



A Comprehensive Review of Supervised Learning Algorithms for the Diagnosis of Photovoltaic Systems, Proposing a New Approach Using an Ensemble Learning Algorithm

Guy M. Toche Tchio^{1,*}, Joseph Kenfack², Djima Kassegne³, Francis-Daniel Menga⁴ and Sanoussi S. Ouro-Djobo^{1,3,*}

- ¹ Regional Center of Excellence for Electricity Management (CERME), University of Lomé, Lome 01 BP 1515, Togo
- ² Laboratory on Small Hydroelectricity and Hybrid Systems, National Advanced School of Engineering of Yaoundé (NASEY), University of Yaoundé 1, Yaoundé P.O. Box 8390, Cameroon; joskenfack@yahoo.fr
- ³ Solar Energy Laboratory, Department of Physics, Faculty of Sciences, University of Lomé, Lome 01 BP 1515, Togo; djimakassegne@gmail.com
- ⁴ National Committee for Development of Technologies (NCDT), Yaounde BP 1457, Cameroon; danielmenga6@gmail.com
- * Correspondence: mtoche6@gmail.com (G.M.T.T.); sourodjobo@univ-lome.tg (S.S.O.-D.)

Abstract: Photovoltaic systems are prone to breaking down due to harsh conditions. To improve the reliability of these systems, diagnostic methods using Machine Learning (ML) have been developed. However, many publications only focus on specific AI models without disclosing the type of learning used. In this article, we propose a supervised learning algorithm that can detect and classify PV system defects. We delve into the world of supervised learning-based machine learning and its application in detecting and classifying defects in photovoltaic (PV) systems. We explore the various types of faults that can occur in a PV system and provide a concise overview of the most commonly used machine learning and supervised learning techniques in diagnosing such systems. Additionally, we introduce a novel classifier known as Extra Trees or Extremely Randomized Trees as a speedy diagnostic approach for PV systems. Although this algorithm has not yet been explored in the realm of fault detection and classification for photovoltaic installations, it is highly recommended due to its remarkable precision, minimal variance, and efficient processing. The purpose of this article is to assist technicians, engineers, and researchers in identifying typical faults that are responsible for PV system failures, as well as creating effective control and supervision techniques that can minimize breakdowns and ensure the longevity of installed systems.

Keywords: diagnosis; faults; photovoltaics; machine learning; supervised learning; extra trees

1. Introduction

In recent years, renewable energy sources have gained popularity, with photovoltaic solar energy ranking as the third most developed technology behind hydroelectricity and wind power. According to the "TrendForce Feb2023" report, photovoltaic solar energy is experiencing remarkable growth, with an estimated world installed capacity of 350.6 GW by 2023 [1]. The annual evolution of the global installed capacity of PV systems is shown in Figure 1. This growth can be attributed to various factors, including reduced production costs, government support policies, reliability, and the desire for localized energy production. However, despite these benefits, photovoltaic installations may face challenges related to aging and environmental constraints that can impact their efficiency and long-term safety. Exposure to difficult environmental conditions can lead to malfunctions and anomalies that result in power losses or even the risk of fire, depending on the severity of the issue [2]. When the surface of a solar panel system is covered with dust for two months, its performance can be reduced by 8.4% compared to a clean system according to studies [3,4].



Citation: Toche Tchio, G.M.; Kenfack, J.; Kassegne, D.; Menga, F.-D.; Ouro-Djobo, S.S. A Comprehensive Review of Supervised Learning Algorithms for the Diagnosis of Photovoltaic Systems, Proposing a New Approach Using an Ensemble Learning Algorithm. *Appl. Sci.* 2024, 14, 2072. https://doi.org/10.3390/ app14052072

Academic Editors: Seongsoo Cho and Bhanu Shrestha

Received: 3 February 2024 Revised: 15 February 2024 Accepted: 16 February 2024 Published: 1 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



Therefore, it is crucial to be aware of any faults, control them to minimize their occurrence, recover the maximum amount of energy produced, and reduce maintenance costs for the PV system.

Figure 1. Annual evolution of the global world capacity of PV installations (TrendForce, 2023).

Several research studies have been conducted to identify the categories of faults and diagnostic techniques for detecting various faults in photovoltaic (PV) systems. Some of these techniques use climate data-independent methods based on the circuit resistance, inductance, capacitance (RLC) and a signal generator to predict faults in PV systems, while others rely on electrical parameters based on current and voltage indicators [3,5,6]. It is noteworthy that these methods are not affected by climate data. In recent years, there has been a renewed interest in the industrial applications of digital methods, such as the use of machine learning for vehicle autonomy on public roads and fault diagnosis using data [7]. In the field of photovoltaics, various machine learning models, such as artificial neural networks (ANN), k nearest neighbors (kNN), the Adaptive Neuro-Fuzzy Inference System (ANFIS), Naïve Bayes (NB), decision trees (RF), and fuzzy logic, have been successfully employed for fault diagnosis [3,8–13]. Several articles have demonstrated the effectiveness of supervised learning algorithms in improving the diagnosis of PV systems with the application of artificial intelligence [14,15]. Compared to traditional techniques that require more computing time and human expertise, Machine Learning (ML) and Deep Learning (DL) supervised learning algorithms are faster and more efficient in providing diagnostic solutions [14,16–18]. For example, Amiri et al. proposed a Deep Learning algorithm that combines convolutional and bidirectional recurrent neural networks to detect faults in a PV system [19]. Additionally, several authors have conducted reviews to highlight the effectiveness of Machine Learning and Deep Learning algorithms in diagnosing PV systems, as they accelerate and improve diagnostic solutions for PV systems [20–27]. This article specifically focuses on supervised machine learning algorithms. To make this happen, the authors propose an ANN model to detect short-circuit faults in a grid-connected PV module. They use Levemberg's algorithm Marquardt and the ANN in MATLAB/Simulink. The authors conclude that the algorithm effectively recognizes short-circuit faults through the trained data [11]. Similarly, Lu et al. propose the random forests algorithm to detect partial shading, short circuit, open circuit, and aging faults on a simulated PV field. The study utilized irradiation, temperature, current, and voltage at maximum power point (MPPT) as inputs for the model. A comparison of the results demonstrated the superiority of the Random Forest (RF) model over the kNN and SVM models [28]. The article proposes the use of a kNN model to detect and classify faults in a PV system, including bypass diode, line-to-line, and open-circuit faults. The results from both simulated and experimental data show an accuracy of 98.7% [3]. Badr et al. demonstrated the effectiveness of the SVM model in identifying various faults, such as line-to-line faults, open circuit, partial shading, and MPPT failure. The algorithm used current and voltage data from the PV

