development area, as they are responsible for most of the failure events of a solar plant, ranging from roughly 30% up to 90% in some scenarios [36] [47] [67]. It is worth to mention that some studies consider an inverter-fault problem even if the problem's origin is associated with the PV modules, which eventually lead to an inverter fault. To model the inverter as a "black-box" will diminish the evaluation capability of the digital twin. Having current, voltage, and temperature measurements can provide a lot of data to be analysed for trending behaviours of the inverter that can lead to failure, for instance. Thus, the next step of DT for PV O&M should be focused on improving the modelling of the power electronics equipment.

3.4 THE GAPS IN THE SOTA

It can be observed that the presented solutions lack the evaluation of the input and output current and voltages of the power converters, which can lead to the detection and diagnosis of modulation fault (over- and sub-modulation), the evaluation of the THD, PF, etc. These parameters could be evaluated as possible indicators of malfunctioning.

In the area of AI for O&M of solar plants, a major challenge surfaces on the diagnosis [9] [34] on how it is possible to discriminate faults that have the same results (i.e., soiling and cracks might both lead to a reduction of the power output). For instance, for a study that accounts only for the input irradiance and output power on the grid, in case of a fault or failure detection, it does not have enough data to diagnose the problem as an inverter-fault, DC-DC converter-fault, PV module hotspot, PV module crack, PV module disconnection, soiling, etc. If some sensors are added to the inverter, for instance, now this data can be used to verify if, in case of detection, the inverter is presenting some problem or not. Even more: by having thermal data of the panels, it can be diagnosed if a PV module is overheating, thus causing the failure. Having aerial images of the PV modules may also help the diagnosis if the problem is a soiling problem.

There are multiple types of equipment in a solar plant, all of them are prompted to present faults or failures, and it is important to collect data from all over the plant, as long as it is relevant data. If the collected data is not relevant, it will only slow the algorithms, decreasing the failure detection and diagnosis efficiency [35] [48] [69].

Even though both academia and industry have achieved great improvements, there is much to explore. The gaps are present on both sides, and there is also a gap between them. In this sense, the future research should address the gaps in academia, industry, and between those two. Whilst the industry still has to improve the detailing and modularity of the PV plant to develop a RCA of the problems, academia still has to improve the experimental verification of their solutions. Of course, custom-made controllers and power converters are not trivial to implement in a real PV plant, on the other hand, to neglect the information that the power electronics equipment data can provide for the fault and failure detection and diagnosis is a waste. The main gaps that were identified in the literature review of academia solutions and commercial solutions regarding AI and DT for PV plants O&M are listed in Table 3.2.

COMPETE 2020 PORTUGAL 2020 CDTI @CDTIoficial European Union European Regional Development Fund

TABLE 3.2: IDENTIFIED GAPS IN PV PLANTS O&M SOLUTIONS.

	Artificial intelligence	Digital twin
Academia	Training and validation using real data Meteorological data consideration	Generalized approach Scalability of the solutions
Commercial/Industry	Recommendation system Root cause analysis	Modularity PV plant level of detail
Both	Power electronics equipment low level of detail	

Whilst most of the gaps in Table 3.2 are already being addressed, the lack of power electronics equipment investigation is a key point for innovative work. Being this asset responsible for most of the faults, failures, and power losses, the inverter should have more focus on this area of research. Interesting variables of an inverter that could be analysed are over- or sub-modulation, THD, and power factor. All of this information can be retrieved from output currents and voltage measurements. Regarding input current and voltages, some variables can be included: efficiency, maximum power point current and voltage, capacitor RUL.

It would be recommended to segment (modular approach) the solar plant in modules, such as PV modules surface, PV module connections, DC-DC converter, and DC-AC converter (inverter). Each of those modules may be sub-divided, such as PV modules surface can be divided into soiling, hotspot, or crack; PV module connection can be divided into module-to-module, strings, and diodes; DC-DC converter can be divided into PI (Proportional-Integral) control, MPPT, temperature, switches (IGBTs, MOSFETs, diodes, etc.), reactive elements (capacitors and inductors), and current and voltage sensors; the DC-AC converter can be divided into PI control, PLL (Phase-Locked Loop), temperature, switches (IGBTs, MOSFETs, diodes, etc.), reactive elements (capacitors and inductors), grid connection, and current and voltage sensor. It is worth noting that some data or information may not be accessible, such as the MPPT algorithm of the DC-DC converter, the control signal outputs of the inverter, the PLL algorithm for grid-synchronization, etc. Thus, it must be evaluated which data is available for processing and which data is worth processing for fault and failure detection, diagnosis, and recommendation.

4. AI4PV SOLUTIONS

The employment of these two technologies (AI, ML, and DT) to PV systems O&M procedures is interesting. They will help the operators monitoring the PV plants, detect faults in early stages, provide a fast response to failure, and help the scheduling of maintenances ahead of a problem occurrence, reducing the downtime. Being validated as a powerful tool by multiple applications in industry and academia, the use of AI, ML, and DT quickly reached the PV systems and power electronics systems. However, it can be noticed that there is a gap between academy and industry. Also, there is a gap between AI solutions for PV systems and AI solutions for power electronics systems regarding O&M of PV plants.

A preliminary analysis suggests that the equipment that usually presents the highest failure rating (i.e., the inverter or the power electronics systems involved [34] [35] [36] should have deeper modelling, whilst equipment less prompt to failure should be less detailed. The development time should be proportional to these failure ratings. Whilst some papers present a detailed digital twin model of the power electronics equipment, they lack experimental verification on real PV plants. On the other hand, the DTs presented by some O&M companies lack a high level of detailing of the power electronics equipment, presenting an effective DT but that cannot perform RCA, for instance.

As the goal of the project is to develop a generalized solution that can be applicable to multiple PV plants, it is interesting to develop a highly modular approach. This is a directive already recommended for DT development and becomes even more important for the use cases of this project. Developing a model of the multiple types of equipment of a PV plant (PV modules, connectors, inverters, transformers, etc.) allows to easily export the DT to other PV plants, as long as the different equipment (PV modules or inverters, as an example) can have its DT module replaced by the correct model.

Besides this modular approach, which is a recommended practice for the digital twins, when studying solar power, it may be interesting to have multiple models and to process the historical data accordingly to the seasons and PV plant location. Depending on the site of the solar plant, it may go through all four seasons, each season presenting different meteorological data patterns (solar irradiance, ambient temperature, dust, precipitation, snow, etc.). This may also include the aging of the solar plant, as all equipment are subject to deterioration. For instance, the modules are directly exposed to the Sun, rain, dust, snow, high temperatures, etc., whilst the inverter, even though being housed to be protected from the weather, has intrinsic deterioration as its components have a limited lifetime, such as the capacitor and sensors (mostly calibration degradation) [47] [75].

It is worth remembering that an evaluation of the available data must be made to define the use cases. Usually, the inverter cannot provide access to its firmware, sensor gains, MPPT, or PLL algorithms. However, readings as input and output current and voltage, temperature, estimated remaining useful life (RUL) of capacitors, etc., may be easily collected. The same applies to PV modules: multiple parameters may help the fault and failure detection and diagnosis, however, not all of them are available in a real scenario outside of the controlled scenarios of laboratories.

The initial approach suggests dividing the entire AI4PV solution into three modules: AI-ML algorithms, DT, and recommendation systems. The final goal is the recommendation system, but it is not a trivial solution that depends on the development of AI, ML algorithms, and DT. The three modules are:

- *Descriptive analytics-module*: PV plant DT for fault and failure detection and diagnosis;
- *Prescriptive analytics module for O&M*: PV plant data analytics for fault and failure detection and diagnosis;
- *Cost-optimised predictive maintenance module*: PV plant O&M recommendation system.

The first module concerns the study of a digital twin (DT) tool for early fault and failure detection and diagnosis of PV plants. Based on electrical data and meteorological data, a DT system will help the supervisor of the plant to detect the most common problems that may happen in solar parks and pinpointed in the Use Case paragraph in Section 4.1.4.

The second module concerns the study of AI, ML solutions for early fault and failure detection, and the diagnosis of PV plants. Based on historical data analytics, and AI, ML algorithm will help the supervisor of the plant to detect the most common problems addressed by the UCs detailed in Section 4.2.4.

The third module envisions the development of a recommendation system to support the O&M team of PV plants. Based on historical data analytics by AI, ML algorithms combined with a DT tool will help the supervisor of the plant with causes and solutions for the UCs in exam. Besides the AI, ML, and DT tools, the usage of previous maintenance reports will play a key role in the development of the recommendation system.

It is expected that are available data sets and technical data of the power electronics equipment, meteorological readings, PV modules, etc. Also, the input from previous maintenance reports can provide useful information to develop the recommendation system. All of the modules have a strong synergy, as all of them have the same goal. The complete diagram of the systems and the cooperation between the AI4PV modules is seen in Figure 4-1.

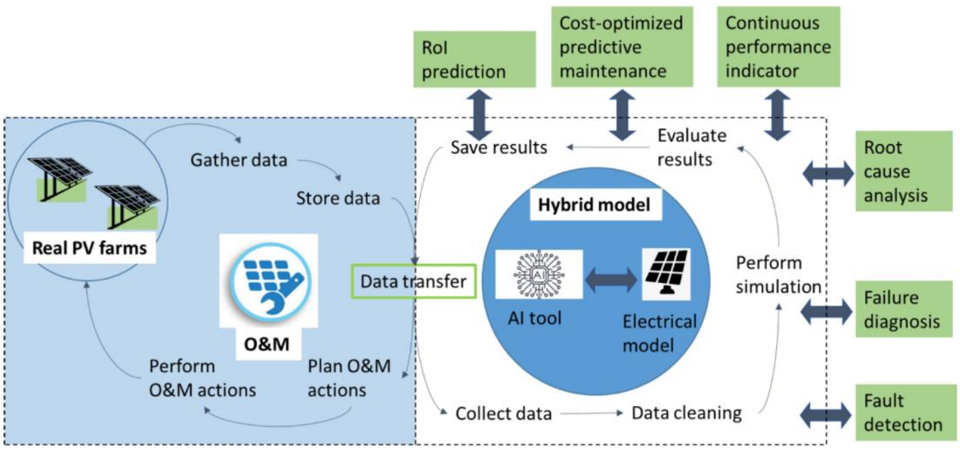


FIGURE 4-1: COMPLETE DIAGRAM OF THE AI4PV SOLUTION.

4.1 DESCRIPTIVE ANALYTICS MODULE OF PV POWER PLANT COMPONENTS AND OPERATION: PV PLANT DT FOR FAULT AND FAILURE DETECTION AND DIAGNOSIS

The first AI4PV module entails the study of a digital twin (DT) tool for early fault and failure detection and diagnosis of PV plants. Based on electrical data and meteorological data, a DT system will help the supervisor of the plant to detect the most common problems that may happen in solar parks representative of the UCs addressed within the project (and described in Section 4.1.4). This module combined with the others will result in fully automated recommendation system, that detects, diagnoses and recommends possible solutions for a problem that may occur in a PV plant. To achieve such a goal, the recommendation systems will rely on AI data analytics and DT modelling. By using these tools, the system will detect, diagnose and recommend possible fixes for the fault or failure in analysis. For instance, if the systems detect an underperformance of the solar park, it will investigate the issue and may diagnose that it is a soiling problem, then, it will recommend an optimal scheduling for the cleaning that will minimize the operational cost.

4.1.1 OBJECTIVES

The ultimate goal of this module is to detect and diagnose fault and failure supporting the O&M team. The following objectives are therefore pursued:

- To detect and diagnose the most common faults, in accordance with the UCs addressed (see 4.1.4);
- To increase the reliability, thus availability, of PV plants by reducing the downtime;
- To increase the number of predictive maintenances and to reduce the number of corrective maintenances;
- To perform the simulation and evaluation of different scenarios to understand what can be improved in the PV plant or what is dragging a better performance of the asset.

In Table 4.1 are shown and described the different actors involved in the DT module.

TABLE 4.1: ACTORS INVOLVED IN THE DT MODULE.

Name	Type	Description
Real PV plant	Facility	Asset to be monitored using the AI, ML algorithms. It is where the data is gathered (currents, voltages, temperatures, solar irradiance, etc.)
O&M Team	Technicians	Technicians that are responsible for the continuous operation of the PV plant and eventual maintenances. They will be assisted on the O&M of the solar farm by the AI, ML algorithms

AI tools	Algorithm	AI, ML algorithm responsible for supporting the O&M team on the supervising of the PV plant. It will be responsible for the faults and failures detection and diagnosis, followed by the recommendation of a solution for the problem.
Digital twin	Virtual model	A virtual object that is a representation of a physical object. It is fed real-time data so it can mimic the behaviour of its real counterpart
PV modules	Asset	The power source of the PV plants that generates electricity via the photovoltaic principle. Are one of the main sources of problems on a PV plant. Requires constant monitoring to ensure the best possible conditions for its operation
Inverter	Asset	Power converter that is the interface/gateway between the PV modules and the grid or ac-load. Encapsulates multiple power electronics technologies, such as dc-dc converter, dc-ac converter, ac-filters, transformers, etc.
Meteorological station	Asset	Data acquisition systems for valuable meteorological data, such as solar irradiance and temperature at the vicinity of the PV modules
SCADA	Virtual interface	A virtual interface that summarizes all valuable data and information of the PV plant. Provides critical information for the operators of a PV plant, supporting the supervision of the facility

4.1.2 MODULE DESCRIPTION

This module addresses the operation and maintenance of photovoltaic plants (solar farms) based on the digital representation of the physical asset via a digital twin tool. This digital twin will be fed with real-time data to generates results as close as possible to the real conditions of the assets considering available data, periodicity, etc.