field as inputs, resulting in a fault classification accuracy of 99.4% for detection and 98% for diagnosis [29]. Dhimish et al. developed a Mamdani fuzzy logic controller to detect bypass diode faults and hot spots using voltage drop, percentage of open circuit voltage, and short circuit current as input data. The obtained results demonstrate that the proposed method can accurately identify 13 types of defects, including the hot spot defect, with 96.7% accuracy [30–32]. Additionally, a C4.5 decision tree-based algorithm is proposed to diagnose and detect string, short circuit, open circuit, and line-to-line faults in a grid-connected PV system. The results indicate that the proposed approach correctly classifies defects with an overall accuracy of 99% [33]. In the literature, several articles have widely discussed several aspects of machine learning in fault diagnosis by highlighting the models used, their advantages and disadvantages, the parameters studied and the results obtained [34–37]. For instance, a systematic review of the use of Artificial Intelligence (AI) techniques in photovoltaic (PV) fault diagnosis and identification revealed the significant role of AI in image analysis, anomaly detection, and optimization. The authors concentrate on AI techniques such as Machine Learning, Deep Learning, Machine Vision, and Natural Language Processing (NLP) [37]. An analysis of the reviews employed in this paper indicates that machine learning techniques are extensively utilized in the diagnosis of PV systems. However, the implementation rate of ensemble algorithms remains very low. This article proposes a new classifier Extra Trees (ETC) and its algorithm for the rapid diagnosis of faults in photovoltaic systems. As demonstrated in various publications across different fields, including economics [38], medicine [39,40], hydraulic engineering, and telecommunications [41-43], the Extra Trees algorithm has shown robustness to noise, a significant reduction in bias errors, and lower variance compared to other models such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), and Decision Trees (DT) [39]. Furthermore, this algorithm exhibits a lower computational complexity rate compared to other Machine Learning (ML) classification models, such as DT, Adaptive Boosting (AdaBoost), Naïve Bayes (NB), SVM, RF and KNN [39]. To the best of our knowledge, previous research on fault detection and classification in PV systems has not yet considered the Extra Trees Ensemble (ETC) algorithm. Therefore, in Section 5, we provide a detailed presentation of the Extra Trees model. This paper examines the types of faults in PV systems, their causes and consequences, and presents the most popular supervised learning methods for PV fault diagnosis in recent years. It also summarizes the reviews published on Artificial Intelligence methods for PV fault diagnosis from 2016 to 2023 [37,44–52]. Finally, the Extra Trees algorithm is proposed as a new robust classifier capable of improving the inadequacies of other classifiers in the fault diagnosis of photovoltaic (PV) systems. The purpose of this paper is to provide technicians, engineers, and researchers with information to guide them in identifying the primary faults to check in the event of a PV system failure. Additionally, it aims to assist in selecting the appropriate model for developing control and supervision tools for PV installations to reduce the rate of outages. Section 2 of this paper describes the most common types of faults encountered in PV installations, their origins, and their impact on system performance, Section 3 presents Machine Learning and the most commonly used models for detecting faults in photovoltaic systems, as well as their contributions, Section 4 discusses these methods and introduces a new method for diagnosing faults, Section 5 presents the Extra Trees model and Section 6 contains the conclusion.

2. Faults in a PV System

In photovoltaic systems, various types of faults can cause power loss in some way. To classify the faults in PV systems, some authors have categorized them according to the components involved [5]. Figure 2 represents the description of faults likely to occur in a photovoltaic system.



Figure 2. Classification of fault types in a PV system.

2.1. Photovoltaic Generator Faults

In a photovoltaic system, a fault refers to an atypical behavior that signals a potential loss of power or complete system unavailability. Given the challenging environmental conditions that photovoltaic systems operate in, defects can arise across various components, each with a unique set of issues. These faults may manifest themselves in the photovoltaic array and/or in the inverter, which can partially identify them [53]. PV generators can experience different types of faults, which are classified based on the area they affect. These categories comprise electrical, physical, and environmental faults [54]. However, faults can occur in photovoltaic systems, with the most common being on the solar panel side. These include shading, mismatch, potential-induced degradation, hotspot, open circuit, short circuit, line-to-line, line-to-ground, arc, bypass, and anti-reverse diode faults. Also, there are different types of faults that can occur in an inverter, which include open-circuit faults, short-circuit faults, insulation faults and so on [36,55]. The following section provides a detailed explanation of the most common faults provides a detailed explanation of the most common faults provides a detailed explanation of the diagnostic process.

Based on the analysis of Figure 2, the following subsections describe some of the most common faults that can affect a PV installation.

2.1.1. Ground Fault

Ground fault (F_1) is an accidental short circuit between one or more current conductors and the earth. It is the most common type of fault that occurs due to cable insulation failure. This fault poses a serious risk as it can produce current arcs at the points of failure, leading to electric shocks. Furthermore, it causes an increase in current in the affected conductors, resulting in imbalances and changes in the architecture of the PV array [56].

2.1.2. Short-Circuit Fault (SCF)

A short-circuit fault occurs when two points in a circuit of different potentials accidentally connect [57]. This fault can happen within the same module string (intra-string fault F_2) or between two modules of different strings (inter-string fault F_3). Poor wiring between the generator PV and the inverter, animal damage to cables, and water infiltration into the PV modules are the causes of this fault [58]. Short-circuited modules result in a drop in network voltage while the current significantly increases. Generally, a short-circuit fault circuit causes a line-to-line fault [59].

2.1.3. Line to Line Fault

According to Pillai et al. [36], a line-to-line fault happens when there is an unintended short circuit between two points of a PV array with different potentials. This type of fault can occur between modules of the same string or between modules of adjacent strings. It can also occur between conductors of the same circuit with different potentials, without involving any earthing point. Furthermore, when this fault occurs between two modules of the same order from different strings, it is sometimes referred to as a bridging fault [60]. The outcome of this fault is a decrease in the open circuit voltage, while the short circuit current may remain unchanged. This voltage reduction results in a modification of the current-voltage characteristics of the photovoltaic field. Please see Figure 3 for a summary of the most common faults in a PV system.



Figure 3. (a) Partial shading of a PV module (b) Total shading of a PV module.