This will improve the work of the operation and maintenance team by having a benchmark of the PV plant. Based on the expected results for the current conditions, the AI-module through AI, ML algorithms can evaluate the real output and the virtual output. If a deviation is detected, it may indicate a problem at the real asset and trigger alarms.

To achieve such a level of fidelity, it is necessary a data streaming of the main variables of the PV plant, i.e., solar irradiance, temperature, combiner boxes currents and voltages, inverter currents and voltages, power electronics temperatures, etc. Not all this data may be available in a PV plant, thus it is necessary to understand what is available and what is possible to be streamed in real-time (or as close as possible to that). Then, the DT model can be developed to be as precise as possible, depending on the available data. Thus, the PV plant will feed the SCADA, consequently the digital twin will feedback valuable information, scenario analysis, predictions, etc., to the real asset (or to its operator).

It is worth noting that the digital twin should be modular to achieve better modelling of the real asset, but this is directly linked to the available data to feed the digital twin. More than that, it is interesting to have more detailed modelling of the most common assets that are prompted to failure, i.e., PV modules and power electronics components.

The main goal of this module is to support the O&M technicians responsible for the PV plant, to ensure an early-fault detection, reducing downtime of the asset, thus, increasing its productivity. The proposed approach relies on three sequential steps:

i. *Data collection and streaming in real-time*: it is a mandatory condition to develop a digital twin. As the DT will be a benchmark of the physical asset (PV plant), to have an early-fault detection, it is needed to have efficient data streaming from the multiple sensors of the PV plant: pyranometers, temperature sensor, current and voltage sensors, etc. Having this data will help in the simulation of the real asset in real-time;
ii. *Data processing*: having the input of multiple variables of the PV plant, the DT model will process this data by using mathematical equations that model the PV module, inverter, transformer, etc., resulting in benchmarked outputs (voltage, current, power, etc.). Having this benchmark allows the AI-module to compare it with the outputs of the real assets;
iii. *Failure detection and diagnosis*: if a deviation is noted, it may indicate a fault or failure. The more measurement points the PV plant has, the more modular the modelling of the DT will be, thus allowing to perform a RCA. Identifying the exact cause and location of a fault or failure is not trivial, thus the use of a DT with AI, ML algorithm is necessary.

The ***real-time monitoring*** is the first step and it aims to constantly read the data fed to the SCADA. This monitoring aims to feed the digital twin with real-time data, so it can develop the benchmark results that will be compared with the real asset measurements. The period and discretization of the data to be provided by the SCADA system must be compatible with the requirements of the DT. Different data should be collected:

a) *meteorological data*: The meteorological data of the surroundings of the PV plant is very important to understand the conditions of the input power of the system.
b) *PV modules data:* The PV modules are the power source of a PV plant, thus, have readings such as current, voltage, and temperature can be very valuable to feed the DT model and understand and retrieve the operating conditions.
c) *Inverter data:* The inverter connects the PV modules to the grid or ac load; thus, it is the gateway of the produced energy of a PV plant. This is one of the main assets of a PV plant and usually it is where most of the failures occur.
d) *Transformer data:* Data on the grid-connection point are necessary to model and retrieve the conditions of the transformer in order to point out and spot eventual faults and failures of this component.

The ***simulation*** is the second step and it consists of the processing of the collected data by the mathematical modelling of the DT. It is a very important step of the DT loop, as a DT is as good as its model. The model takes into consideration the available data that can be fed to the DT. However, it is very important to not use "not-so-valuable-data", as it will increase the processing time and will result in a poor model. The irradiance, temperatures, currents, voltages, etc., are processed and a benchmarked output is generated. This is the expected output for an operation without problems. The smaller the error between the real output and the virtual output, the better, and this indicates that all the assets are operating as expected. However, some deviations will eventually appear

Finally, the third step is the ***comparison*** between the expected and real output. This can be done by an error analysis; however, the more data is involved, the more complex this process is. That's why the use of AI, ML algorithms can be an interesting tool for this analysis. Based on threshold evaluation, the algorithm may detect a fault or failure, that based on the historical data analytics done by the AI-module, combined with the information provided by the DT, can be diagnosed. This is a challenging task, as most of the DT solutions nowadays are for reduced-scale PV plants or have a simple model that is not capable of performing a RCA. Besides that, the DT can be used to test different scenarios of the real asset. That's why its modularity is so important: to evaluate multiple scenarios is a powerful tool for O&M and business planning.

4.1.3 MODULE REQUIREMENTS AND CONDITIONS

The following inputs must be available:

- Electrical power electronics assets data available for real-time feed, such as current and voltage readings from the PV modules, inverter, transformers, temperature measurements, etc.;
- Meteorological (environmental) measurements in the vicinity of the PV modules/solar farm, or there are some readings from satellites, such as solar irradiance and temperature in real-time.

The following prerequisites must be achieved:

- Data streaming of valuable data from the PV modules, inverter, and meteorological parameters;
- Technical information on the electrical components (datasheets), such as PV modules ratings and inverter ratings.

4.1.4 USE CASES FOR THE DIGITAL TWIN

Table 4.2 summarises the UCs that will entail the use of the DT for descriptive analytics, as well as the partner(s) responsible of the development and the task of the project in which they will be tackled. In particular, through machine learning and optimization algorithm, it will be determined the parameters of the "normal" operation of the elements of the PV plant in different situation of operations (weather, month, irradiance, clouds, etc). Automatic out of normality detection algorithms will be developed, so real data can be matched with the models with different combinations of parameters and data analytics comparison for diagnosis of the problems (i.e., soiling and inverter malfunction) and current situation of the assets.

TABLE 4.2: SUMMARY OF THE DESCRIPTIVE ANALYTICS-RELATED USE CASES.

Module	Application/ Use case		Responsible	Participant	Task
Descriptive analytics	PV panels	Soiling	ISOTROL	EDP	2.3
	Inverter	Inverter shutdown, temperature disconnection, maintenance stop, late awakening. clipping, MPPT optimal point. Out of normality analysis,	ISOTROL	EDP	2.3
	Solar field problems	Solar field incidences detection: string & stringbox disconnection, tracker blocking, tracker misalignment, panel ageing	ISOTROL	EDP	2.3
		Model based sensor malfunction detection & measurement correction (pyranometers, currents, power)	ISOTROL	EDP	2.3

4.2 PRESCRIPTIVE ANALYTICS MODULE FOR O&M: PV PLANT DATA ANALYTICS FOR FAULT AND FAILURE DETECTION AND DIAGNOSIS

This module entails the use of AI and ML solutions for early fault and failure detection, and the diagnosis of PV plants. Based on historical data analytics, and AI, ML algorithms will help the supervisor of the plant to detect the most common problems that may happen in solar parks in accordance with the identified use cases. The combination of the AI-module and the DT-module will result in a fully automated recommendation systems, that detect, diagnose, and recommend possible solutions for a problem that may occur in a PV plant.

4.2.1 OBJECTIVES

The ultimate goal of this module is to detect and diagnose faults and failures that may occur in a PV plant, supporting the O&M team in the troubleshooting. The following objectives are therefore pursued:

- To detect and diagnose the most common faults, in accordance with the UCs addressed (see Section 4.2.4);
- To increase the reliability, thus availability, of PV plants by reducing the downtime;
- To increase the number of predictive maintenances and to reduce the number of corrective maintenances;
- To perform the simulation and evaluation of different scenarios to understand what can be improved in the PV plant or what is dragging a better performance of the asset.

4.2.2 MODULE DESCRIPTION

This module aims at supporting the operation and maintenance of photovoltaic plants (solar farms) through the employment of AI, ML algorithms for monitoring multiple readings from the solar farm sensors in real-time (or as close as possible).

This will improve the work of the operation and maintenance team, allowing the early detection of problems that may surface in the PV plant, resulting in a faster response for predictive and preventive maintenance.

The ML, AI algorithms take into consideration the available data that feed the SCADA system to recognize patterns and/or trending on the data that may lead to a fault or failure. This analysis takes into consideration power electronics data (current and voltage readings from the inverter and/or the photovoltaic modules, power converter temperatures, etc.) and meteorological data (solar irradiance, temperature, wind speed, etc.).

In real-time (or as close as possible), the state of the photovoltaic plant is analysed ad if there is any significant deviation from the normal scenario/operation, the AI, ML algorithms will detect it before a malfunction. If the early detection of the AI, ML algorithm fails and a failure comes to happen, the algorithm should alert the operator, nevertheless. Most common problems can also have a recommendation attached to their diagnosis, helping in the maintenance process.

The main objective of the AI-module is to monitor the available data of the SCADA system and ensure that any out-of-normality scenario is notified to the operators. In case of any deviation (such as overheating, disconnection, soiling, etc.), a set of steps is taken by the algorithm to identify the problem and may recommend some possible solutions/fixes. The proposed approach will rely on two different and sequential stages:

i. *Data processing and fault detection*: In the first stage, the AI, ML algorithm will be monitoring the data from the SCADA system in real-time. The data is composed of common measurements of a PV plant, such as current, voltages, irradiance, temperatures, etc. Whenever potential deviations are identified, a detection is triggered. The detection itself though is not capable of identifying the problem, as this is the second step of the algorithm operation;

ii. *Fault diagnosis:* In the second stage, the AI, ML algorithms will be working on the identification of the problem. Based on a set of pre-determined scenarios in comparison with the real-time fed data, the AI, ML algorithm will identify and diagnosis the fault or failure, pointing out as specifically as possible (depending on the available data) where is the problem and what caused it. Then, based on the historical data analysis in combination with previous maintenance reports, the AI, ML algorithm may suggest some recommendations for the possible problem.

The ***real-time monitoring*** is the first step of the analysis. It aims to constantly read the data fed to the SCADA. This monitoring aims the mitigation of eventual problems that may occur in the PV plant, doing the early detection of faults and/or failures. The period and discretization of the data to be provided by the SCADA system must be compatible with the requirements of the AI, ML algorithm:

1. *meteorological data*: The meteorological data of the surroundings of the PV plant is very important to understand the conditions of the input power of the system.
2. *PV modules data:* The PV modules are the power source of a PV plant, thus, have readings such as current, voltage, and temperature can be very valuable to feed the DT model and understand and retrieve the operating conditions.
3. *Inverter data:* The inverter connects the PV modules to the grid or ac load; thus, it is the gateway of the produced energy of a PV plant. This is one of the main assets of a PV plant and usually it is where most of the failures occur.
4. *Transformer data:* Data on the grid-connection point are necessary to model and retrieve the conditions of the transformer in order to point out and spot eventual faults and failures of this component.

The ***detection***, second step of the process, consists of the analysis of multiple variables from the PV plant, especially the electrical ones. It is the first step on the diagnosis and, later, the recommendation for the operation and maintenance team. The electrical variables are monitored, and any out-of-normality behaviour should be noted. If a pre-defined threshold trespasses, a problem is detected. However, a problem may have multiple causes and lead to multiple consequences, thus a deeper analysis of what is happening must be made.

Finally, after one or multiple deviations are detected by the AI, ML algorithm, the ***diagnosis*** stage starts. The analysis of the variables that present deviation will lead to the identification of the fault or failure. In this stage, it is very important to understand the correlation between the variables, their causes, and consequences. Whilst the detection is a relatively complex stage, the diagnosis is even more complex. False diagnosis may lead to dangerous scenarios, larger PV plant downtime, and wrong recommendation. Once a problem is detected and diagnosed, a list of recommendations may be provided to the operation and maintenance team based on previous maintenance reports. This aims to provide an initial discussion on the problem and how to fix it, saving time for the operation and maintenance team and, consequently, reducing the downtime.

4.2.3 MODULE REQUIREMENTS AND CONDITIONS

The following inputs must be available:

- Electrical power electronics assets data available for real-time feed, such as current and voltage readings from the PV modules, inverter, transformers, temperature measurements, etc.;
- Meteorological (environmental) measurements in the vicinity of the PV modules/solar farm, or there are some readings from satellites, such as solar irradiance and temperature in real-time.
- Maintenance reports that can provide useful data for the recommendation system.

The following prerequisites must be achieved:

- Dataset for training and validation of the algorithms (power electronics assets, PV modules, transformers and meteorological data);
- Technical information on the electrical components (datasheets), such as PV modules ratings and inverter ratings.

4.2.4 USE CASES FOR O&M PRESCRIPTIVE ANALYTICS TOOLS

Table 4.3 summarises the UCs that will entail the use of the AI algorithms for perspective analytics, as well as the partner(s) responsible of the development and the task of the project in which they will be tackled. In particular faults and failures of the three main components (PV panels, inverter and transformer) will be addressed.

TABLE 4.3: SUMMARY OF THE USE CASES RELATED TO THE PERSPECTIVE ANALYTICS TOOL.