2.1.4. Open-Circuit Fault (OCF)

An open-circuit fault (F_4) occurs when a cable inside a module or a PV module string accidentally disconnects. This fault affects the total resistance of the PV generator and causes a significant increase in the short circuit current [61]. However, an open-circuit fault is more damaging than a short-circuit fault due to the increased current flow. The breakage of connection wires between cells or PV modules, faulty diodes, and the deterioration of connection cables usually cause this fault [59]. An open-circuit fault is a result of the line-to-line fault, which itself is caused by the short-circuit fault [22].

2.1.5. Arc Fault

An arc fault is a type of fault that occurs when an electrical current passes accidentally through air or another dielectric material [55]. Detecting arc faults is a complex process because they occur intermittently. Arc faults can happen within a single conductor (series arc fault F_5) or between two parallel conductors (parallel arc fault F_6). Additionally, faults can occur due to the breakage of insulation cables, which can cause significant noise in the output voltage and current of the PV network [56].

2.1.6. Mismatch/Shading Defects

A mismatch fault occurs when a group of photovoltaic cells has different electrical characteristics [62]. This type of fault can be permanent, like an open-circuit fault, or temporary, like partial shading. Partial shading is a specific type of mismatch fault and is one of the main causes of failures in a PV system. The shading phenomenon can be classified as uniform or non-uniform [62]. The source of uniform shading can be adjacent buildings, passing clouds, trees, other signs, bird droppings, dirt and so on. Non-uniform or partial shading defects occur when some cells or modules receive direct irradiation and temperature in a non-uniform manner. On the other hand, uniform or total shading occurs when all cells or modules receive uniform but reduced exposure, resulting in a constant reduction in the output current and voltage of individual cells in a string. It is important to note that technical abbreviations should be explained the first time they appear [63]. Figure 3 shows a partially shaded and fully shaded module.

The setting of the sun causes shading of the photovoltaic (PV) module, which reduces its power output. It is important to note that although shading has a negligible impact on the PV module's overall performance, it should still be avoided. Shaded cells can become reverse polarized, consuming energy instead of producing it, leading to a drop in power and the hotspot phenomenon [64]. The hotspot phenomenon can accelerate the aging process of the PV system and may even lead to an open-circuit fault or fire risks [65]. Figure 4 provides an illustration of the various faults described above.



Figure 4. Illustration of ground faults F_1 , short-circuit faults F_2 , F_3 , open-circuit faults F_4 and arcing faults (F_5 , F_6 in a PV array).

In the case of a short circuit, the output voltage drops significantly while the output current slightly increases. Short-circuit faults can affect cells, modules, and bypass diodes [66]. Bypass diodes are protection devices against shading and are connected in parallel to each group of cells as show in Figure 5. However, these diodes can be damaged during factory electrical discharge and high reverse voltage due to any fault [67]. If the bypass diode is faulty, there will be a sudden drop in power due to the absence of the voltage chain. The fault may be caused by non-functioning diodes, diodes reversed during assembly, poor diode connection, disconnection, or corrosion of the junction boxes. A bypass diode fault can cause damage such as hot spots, electric arcing, and the risk of fire if the diode is in an open circuit [68].



Figure 5. Illustration of a group of cells with bypass diode disconnected.

PV modules can degrade in several ways, including discoloration of the encapsulant due to exposure to UV rays, which causes the PV cells to turn brown or yellow. Another form of degradation is delamination or the separation of different layers of the PV module. There are also two specific types of degradation to be aware of. Potential Induced Degradation (PID) occurs when there is a potential difference between the metal frame of the PV module and the solar cell, which can significantly degrade the electrical characteristics of the PV cell. Light-induced degradation (LID) is a loss of performance that occurs when the boron–oxygen effect and the boron–iron effect are activated after the PV modules have been exposed to sunlight [5]. In addition to the breakdowns observed at the PV generator level, the photovoltaic inverter is also a vulnerable component with unreliable performance [68]. Therefore, it is necessary to have knowledge of the common faults associated with this component.

2.2. PV Inverter Faults (PVI)

In photovoltaic applications, one of the biggest challenges is ensuring that power electronics are reliable in order to optimize energy production. The inverter serves as the interface between the photovoltaic generator and the network and/or load. Its main function is to convert the continuous energy produced by the photovoltaic modules into alternating energy that is identical to the network. This allows the inverter to access electrical information from the generator and the electrical network, making it an intermediary between the two. Additionally, the inverter is equipped with a high level of data granularity, which enables it to detect electrical anomalies in real-time and alert the user through an audible signal or a message. However, despite its advantages, the inverter is vulnerable and subject to faults. During its operation, it is exposed to overvoltage and overcurrent constraints due to transient operating conditions, mechanical turbulence, temperatures, and humidity [68]. The IGBT (insulated gate bipolar transistor) power switch, being the main energy transfer component, is the most likely source of failure in the photovoltaic inverter [69]. The most common faults that can occur during the inverter's operation are open-circuit, insulation faults and short-circuit faults [70].

2.2.1. Short-Circuit Fault

A fault can occur due to breakage of the connection wire, deterioration of the gate circuit, or overcurrent. However, a short circuit happens very quickly, making it difficult to detect. Shortly after appearing, it transforms into an immediate open-circuit fault [71].

Short-circuit faults automatically shut down the system, making them more dangerous than open-circuit faults.

2.2.2. Open-Circuit Fault

An open-circuit fault can occur in a photovoltaic system due to a disconnection of the jumper wire, overheating, or a device driver fault, resulting in a broken connection. Unlike a short-circuit fault, an open-circuit fault may not immediately affect the inverter, but if left unaddressed, it can lead to serious accidents with other components [71]. This is because an open-circuit fault distorts the output current of the inverter, causing an increase in the total harmonic ratio, which does not meet the grid connection requirements. Table 1 provides a list of main faults that can occur in a photovoltaic system, and Figure 6 illustrates an overview of the open-circuit fault that can occur in the IGBT transistor of a PV inverter.

Table 1. Summarizes the various faults, their causes, and the resulting consequences.