Module	Application/ Use case		Responsible	Participant	Task
Prescriptive analytics tool: root cause analysis	PV panels	Soiling	EDP		3.1
	Inverter	IGBTs, MOSFETs or diodes malfunction	INESC TEC	ISOTROL	3.1
		Reactive components degradation, mainly the capacitors			
		Fans malfunction			

		Inverter shutdown, temperature disconnection, maintenance stop, late awakening	ISOTROL	INESC TEC	3.1
	Transformer	Underperformance	EDP	INESC TEC	3.1
		Open circuit			
		Short circuit			

4.3 COST-OPTIMIZED PREDICTIVE MAINTENANCE MODULE: PV PLANT O&M RECOMMENDATION SYSTEM

This module aims at supporting the O&M team of PV plants. Based on historical data analytics by AI, ML algorithms combined with a DT tool will help the supervisor of the plant with causes and solutions of the most common problems of a PV plant. Besides the AI, ML, and DT tools, the usage of previous maintenance reports will play a key role in the development of the recommendation system.

4.3.1 OBJECTIVES

The objectives of this module is to recommend solutions for faults and failures, so as to support the O&M team. The following objectives are therefore pursued:

- To present the possible causes of a fault or failure at the PV plant;
- To present the possible solutions for a fault or failure at the PV plant, including spare parts, tools, etc., that may be needed for the repair or replacement of the failed components;
- To increase the reliability, thus availability, of PV plants by reducing the downtime;
- To increase the number of predictive maintenances and to reduce the number of corrective maintenances.

4.3.2 MODULE DESCRIPTION

This module will help the O&M team to respond faster to a fault or failure at the PV plant. As a basis, this recommendation system has an AI, ML system that works in cooperation with a DT of the real PV plant.

The recommendation system takes into consideration the inputs from the AI-module and DT-module benchmark analysis, and by combining it with historical information on previous maintenances, can provide insightful directions for the technicians. Base on the development of the AI, ML algorithms, DT, and available maintenances reports, the recommendation system can range from suggestions for solutions to problems related to the PV panels, power electronics converters, transformers, connectors, etc.

Being the last step on the root-cause analysis of a fault or failure, the recommendation system will have detected, diagnosed, and recommended possible causes, solutions, parts, tools, etc., that will help the O&M technicians on the predictive or corrective maintenances of the PV plant and its equipment.

The main objective of this module is to support the O&M team by doing a preliminary analysis of the problem by applying a root-cause analysis. The system will detect, diagnose and provide some recommendations about the problem, supporting the O&M team with information on what caused the problem, when it happened, and where it happened. Besides that, it will provide a list of possible solutions, parts, and tools that may help the technicians to fix the problem. The proposed approach will rely on three different pillars:

i. <u>*Digital twin*</u> to achieve the benchmarked state of the PV plant or its expected state of operation. By having a reference on how the PV plant should behaviour, i.e, currents, voltages, output power, etc., it is possible to do a comparative analysis with the real outputs. Based on that, any deviations from the expected result may indicate that a fault or a failure is happening;
ii. <u>*AI-ML algorithms*</u> will be responsible for the historical data analytics, to perform early-fault or failure detection based on trending, deviations, etc., that may be noticed on the measured data (mostly electrical data). Also, it will verify if the real PV plant is presenting a behaviour like its DT, as a deviation between them appear it may be indicative of a fault or a failure. Combining the AI, ML with the DT, it is possible to detect and diagnose faults or failures at the PV plant and to perform a root-cause analysis;
iii. <u>*Fault diagnosis:*</u> the recommendation system will investigate previous maintenances reports on what causes similar faults or failures similar to the one that has been diagnosed. In that way, it will be possible to provide insightful information to the O&M team, informing replacement parts, tools, causes, etc., of a fault or a failure. Thus, based on the historical data analysis in combination with previous maintenance reports, the AI, ML algorithm may suggest some recommendations for the possible problem.

Detection and diagnosis is the first step of the recommendation process. The AI, ML in cooperation with the DT will be responsible for detecting and diagnosing the faults and failures of the PV plant. They will be fed with electrical readings, irradiance and temperature readings, electrical equipment temperature readings, etc., to process the data and to investigate any trending or deviation between the PV plant and its DT.

The ***root-cause analysis*** consists in the most precise identification of a problem, by clearly describing it, establishing a backtrack of what led to the problem, and developing a graph to identify the root cause of the problem. This is an intermediate step as it will work as a filter for the multiple solutions that the recommendation system may provide to the O&M team.

After identifying the problem, what caused it, where it happened, etc., the ***recommendation of possible solutions*** for troubleshooting will take place., based on previous experience and reports. This may indicate, for instance, methodologies, parts, tools, etc., that were used to solve a similar problem before. This will help the O&M team to start with a strong set of information about the problem, saving the time that would be spent on the investigation of the problem. The system will be based on the most common failures but should be configurable for different PV plants in different scenarios, adding or removing solutions from the database.

4.3.3 MODULE REQUIREMENTS AND CONDITIONS

The following assumptions must be fulfilled:

- There is a DT model of the PV plant that is generating the expected state/results of the asset;
- There is AI, ML algorithms constantly monitoring the PV plant historical data and DT benchmark states, investigating and trending or deviations that may appear between the PV plant and its virtual counterpart;
- There are maintenance reports that can provide useful data for the recommendation system.

The following prerequisites must be achieved:

- Fault and failure detection and diagnosis systems to carry out the root cause analysis of the problems;
- Database with previous solutions for similar problems that may be diagnosed by the AI, ML algorithms in cooperation with the DT.

4.3.4 USE CASES FOR COST-OPTIMISED PREDICTIVE MAINTENANCE

Table 4.4 Table 4.3 summarises the UCs that will entail the use of the AI algorithms for perspective analytics, as well as the partner(s) responsible of the development and the task of the project in which they will be tackled. In particular two main tasks will be studied: the optimal scheduling of the O&M tasks (taking into account also meteorological parameters, i.e., rainfall that might impact on the cost) and asset replacement.

The optimal scheduling will leverage from an AI-based techno-economic model developed in Task 3.2 that will be used to predict the RoI of a single PV power plant, and thus assess the impact of different O&M policies, considering parameters such as: O&M costs, assets lifetime, PV resource.

Moreover this module will produce advices for O&M planning by prioritizing actions, asset replacement and preventive maintenance tasks so as to optimise the overall RoI.

TABLE 4.4: SUMMARY OF THE USE CASES RELATED TO COST-OPTIMISED PREDICTIVE MAINTENANCE.

Module	Application/ Use case	Responsible	Participant	Task
Cost-optimised predictive maintenance approach	Optimal O&M scheduling for RoI optimisation	EDP	ISOTROL	3.2-3.3
	Asset replacement – Action prioritisation	INESC TEC	ALL	3.1 -3.3

5. AI4PV KEY PERFORMANCE INDICATORS

In Table 5.1 are listed some of the KPIs that will be monitored during the AI4PV test campaign so as to validate the developed solutions. Quantifiable targets were also identified as well as methods to compute the addressed KPIs.