Type of Defects			Causes	Effects	Consequences	
	Short circuit		Manufacturing defect	Low impedance, blocked path between internal power rails	Reduction of power produced	
Internal	Cell microcracks		Manufacturing defect	Difference in module	Unable to deliver power to the load	
	Broken modules		Shock during transport			
	Degraded modules		Aging	Drop in power delivered	Low production	
	Bypass diode		Manufacturing defect, wiring defect	Can't drive	Unable to prevent the appearance of hot spots, electric arc, fire risk	
	Open circuit		Manufacturing defect, wiring defect	Lack of access path for the power produced	No power produced	
External	Mismatch fault	Temporary	Temporary shading	Cloud	Drop in production, risk of fires	
		Permanent	Equipment damage	Blackout	No production	
	Shading		Passage of clouds, weather conditions	Uneven distribution of irradiation on the surface of the modules	Drop in power produced	
5			Natural disaster	. Module reverse bias	Hot spot/fire hazard	
		Permanent	Partial shading		1 . ,	
	Short circuit		Bad wiring between inverter and PV field, chewing of cables by animals, water infiltration into modules	Drop in network voltage and increase in current	Drop in production	
	Open circuit		Accidental breakage of connecting cables	Drastic drop in short-circuit current	Drop in production	
PV field	Line to line		Faulty connection link between the different rail circuits	Power loss	Reduction of open circuit voltage, modification of characteristic IV of the PV field	
	Arc fault		Accidental passage of current in a dielectric	Strong noise in currents and voltages	Fire hazard	
	Line to ground		Ground wiring fault, corrosion	Drop in network voltage and increase in current	Risk of electrocution, variable voltage	
PV inverter	Inverter open circuit		Absence of gate control, connection wire breakage due to high short-circuit current, external disconnection due to vibrations	Deterioration of phase current and torque	External radiation	
	Inverter short circuit		High gate voltage, delamination and cracking in the solder layer, static locking and high temperature	Excessive leakage current, affected phase current close to zero	Temperature variation	
	Insulation fault		Humidity, high heat, poor connection in the solar panel junction box, aging of solar panels		No power	
Grid	Grid anomalies		Electrical overload, deterioration of conductive insulators	Network disruption, voltage dips and peaks, harmonics	Interruption of current flow, short circuit	



Figure 6. Representation of an open-circuit fault on an inverter arm.

3. Machine Learning and Fault Diagnosis

Artificial intelligence (AI) has become increasingly popular over the recent years, and is now used in various aspects of human life [72]. AI refers to a set of concepts and algorithms that aim to replicate the reasoning of the human brain [72]. The main applications of AI are automatic learning (ML) and deep learning (DL). Machine learning is a type of AI that enables software applications to predict results of a model with greater accuracy, without the need for explicit programming. However, machine learning, which is the dominant paradigm of AI applications, relies on historical data as input to predict new output values, and can be categorized into three subgroups [72], each using a specific algorithm as shown in Figure 7.



Figure 7. Illustration of the Machine Learning hierarchy.

3.1. Supervised Learning

Supervised learning algorithm is a type of algorithm designed to learn how to classify data based on certain input parameters. The algorithm is trained on a set of data accompanied by the desired outputs or labels. This helps the algorithm to identify patterns and correlations in the data, which can be used to predict or classify new data. The goal of supervised learning is to accurately predict the correct output or label for future observations based on what it has learned from the training data [5]. Supervised learning problems (see Figure 8) can be described mathematically in the following way: we have a set of *n* samples $\{\vec{x_i}\}_{i \in \{1, 2, \dots, n\}}$ in a universe Ω and their respective labels $\{y_i\}_{i \in \{1, 2, \dots, n\}}$ in a universe \mathcal{P} , we define a function $\varphi: \Omega \to \mathcal{P}$ (which is fixed and unknown) that takes the sample as input and produces the labels $y_i = \varphi(\vec{x_i}) + \omega_i$ as output, where ω_i is some random noise involved which can affect the accuracy of the labels generated from the samples. Thus, the data used makes it possible to determine a function $f: \Omega \to \mathcal{P}$ such that

for any pair $(\vec{x}, y) \in \Omega \times \mathcal{P}, f(\vec{x}) \approx y$. The universe on which these data are defined is typically a set of possible inputs for the algorithm $\Omega = \mathbb{R}^N$. As a result, the whole $\left\{ \begin{pmatrix} \vec{x}_i, y_i \end{pmatrix} \right\}_{i \in \{1, 2, ..., n\}}$ makes up the training data set. Figure 8 provides a demonstration of how a learning algorithm operates in a supervised learning environment, which differs from unsupervised learning.



Figure 8. Supervised learning principle.

3.2. Unsupervised Learning

This learning technique does not require labeled data to assign classes. Unlike supervised learning, it identifies the structure of the dataset itself. Various algorithms are employed to enable the machine to scan through the data sets and search for any significant connections [5]. The probability of unsupervised learning enabling the discovery of undetectable models or classifications by humans is high $\{\vec{x}_i\}_{i \in \{1,2,..,n\}}$ in a universe Ω , we learn a function Ω which verifies certain properties as shown in Figure 9, explaining the operation of an unsupervised learning algorithm.



Figure 9. General principle of unsupervised learning.

3.3. Semi-Supervised Learning

This type of learning is a combination of supervised and unsupervised learning. The algorithm learns from a small amount of labeled data to become familiar with it and then explores the data on its own to expand its reasoning on the dataset. However, this type of learning is not commonly used in machine learning applications [5]. It takes into account both labeled and unlabeled data to reduce the dependence on data and improve the performance of the PV system fault detection [73].

In this article, we review the most common supervised learning algorithms used in diagnosing faults in solar photovoltaic installations. Machine learning algorithms are efficient and precise in solving complex and non-linear problems, unlike other methods [23]. In the literature, several ML techniques are used for fault diagnosis in PV systems [72,74]. Figure 10 shows the structure of Machine Learning (ML) techniques in PV system detection and diagnosis.



Figure 10. Illustration of the structure of Machine Learning in PV fault diagnosis.

This particular approach is useful in reducing faults and enhancing the performance of a PV system. However, the accuracy of defect detection may vary depending on the principle and model architecture of the machine learning (ML) used [75]. There are various types of supervised learning models available, with the most common being k-nearest neighbors (KNN), decision trees (DT), artificial neural networks (ANN), fuzzy logic (FL), random forest (RF), support vector machines (SVM), and so on [31,34,57,76–78]. Depending on the expected result, an author can choose to use a single model or combine two or three models to assess the relevance of their work. As a result, the machine learning models discussed in this article are categorized into simple models and hybrid models, among other things.

3.4. Simple Fault Diagnosis Models

3.4.1. Artificial Neural Networks (ANN)

An artificial neural network is a type of supervised learning model that is based on a simplified model of the human brain. These networks consist of highly connected elementary neurons that operate in parallel. Each neuron can have multiple inputs and calculate a single output based on the received information. A hierarchical network structure is always a network, consisting of an input layer, several hidden layers, and an output layer with numerous neurons in each layer, as depicted in Figure 11.