TABLE 5.1: AI4PV'S LIST OF KPIS.

#	Name	Description	Formula	Target
KPI1	Root mean squared error (RMSE) between empirical and reproduced I-V curve	It represents the difference between the empirical I-V curve provided in the datasheet of the PV module and the reproduced curve through the DT modelling	$$RMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I - I_i)^2}}{Isc}$$ Where: • $I_i, \hat{I}_i$ are the real and modelled output current of the PV module. • N is the number of samples of the empirical I-V curve • Isc it's the short circuit current of the PV module	<0.6
KPI2	Reduced soiling losses (RSL)	It represents the ratio between the energy of the soiled PV panel and the cleaned one. The higher it is, the more cleaned the PV is for a long period of time. It considers losses due to both dust or organic soiling	$$RSL = \frac{\int_0^T P_{PV\_soiled}\,dt}{\int_0^T P_{PV\_cleaned}\,dt}$$ Where: • $P_{PV\_soiled}, P_{PV\_cleaned}$ are the output power of the soiled and cleaned PV; • T is the observation time, it can be 1 week, 1 month, etc	>80%
KPI3	Number faults and/or failures detected automatically through data analysis	The inspection of the SCADA and sensor data of the inverter by AI, ML algorithms will detect trending and deviations in the measurements	n.a.	8

		that may indicate a fault or a failure in the PV plant		
KPI4	Fault Detection accuracy	It's the ratio between true faults detected by AI4PV and real faults	$FDA = \frac{N\_true\_positive\_fault}{N\_true\_positive\_fault + N\_false\_positive\_fault}\%$	>80%
KPI5	Number of maintenance actions at validation site	Depending on the output of the recommendation system, predictive maintenance may be carried out to avoid failures. It is the number of interventions advised to the O&M team by AI4PV recommender system.	n.a.	10 actions/month
KPI6	Recommendation accuracy (RA)	Number of correct recommendations	$RA = \frac{N\_good\_recommendation}{N\_tot\_recommendation}\%$	>70%
KPI7	Percentage of losses & degradation underperformance quantification (AEL_UD)	The early detection of faults in the PV plant is important to avoid power losses that, otherwise, would be undetected until a failure occurs	$AEL\_UD = \frac{\int Psaved\ dt}{\int Ptot\ dt}\%$	< 5%
KPI8	Avoided energy losses due to early detection	Avoided energy losses due to fault detection at early stage	$AEL\_ED = \frac{\int Psaved\ dt}{\int Ptot\ dt}\%$	4%

	problems (AEL_ED)			
KPI9	Reduce unexpected outages (RUO) in the transformer stations	It's the ratio between the outages registered with the AI4PV solutions in place, and the ones registered without AI4PV. The outages are avoided through early detection of failures that would allow to intervene before the worsening of the failure.	$$RUO = \frac{Out\_AI4PV}{Out\_noAI4PV}$$ Where: • Out_AI4PV, are the outages registered with AI4PV solutions in place • $Out\_noAI4PV$ are the outages registered without AI4PV solutions in place	<96%
KPI10	Reduce response time	It is the time between failure occurrence and detection	$$RRT = \frac{RT_{AI4PV}}{RT_{conventional}}\%$$ Where: • RT_{AI4PV} is the response time with AI4PV in place, for a particular failure; • $RT_{conventional}$ is the conventional response time (without AI4PV) for a particular failure.	<90%
KPI11	Plant availability increase (PAI)	It is the number of working hours ensured by AI4PV by reducing the number of downtimes.	$$PAI = \frac{N_{hours\_availability\_wAI4PV} - N_{hours_{availability_w}/outAI4PV}}{N_{hours_{availability_w}/outAI4PV}}$$ Where: • $N_{hours\_availability\_wAI4PV}$ is the number of working hours of the PV plant with AI4PV solutions in place • $N_{hours_{availability_w}/outAI4PV}$ is the number of working hours of the PV plant without AI4PV solutions in place	>5%

6. CONCLUSION

This report presented firstly the most pressing challenges that PV operators face during the operation and maintenance of PV parks. A set of possible applications, in which the application of AI solutions can be key to fulfil user expectation and needs, were identified. A review on AI, ML algorithms and DT solutions for O&M of PV plants is presented as well. A generalized approach is done for those technologies, later focusing on power, PV systems and its solutions for O&M usage. Both white and scientific papers were reviewed always looking for the categorization of the advancements and challenges of the technology applied to PV plants O&M. Base on that, AI4PV solutions and modules were proposed updating the project proposal with the most recent discussing regarding this topic. The modules were detailed in objectives, conditions, and use cases addressed within the project, resulting in the road map for the next steps of the project. Finally, a set of KPIs to be monitored during the validation phase were defined, so as to validate and benchmark AI4PV final solutions against State-of-the-Art technologies.

7. REFERENCES

[1] Photovoltaic Power Systems Programme, *Snapshot of Global PV Markets 2021*, April 2021.

[2] G. D. Lorenzo, "Review of O amp;M Practices in PV Plants: Failures, Solutions, Remote Control, and Monitoring Tools.," *IEEE Journal of Photovoltaics 10.4 (2020)*, pp. 914-926.

[3] Rolf Isermann, "Fault-Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance," *Springer-Verlag Berlin Heidelberg*, p. 28, 2006.

[4] Shuai Zhao, Frede Blaabjerg, and Huai Wang., "An Overview of Artificial Intelligence Applications for Power Electronics.," *IEEE Transactions on Power Electronics*, no. 36.4, pp. 4633-4658, 2021.

[5] Huaming Pan et al., "Digital Twin and Its Application in Power System," in *5th International Conference on Power and Renewable Energy (ICPRE)*, 2020.

[6] Barbara Rita Barricelli, Elena Casiraghi, Daniela Fogli, "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications," *IEEE Access*, no. 7, pp. 167653-167671, 2019.

[7] Adel Mellit, Soteris A. Kalogirou, "Chapter II-1-D - A Survey on the Application of Artificial Intelligence Techniques for Photovoltaic Systems," *McEvoy's Handbook of Photovoltaics (Third Edition)*, pp. 735-761, 2018.

[8] Giuseppe Marco Tina et as, "A State-of-Art-Review on Machine-Learning Based Methods for PV.," *Applied Sciences*, no. 11.16, 2021.

[9] Adel Mellit, Soteris Kalogirou, "Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: Challenges, recommendations and future directions," *Renewable and Sustainable Energy Reviews*, no. 143, p. 110889, 2021.

[10] Li Bian et al, "A multifunctional data acquisition system for photovoltaic plants," *International Conference on Systems and Informatics (ICSAI2012)*, pp. 598-602, 2012.

[11] Kurukuru, V.S.B.; Haque, A.; Khan, M.A.; Sahoo, S.; Malik, A.; Blaabjerg, F, "A Review on Artificial Intelligence Applications for Grid-Connected Solar Photovoltaic Systems..," *Energies*, no. 14, 2021.

[12] A. Tived, "Artificial Intelligence in the Solar PV value chain : Current applications and future prospects," *Dissertation, KTH*, 2020.

[13] A. Mellit, A. M. Pavan, "24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy.," *Solar Energy*, vol. 84, pp. 807-821, 2010.

[14] Whang, Zhe, Wang, Fei, & Su, Shi, "Solar Irradiance Short-Term Prediction Model Based on BP Neural Network," *Energy Procedia*, vol. 12, pp. 488-494.

[15] Chen, S.X., Gooi, H.B., & Wang, M.Q., "Solar radiation forecast based on fuzzy logic and neural networks.," *Renewable energ*, vol. 60, pp. 195-201, 2013.

[16] K. Baharin, Abd Rahman, Hassan, & Chin Kim Gan, "Hourly irradiance forecasting in Malaysia using support vector machine," in *IEEE Conference on Energy Conversion (CENCON)*, 2014.

[17] G. Li, H. Wang, S. Zhang, J. Xin & H. Liu, "Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach," *Energies:*, vol. 12, 2019.

[18] J. Liu, W. Fang, X. Zhang, & C. Yang, "An Improved Photovoltaic Power Forecasting Model With the Assistance of Aerosol Index Data. ,," *IEEE Transactions on Sustainable Energ*, vol. 6, pp. 434-442, 2015.

[19] Yuqi Tao and Yuguo Chen, "Distributed PV power forecasting using genetic algorithm based neural network approach," in *Proceedings of the 2014 International Conference on Advanced Mechatronic Systems*, 2014.

[20] F-M. Petcut & T- Leonida-Dragomir, "Solar Cell Parameter Identification Using Genetic Algorithms," *CEAI*, vol. 12, no. 1, pp. 30-37, 2010.

[21] S. Ebrahimi, Salahshour, Malekzadeh, & Francisco Gordillo, "Parameters identification of PV solar cells and modules using flexible particle swarm optimization algorithm," *Energy*, vol. 179, p. 358–372, 2019.

[22] R. Wang, Y. Zhan, & H. Zhou, "Application of Artificial Bee Colony in Model Parameter Identification of Solar Cells," *Energies*, vol. 8, no. 8, p. 7563–7581, 2015.

[23] Vatankhah, R.; Ghanatian, M., "Artificial neural networks and adaptive neuro-fuzzy inference systems for parameter identification of dynamic systems," *Journal of Intelligent & Fuzzy Systems*, vol. 39, pp. 6145-6155, 2020.

[24] Köntges M, Kurtz S, Packard CE, Jahn U, Berger K, Kato K, et al., "Review of Failures of," *Photovoltaic Power System Program Task 13, IEA-PVPS*, Vols. T13-ol, 2014.

[25] Köntges M, Oreski G, Jahn U, Herz M, Hacke P, K. A Weiss, et al., "Assessment of Photovoltaic Module Failures in the Field," *Photovoltaic Power System Program Task 13 - subtask 3, IEA-PVPS*, 2017.

[26] Dhimish, Mahmoud, Holmes, Violeta, Mehrdadi, Bruce, Dales, Mark and Mather, Peter, "Detecting Defective Bypass Diodes in Photovoltaic Modules using Mamdani Fuzzy Logic System.," *Global Journal of Researches in Engineering: Electrical and Electronics Engineerin*, vol. 17, no. 5, pp. 33-44, 2017.

[27] X. Li, J. Wang, and Z. Chen, "Intelligent fault pattern recognition of aerial photovoltaic module images based on deep learning technique," *J. Syst. Cybern. Inf*, vol. 16, p. 67–71, 2018.

[28] Parikh, H.R.; Buratti, Y.; Spataru, S.; Villebro, F.; Reis Benatto, G.A.D.; Poulsen, P.B.; Wendlandt, S.; Kerekes, T.; Sera, D.; Hameiri, Z., "Solar Cell Cracks and Finger Failure Detection Using Statistical Parameters of Electroluminescence Images and Machine Learning," *Applied Science*, vol. 10, 2020.

[29] Mingjian Sun, Shengmiao Lv, Xue Zhao, Ruya Li, Wenhan Zhang and Xiao Zhang, "Defect detection of Photovoltaic modules based on Convolutional Neural Network," in *Machine Learning and Intelligent Communications. Second International Conference, MLICOM 2017, Weihai, China, August 5-6*, 2018.

[30] Simal Pérez, N.; Alonso-Montesinos, J.; Batlles, F.J. , "Estimation of Soiling Losses from an Experimental Photovoltaic Plant Using Artificial Intelligence Techniques," *Applied. Science*, vol. 11, p. 1516, 2021.

[31] M. Coello and L. Boyle, "Simple Model for Predicting Time Series Soiling of Photovoltaic Panels," *IEEE Journal of Photovoltaics*, vol. 9, no. 5, pp. 1382-1387, 2019.

[32] M. Yang, J. Ji and B. Guo, "An Image-Based Method for Soiling Quantification," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)," 2020.

[33] Matthew Milton et al. , "Controller-Embeddable Probabilistic Real-Time Digital Twins for Power Electronic Converter Diagnostics," *IEEE Transactions on Power Electronics*, vol. 35.9, p. 9850–9864, 2020.

[34] Palak Jain et al., " A Digital Twin Approach for Fault Diagnosis in Distributed Photovoltaic Systems," *IEEE Transactions on Power Electronics*, vol. 35.1, p. 940–956, 2020.

[35] GreenPowerMonitor, *Position Paper: PREDICTIVE MAINTENANCEOF SOLAR PV PLANTS: THE TIME IS NOW*, DNV GL GROUP TECHNOLOGY & RESEARCH, 2021.

[36] Geoffrey Taylor Klise, Olga Lavrova, and Renee Lynne Gooding, "PV System Component Fault and Failure Compilation and Analysis," 2018.

[37] Z. J. Ma and S. Thomas, "Reliability and maintainability in photovoltaic inverter design," in *Proceedings - Annual Reliability and Maintainability Symposium*, 2011.

[38] X. Wang, B. Yang, Q. Liu, J. Tu and C. Chen, "Diagnosis for IGBT Open-circuit Faults in Photovoltaic Inverters: A Compressed Sensing and CNN based Method," in *IEEE 19th International Conference on Industrial Informatics (INDIN)*, 2021.

[39] Lillo-Bravo, I.; González-Martínez, P.; Larrañeta, M.; Guasumba-Codena, J. , "Impact of Energy Losses Due to Failures on Photovoltaic Plant Energy Balance," *Energies*, vol. 11, p. 363, 2018.