Input layer Hidden layer 1 Hidden layer N Output layer

Figure 11. Structure of an artificial neural network.

Artificial neural networks (ANNs) have two key properties that make them useful in the field of production systems diagnostics. Firstly, they are capable of approximating non-linear functions, and secondly, they are well-suited for pattern recognition tasks, such as PV system diagnosis [79]. ANNs do not rely on mathematical models and are, therefore, applicable to complex systems, making them highly advantageous. However, the learning process, network architecture, and explanation of ANNs have certain limitations [80]. In the literature, ANNs are widely used to characterize PV system failures [81]. For instance, an ANN model has been proposed to identify and detect several faults in a photovoltaic installation, including short-circuit faults, short-circuited bypass diodes, inverted bypass diodes, disconnected bypasses, module open circuits, and connection resistance between modules. The results obtained from the experimental data showed good accuracy in detecting and classifying different defects [79,82]. Additionally, to improve the performance of a PV module, the IV characteristics of a solar cell can be modeled using ANN [83]. A comparison between the electrical equivalent model and the thin film technology model showed good accuracy of the ANN with the crystal technology model. Similarly, a new diagnostic model has been developed that uses current and voltage data at the maximum power point to detect short-circuit and disconnected string faults with the aid of ANN [84]. The comparison between the simulated model values and the actual values showed an accuracy rate of approximately 98.6%. Furthermore, ANNs made it possible to detect the open switch fault of the inverter based on the characteristics of the phase current [85]. This technique was found to provide better control than using basic DTC in terms of sound. A study was carried out using irradiation and temperature data on two PV fields of different capacities (2.2 kWp and 4.16 kWp) to detect partial shading and module chain disconnection faults with the ANN algorithm [10]. The model was found capable of detecting defects with an accuracy that varies between 96.7 and 98.1%, respectively, without and with shading. An accuracy of 97.6% with the 2.2 kWp field and 97% with the 4.16 kWp field was obtained. This decrease is due to the variation in the nature and capacity of the PV field in terms of the amount of data and the number of faults detected. Moreover, the use of ANNs significantly reduced the risks associated with manual repairs and the time necessary for diagnosis when detecting reverse diode faults and partial shading of a photovoltaic module array from simulated data [86]. A technique based on fast Fourier transform and ANN was able to detect open-circuit and short-circuit faults in the 5-level cascaded inverter using the inverter's output voltage [87,88]. Dhimish et al. showed that ANNs can detect partial shading, short circuits, ground faults, and degradation faults, with an efficiency of approximately 99% for correctly classified defects [89]. An ANN-based approach is proposed to detect and classify series, parallel, short-circuit, and open-circuit resistance faults under different irradiation and temperature conditions in a PV system [90]. The results of the simulation and experimental study show a good correlation with a classification error rate of 2.7%. Dhimish et al. also found that ANNs are useful in the detection of bypass diode faults in short circuits and open circuits, with the model being 96.4% and 92.6% accurate in detecting short-circuit and open-circuit bypass diode faults, respectively [89]. In a review of the application of ANNs in the diagnosis of PV systems, Li et al. emphasized the importance of ANNs in the field of solar photovoltaics [21]. Finally, a new approach to diagnosing short-circuits faults and disconnected strings based on ANNs in a PV system was found to have an accuracy of 98.6% [86].

3.4.2. Support Vector Machine (SVM)

SVM is a supervised learning algorithm based on statistical learning theory and the structural principle of risk minimization and was first presented in 1995 by Cortes and Vladimir Vapnik [90,91]. This model is used for classification and regression problems such as medical diagnosis, communication, biology, engineering, etc. The main goal of this model is to search for a hyper-plane in a high-dimensional space that best separates different classes of data with a large margin [92]. This model performs well when the separation margin between classes is observable with large-dimensional spaces. Nevertheless, it shows that the model detected the line-to-line fault with about 95% accuracy. However, it is not suitable for large amounts of data [93]. The structure of the SVM model is given in Figure 12.





The hyperplanes $H_1et H_2$ pass through the closest samples and are parallel to the hyperplane H, respectively. Points located on hyperplanes $H_1et H_2$ are the samples that have all the information used to design the SVM classifier. In the context of fault diagnosis, this model can make decisions on small quantities of data because a large amount of implicit data classification knowledge can be extracted [92]. Several authors have implemented the SVM algorithm successfully in many fields. For example, Kuraku et al. used the SVM algorithm for open-circuit fault detection of the IGBT switch of an H-bridge multilevel inverter [94]. Natarajan et al. proposed the SVM model to classify module crack defects and hot spot defects due to shading and dust accumulation. Intentionally created defects were detected and classified in real time with 97% accuracy [93].

3.4.3. k-Nearest Neighbor Algorithm (kNN)

The kNN model is a method for classifying a new object by examining its distance from the nearest neighbor of training samples in feature space. It belongs to the family of supervised learning algorithms. This algorithm does not require learning strictly speaking but just focuses on storing the training dataset, hence the term lazy algorithm. Indeed, to predict the class of new input data, it will look for their k closest neighbors using the Euclidean or Manhattan distance or others in order to choose the class of the majority neighbors. Two voting schemes exist to determine a label namely, the majority vote assigned to the class that appears most commonly in its k-nearest neighbors and the weighted sum voting where each vote is weighted on the basis that the closer nearest neighbors must count more than distant neighbors [83]. As a result, the kNN classification model depends on three main determinants which are: the training samples, the distance between the training samples (labeled data) and the test samples (unlabeled data) and the value of k [95]. Figure 13 shows the working principle of the kNN model.



Figure 13. Synoptic overview of the KNN algorithm.

The kNN algorithm is straightforward and does not require training before making predictions. Its accuracy is not affected by adjusting several parameters or making additional assumptions [96]. It is also a versatile model that can be used to classify or regress and is suitable for sparse data [96]. Furthermore, this algorithm is threshold value independent and can detect small to medium data samples with high accuracy and speed, all while being relatively low cost compared to other existing methods such as DT and SVM [78,97]. However, as the number of observations and independent variables increases, it can become slower. The effectiveness of the algorithm is limited by the fact that it only has access to one class of healthy operational data [98]. However, this problem can be avoided by only using data that are in optimal condition [99]. Naik et al. used the KNN model to detect faults in the transmission line of a power system [79]. This model made it possible for all the cases tested to have a detection and classification accuracy of 100% [79]. Furthermore, a similar study made it possible to detect and classify all the possible faults of a six-phase transmission system [100]. Two new KNN-based diagnostic methods for classifying eight types of faults of a power transmission system yielded a success rate of 98% of detected faults [101]. Furthermore, Madeti et al. proposed the KNN model to detect and classify bypass diode, line-to-line and open-circuit faults in a PV system. The results generated from the simulated model and the experimental data present an accuracy of 98.7% [3]. This same KNN model was proposed from the experimental data to detect open circuit, line-to-line and partial shading faults to improve the results with an accuracy of 99.84% [102]. Also, a KNN model proposed to detect the open-circuit fault of an inverter using the output current and voltage of the inverter achieved good results with an accuracy of about 99.77% [103]. Similarly, Livera et al. showed that among the machine learning models used for the detection of open circuit, short circuit, bypass diode shorted and inverter failure faults, only the KNN model achieved an accuracy of 100% against the SVM, DT and FL models [8].