[40] Brown, A.J. and Catterson, V.M. and Fox, M. and Long, D. and McArthur, "Learning models of plant behavior for anomaly detection and condition monitoring," *Engineering Intelligent Systems for Electrical Engineering and Communications*, vol. 15, no. 2, pp. 61-67, 2017.

[41] Illias, Hazlee Azil et al., "Transformer Incipient Fault Prediction Using Combined Artificial Neural Network and Various Particle Swarm Optimisation Techniques," *PloS one*, vol. 10.6, 2015.

[42] Sahoo S., Chowdary K.V.V.S.R., Das S, "DGA and AI Technique for Fault Diagnosis in Distribution Transformer," *Advances in Smart Grid and Renewable Energy. ETAEERE 2020. Lecture Notes in Electrical Engineering*, vol. 691, 2020.

[43] J. &. S. S. (. Maan, "Transformer Failure Analysis:Reasons and Methods.," *International Journal of Engineering Research & Technology (IJERT)*, vol. 4, no. 15, 2016.

[44] Tahir, M.; Tenbohlen, S, " Transformer Winding Condition Assessment Using Feedforward Artificial Neural Network and Frequency Response Measurements," *Energies* , vol. 14, 2021.

[45] Li, Z.; Zhang, Y.; Abu-Siada, A.; Chen, X.; Li, Z.; Xu, Y.; Zhang, L.; Tong, Y, "Fault Diagnosis of Transformer Windings Based on Decision Tree and Fully Connected Neural Network," *Energies*, vol. 14, 2021.

[46] Oonsivilai, A., & Marungsri, B, "Application of artificial intelligent technique for partial discharges localization in oil insulating transformer," *WSEAS TRANSACTIONS on SYSTEMS archive*, vol. 7, pp. 920-929, 2008.

[47] Janine M. Freeman et al, "Evaluating Energy Impacts and Costs from PV Component Failures," in *IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC)*, 2018.

[48] ABB., "White Paper: ABB Review: Digital twins and simulations," *Tech. rep. ABB*, 2019.

[49] SunSniffer, "White Paper: Transforming PV O&M with Advanced Analytics," Tech. rep. SunSniffer GmbH & Co. KG, 2020.

[50] Christoph Brosinsky, Dirk Westermann, and Rainer Krebs., "Recent and prospective developments in power system control centers: Adapting the digital twin technology for

application in power system control centers," in *IEEE International Energy Conference (ENERGYCON)* , 2018.

[51] Yong Cui et al. , "Digital Twin for Power System Steady-state Modelling, Simulation, and Analysis," in *IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, 2020.

[52] William Danilczyk, Yan Lindsay Sun, and Haibo He, " Smart Grid Anomaly Detection using a Deep Learning Digital Twin," in *52nd North American Power Symposium (NAPS)*, 2021.

[53] Paolo Pileggi et al., "A Digital Twin for Cyber-Physical Energy Systems," in *7th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, 2019.

[54] Sam Yang, Bjorn Vaagensmith, and Deepika Patra, "Power Grid Contingency Analysis with Machine Learning: A Brief Survey and Prospects," in *Resilience Week (RWS)*, 2020.

[55] J. Y. a. D. F. Mike Zhou, " Digital twin framework and its application to power grid online analysis," *CSEE Journal of Power and Energy Systems*, vol. 5.3, p. 391–398, 2019.

[56] M. Grieves, ""Virtually intelligent product systems: Digital and physical twins," *American Institute of Aeronautics and Astronautic. Complex Systems Engineering: Theory and Practice*, pp. 175-200, 2019.

[57] Adil Rasheed, Omer San, and Trond Kvamsdal, "Digital Twin: Values, Challenges and Enablers From a Modeling Perspective," *IEEE Access*, vol. 8, pp. 21980-22012, 2020.

[58] Jiaju Wu et al., " The Development of Digital Twin Technology Review," in *Chinese Automation Congress (CAC)*, 2020.

[59] Aidan Fuller et al, "Digital Twin: Enabling Technologies, ges and Open Research," *IEEE Access 8*, p. 108952–108971, 2020.

[60] M. Mazhar Rathore et al, "The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities," *IEEE Access*, vol. 9, p. 32030–32052, 2021.

[61] Andrea Castellani, Sebastian Schmitt, and Stefano Squartini, "Real-World Anomaly Detection by Using Digital Twin Systems and Weakly Supervised Learning," *IEEE Transactions on Industrial Informatics*, vol. 17.7, p. 4733–4742, 2021.

[62] A. K. e. al, "Digital-Twin based Services for advanced Monitoring and Control of future power systems," in *IEEE Power Energy Society General Meeting (PESGM)*, 2020.

[63] Hesam Mazaheri et al, "Nonlinear Stability Analysis of DC-DC Power Electronic Systems by Means of Switching Equivalent Models," *IEEE Access*, vol. 9, p. 98412–98422, 2021.

[64] Jason Poon et al, "Fault Prognosis for Power Electronics Systems Using Adaptive Parameter Identification," *IEEE Transactions on Industry Applications*, vol. 53.3, p. 2862–2870, 2017.

[65] Jiawang Xiong et al, "Real-time FPGA-digital twin monitoring and diagnostics for PET applications," in *6th Asia Conference on Power and Electrical Engineering (ACPEE)*, 2021.

[66] Massimiliano De Benedetti et al, "Anomaly detection and predictive maintenance for photovoltaic systems," *Neurocomputing*, vol. 310, p. 59–68, 2018.

[67] Thushara Gunda et al, "A Machine Learning Evaluation of Maintenance Records for Common Failure Modes in PV Inverters," *IEEE Access 8*, p. 211610–211620, 2020.

[68] Jingwei Zhang et al, "Model Parameter Analysis of Cracked Photovoltaic Module under Outdoor Conditions," in *47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020.

[69] Kamel Arafet and Rafael Berlanga, "Digital Twins in Solar Farms: An Approach through Time Series and Deep Learning," *Algorithms*, vol. 14.5, 2021.

[70] Fouzi Harrou et al, "An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine," *Solar Energy*, vol. 179, p. 48–58, 2019.

[71] João Pereira and Margarida Silveira, "Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention," in *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018.

[72] D. N. W. a. S. H. Wihan Booyse, " Deep digital twins for detection, diagnostics and prognostics," *Mechanical Systems and Signal Processing* , p. 106612, 2020.

[73] G. R. Energy, "White Paper: Solar Plant Asset Performance Management (APM)," Tech. rep. General Electric, 2019.

[74] Reuniwatt, "White Paper: Satellite-based Digital Twin: Monitoring and improving remote assets," Reuniwatt, 2020.

[75] Jason Poon et al, "Model-Based Fault Detection and Identification for Switching Power Converters," *IEEE Transactions on Power Electronics*, vol. 32.2, p. 1419–1430, 2017.