3.4.4. Fuzzy Logic (FL)

Fuzzy logic is a branch of mathematics that allows a computer to model the real world in the same way that people do. It makes it possible to improve expert systems (appropriate where humans can linguistically describe the solution to the problem) because typically human knowledge is imprecise and vague. It was set up by Lofti Zadeh in 1965. The principle of a fuzzy system is to be able to calculate output parameters by providing the system with a set of rules formulated in natural language and is composed of three parts as illustrated in Figure 14.



Figure 14. Synoptic overview of a fuzzy logic system.

The fuzzy logic model has the advantages of being simple, does not need a large amount of training data, and provides easy ways to deal with antagonisms in the well-defined knowledge base [30]. With fuzzy logic, rules can be generalized to cover a larger number of situations. However, in fault diagnosis, fuzzy systems are useful due to the fact that diagnosis often needs knowledge-based processing [30]. In practice, it is almost impossible to obtain adequate representations of the complex and highly nonlinear behavior of faulty systems using quantitative models. Several authors use the fuzzy logic controller

also to resolve the uncertainties and inaccuracies associated with the description of the system [103]. In the work of Varga et al. The fuzzy logic algorithm is used for the diagnosis of renewable energy systems. A comparison with the PSD (power spectral density) method showed better accuracy compared to PSD [104]. Tamissa et al. examined the possibility of open circuit diagnosis in a three-phase inverter using the fuzzy logic controller. The simulation results show a better performance of the algorithm in the classification of desired defects [105]. Furthermore, Mehta et al. used fuzzy logic control to detect an open-circuit fault in a five-level cascaded inverter. The total harmonic distortion (THD) of the output voltage and the average output voltage are used as diagnostic variables. After simulation, the results obtained show the fault tolerance of the proposed method [106]. Mamdani-type fuzzy logic has been proposed to detect hot spot faults in PV systems. Three parameters are used as diagnostic variables (voltage drop, open circuit voltage, short circuit current). After simulation, the model shows an accuracy of 96.7% in detecting hot spots [30]. A fuzzy logic algorithm is used by Zaki et al. to detect and identify eight types of faults in a photovoltaic installation. It uses current ratios, maximum operating point voltage and open circuit voltage ratio as input parameters. All experimental defects were identified and classified with 99% accuracy [107]. Furthermore, it is also shown that fuzzy logic can be used to detect and classify faults in a solar photovoltaic system. The authors demonstrate that the proposed technique can accurately diagnose various types of faults with 98% [107].

3.5. Hybrid Models for Diagnosing PV Systems

At this stage, most of the models used provide accurate results in fault diagnosis, each with its own strengths and weaknesses. However, these models, although effective, do not realize the full potential of ML on all data samples. One way to improve the performance of ML in task execution is to combine the advantages of simple algorithms in order to solve problems that they are incapable of solving alone. As a result, some authors have proposed improved and/or hybrid versions of simple algorithms to carry out diagnostic work in photovoltaic systems. The ANFIS model (Adaptive Neuro Fuzzy Inference System) is an adaptive neuro fuzzy inference system that combines the advantages of both fuzzy logic and artificial neural networks. The detection and classification of system faults is proposed by deploying two subsets of the ANFIS model, such as the ANFIS partition network (ANFIS GP) and ANFIS subtractive clustering (ANFIS SC). The results of the statistical analysis based on RMSE (Root Means Square Error) show that ANFIS SC is better than ANFIS GP [108]. De campos Souza et al. proposed the ANFIS model for modeling and the identification of the PV system to detect short-circuit and open-circuit faults in the single-phase photovoltaic inverter. The model detected the simulated defects with an accuracy rate of 100% with a fast execution time [109]. A review based on the applications of the fuzzy neuro network algorithm and its variants is proposed in order to show the effectiveness of the algorithm in the construction of systems [110]. Likewise, a new detection method based on the combination of the three algorithms SVM, KNN, and NB made it possible to extract the electrical characteristics in order to analyze and detect the line-line fault on the basis of the current-voltage curve of the PV system [97]. A grid-connected PV fault detection and diagnosis technique based on the combination of three algorithms KNN, DT and SVM showed a better performance with an accuracy rate of 99.96% [74].

To improve the performance of a PV system, a hybrid algorithm based on principal component analysis and support vector machine is proposed to detect bypass diode and series resistance faults. The results obtained after the detection of all types of defects show an accuracy of 99.96% and 99.93%, respectively [96,98]. Also, a fuzzy k-nearest neighbors (FKNN) model was used to detect short-circuit, open circuit and irradiation faults in a spacecraft powered by a solar system. The results obtained are significantly better compared to the KNN model, i.e., an accuracy of 99.4% compared to 91.8% for KNN [99]. In the literature, several reviews on the diagnostic methods for PV systems have been carried out in recent years [7,111,112]. In these reviews, the presence of hybrid models such as

ANFIS, FKNN and so on is remarkable, as shown in Table 2. Another type of hybridization involves combining weak classifiers to obtain a strong classifier for better prediction. This algorithm is based on the principles of Bagging and Boosting, using decision trees as a basic algorithm. It proceeds by voting or averaging the individual performances of each weak classifier to obtain an optimal performance. Therefore, it is called an ensemble learning algorithm.

3.6. Ensemble Learning Algorithms

3.6.1. Decision Algorithm Tree (DT)

The decision tree algorithm is a hierarchical representation of the data structure in the form of decision sequences in order to predict a result or a class [78,113,114]. Indeed, it allows the prediction of a target variable from other so-called explanatory variables (model). The principle of the DT algorithm consists of determining the best possible characteristic for a set of data; separating the data into subsets containing the values of the best characteristic [115]. It also allows you to recursively generate new decision trees using the subset of data created and make the decision when you can no longer classify the data [111]. In other words, the DT model has decision nodes that have several branches and are used to make decisions and leaf nodes that represent the result or class of these decisions [75]. Each node tests a condition on a variable and each of its child nodes corresponds to a possible outcome for that condition. The label or class of an observation is predicted by following the test results from root to leaf as shown in Figure 15. This model is simple to visualize and understand, requires little data preparation, uses categorical and numerical data and handles multi-class problems with robustness and resistance to noisy data [74]. However, it is a lazy algorithm with generally low modeling capacity and instability.



Figure 15. Structure of a DT classification model with six labeled classes (N, S, O, L, A, R).

The start node is called the root of the tree and the path from the root to the terminal node represents the classification rule [113]. Classification and decision tree regression trees are non-parametric algorithms whose regression output is numerical, while that of classification is a function of the input data used for training and testing [110]. This model is proposed to predict the output powers of photovoltaic and wind systems [113]. The simulation results demonstrate that the model is capable of correctly predicting dam classes. An approach using the decision tree algorithm made it possible to detect short-circuit faults and string faults in a photovoltaic installation connected to the grid at an accuracy rate of 99% [33].

3.6.2. Random Forest (RF) Algorithm

The RF algorithm is a classification and regression algorithm whose learning is based on the decision tree. This algorithm was first proposed in 2001 by Leo Breiman [116,117]. Furthermore, the RF algorithm is considered a black box that combines several randomly constructed decision trees in parallel [115]. This model is easy to evaluate, robust to outliers, capable of handling efficient prediction on a large dataset without removing variables, and does not need cross-validation. In addition, the algorithm is presented as a reference in Machine Learning competitions and is more efficient than decision trees [112]. The disadvantages of this algorithm lie in the fact that it is not easy to interpret (explain how the forecast is calculated) since it is obtained from a large number of trees which are very deep, it is also difficult to improve because it is considered a black box and trains slowly [118]. Random forests are tree ensemble methods, they aggregate the predictors of several trees, each of which is trained separately [28]. It offers RF to detect line-to-line faults, degradation, open circuits and partial shading. Random forests, thanks to the bootstrap resampling technique (sampling technique with replacement which gives the selection procedure the particularity of being random) can repeatedly extract n different samples from a set of data. This technique allows the creation of a new training sample for training the decision tree in order to generate n decision tree classifiers [28]. Therefore, Liu et al., proposed the adaptive period electrical partition (AEPP) and random drill algorithm to detect the open-circuit fault in a multilevel NPC inverter. A comparison of the results obtained with traditional methods shows a good accuracy of 99.21% and 99.38%, respectively [119]. The construction of random forests is conducted in four steps as shown in Figure 16.



Figure 16. Steps for generating random forests.

Several fault diagnosis methods based on the random forest algorithm are proposed in the literature [5,115,120]. A proposed RF model-based approach to detect the open-circuit fault of an inverter using the output current and voltage of the inverter gave an accuracy of about 96% [103].

3.6.3. Adaptive Boosting Algorithm (AdaBoost)

The AdaBoost algorithm is a type of machine learning algorithm that combines weak classifiers, such as one-split decision trees, to create a strong classifier. The strong classifier is obtained by adjusting the weights of observations in the dataset based on previous prediction errors. In other words, the algorithm gives more weight to observations that were misclassified in previous iterations, so that they are more likely to be correctly classified in the future [121]. The AdaBoost algorithm was invented by Yoav Freund and Robert

Schapire in 1996 [122]. AdaBoost was the first algorithm to show that boosting ideas can be implemented efficiently and simply. Today, it remains the most widely used algorithm in various fields of application [122]. Lodhi et al. have successfully implemented the AdaBoost model for the detection and classification of short circuits, open circuits, and degradation faults in a photovoltaic system. A comparative study between the proposed model and other models such as KNN, SVM, and RF shows that the AdaBoost model has a superior accuracy rate of 97.84%, compared to 91.29%, 94.34%, and 96.76%, respectively, for the KNN, SVM, and RF models [123]. Similarly, the AdaBoost model has been able to correctly detect and classify faults in a simulated 250 kW PV array with 95% accuracy [124]. The implementation process of the AdaBoost model is divided into four steps, which are: the collection of data, the generation of a strong classifier from weak classifiers using the training, test, or validation data of the classifier, and the application of the classifier for engineering problems [125]. The decision tree, or classification and regression tree (CART), is used to generate the weak classifiers, which produces a strong classifier through majority voting for classification or through arithmetic mean for regression [125]. Figure 17 shows the flowchart of the AdaBoost model implementation procedure.



Figure 17. Flowchart for implementation of AdaBoost approach [121].

In this paper, most of the articles used are written in English and extracted from different databases such as MDPI, IEEE Xplore, ScienceDirect, Springer, Wiley, and Hindawi. Thus, some reviews published in recent years on diagnostic methods using an artificial intelligence approach are represented in Table 2 [126–128].

Authors	Year	Type of Techniques	ML Models	PV System Component	Contribution	Remarks
Youssef et al. [127]	2016	AI	ANN, FL, ANFIS, GA, GA-fuzzy, NN-fuzzy	PV field	This text demonstrates the importance of AI in modeling, sizing, forecasting, and diagnosing fault in PV systems.	The text compares the accuracy of different AI techniques with traditional methods in each application. However, it does not specify the monitoring parameters of each method.
Daliento et al. [59]	2016	Electrical and AI	ANN, SVM, ANFIS, RBN	PV field	This text presents a review of the various methods used to monitor PV systems.	The text is already well-written and adheres to the desired characteristics. Therefore, no changes have been made to the original text.
Madeti et al. [57]	2017	Conventional and AI		PV field	A Review of Detection Methods for Grid-Connected Photovoltaic Systems	The text already meets the desired characteristics. No changes were made.
Mellit et al. [33]	2018	Electrical and ML	ANN, FL, GA, HS	PV field	A comprehensive review on detection methods for grid-connected PV systems	The author's work focuses on using electrical methods to diagnose faults.
Mellit et al. [82]	2016	Electrical and ML	ANN, FL, MSD	PV field	This text discusses PV fault information and diagnosis methods,	However, the main scope of the work is based on identifying defects.
Pillai et al. [36].	2018	IRT, ML, Others	ANN, LAPART,	PV field	including a review of almost all PV faults and advanced detection techniques.	However, his discussion is focused on the flaws
Abdulmawjood et al. [60]	2018	Visual, Thermal and ML Methods	SVM, k-Means, HMM, BN, ANN, GMM (Gausian mixture model)	PV field	It also covers different types of faults and detection techniques in PV fields,	The discussion is focused on electrical faults, but the parameter used for fault detection is not specified for each method.
Appiah [47]	2019	IRF, ML, DL,	ANN, LAPART, KELM, ANFIS	PV field	as well as reviews on types of defects, their origins, and traditional and intelligent detection methods.	The text is clear and concise, but lacks complexity, precision, and input data.
Li et al. [21]	2020	M.L.	ANN	PV field	The text identifies work specifically applied to ANN and hybrid methods with ANN to analyze defects, type and amount of data used, model configuration, and effectiveness	A comparison of ANNs with other ML models shows the superiority of ANNs. However, a comparison between ANN models is not mentioned to identify the most efficient model
Ghaffarzadeh et al. [125]	2019	Electric, ML	ANN, SVM, DT, FL, Kalman filter	PV field	It explains the types of defects over a broad spectrum	It focused on current fault on the DC and AC side of the PV system.
Venkatesh et al. [126]	2020	Visual method, IRT, EL, ML	ANN, SVM, NC-NFC, CNN, DT, KNN, FL	PV field	Lists four types of visual defects and detection methods	Failure to take into account non-visual defects, no precision
Kurukuru et al. [46]	2021	ML, DL	ANN, ANFIS, PSO, FL, GA, ABC, CNN, SVM, KNN, LTSM	PV field	Review to show the impact of AI on the PV value chain.	The precision of each technique is not made
Zenebe et al. [48]	2021	ML, DL	SVM, DA, BN, ANN, KNN, RF, DT, CNN	PV field, Inverter	Presents a review on ML-based detection methods to show that ANN and MLP are the most promising models in terms of simplicity and accuracy	However, its main field of action was based on defects and detection methods
Mansouri et al. [25]	2021	D.L.	DBN, CNN, RFCN, R-CNN	PV field	Review of Deep Learning applications in solar cell fault detection	The article examines defects related to cell discoloration, cracking, and delamination in PV systems.
Rodrigues et al. [129]	2017	M.L.	DT, RF, FL, ANN, GA, Bayesian, KNN, GA-ANN, ANFIS, RVM, k-Means	PV field	Review of the articles that deal with the prognosis and diagnosis of defects and the number of themes covered in the study	It reviews the types of studies conducted, the types of faults studied, the input parameters used, and the types of PV systems analyzed. However, it does not provide an evaluation of the effectiveness of each method based on these parameters.
Abubakar et al. [5]	2021	AI, ML	ANN, SVM, LAPART, RBF-ELM, FL, GBSSL, ANFIS, DT	PV field	Characteristics of AI methods, their speed and effectiveness in detecting defects with minimal errors	The authors did not justify the interest in including articles from the last 15 years, nor did they include the accuracy rate of each model.
Gaviria et al. [45]	2022	D.L.	ANN, LTSM, CNN, SVM, RF	PV field	Review on the interest of ML in PV systems, it gives the resources to find the data sets, the source codes and presents each theme with the source code and the data	The text lacks objectivity and precision in presenting the results. Additionally, there are several articles on the diagnosis and detection of defects using machine learning approaches that are not significant enough.

Table 2. Summary of the few review articles from the 2016–2023 period on fault diagnosis of PV systems.

Authors	Year	Type of Techniques	ML Models	PV System Component	Contribution	Remarks
Autors	Ical	Type of feeliniques	WIE WOUCHS	I V System component	Contribution	Kentarkö
Hammoudi et al. [49]	2022	D.L.	CNN and LTSM	PV field	Survey on the interest of Deep Learning and IoT in the maintenance of PV systems	The text is limited to discussing deep learning in preventive maintenance on the DC side.
Yuan et al. [80]	2022	M.L.	ANN	PV field	Review on the progress of ANN in fault diagnosis	The text lacks information on the precision and complexity of each type of ANN.
Forootan et al. [50]	2022	ML, DL	SVM, DA, BN, ANN, kNN, RF, DT, CNN, FL, ANFIS, GA, LTSM, RL, MLR, SLR, k-Means etc.	PV field	Review on the use and application of ML and DL algorithms in energy systems	Additionally, it fails to consider non-visual defects and lacks precision.
Berghout et al. [22]	2022	ML, DL	SVM, kNN, MLP, LTSM, CNN, Gans	PV field	Presents the various works on monitoring PV systems and shading and degradation defects	However, it focuses on ML categories, detection techniques, and two types of defects. The accuracy of each model is not provided.
Puthiyapurayil et al. [70]	2022	AI, signal-based method	ANN, BPNN, SVM, CNN	Inverter	Lists work on the different methods of diagnosing the open-circuit fault in an NPC inverter	The text only focuses on single switch open-circuit faults as three switch faults are rare.
Engel et al. [44]	2022	ML, DL	ANN, CNN, ANFIS, YOLOv4, k-NN, DT, SVM, RF, NB	PV field	Review of ML advances in prediction, forecasting, sizing and diagnosis of PV systems	methods shows that the DNN model provides better information and performance in the diagnostic process compared to the non-neural model.
Ying-Yi et al. [51]	2022	Visual and thermal	SVM, kNN, MSD, DT, RF, ANFIS, ANN	PV field	Presents the different traditional methods on the detection and classification of PV faults and a projection on AI techniques	The study focuses on traditional methods and demonstrates the potential of ML techniques.
Osmani et al. [52]	2023	Conventional methods, AI	SCADA, ANN, KELM (kernel extreme learning machine)	PV field	Critical review of detection methods in the PV field	This text presents the DC and AC side faults of the field, as well as the detection method. It focuses on conventional methods and does not mention any supervised learning methods.
Islam et al. [37]	2023	Artificial intelligence based on ML and DL	AdaBoost, ANN, CNN, RNN, SVM, RF	PV field	Systematic review on identification and diagnosis methods, they compare existing reviews with its review in terms of technical approaches for fault detection	The most effective DL and ML approaches for diagnosing PV faults are identified, and it is shown that DL outperforms conventional approaches. ANN is proposed for diagnosis, but no accuracy rate is provided for different methods.

Table 2. Cont.

4. Synthesis

Based on the studies carried out on more than 100 articles, 133 articles were used in this work. Looking at the summary of the different journals in Table 2 and the research articles used, DC electrical and environmental disturbances are the disturbances that the authors pay the most attention to. In addition, most of the journal articles listed in this article have focused their work on either the DC part (PV field) or the AC part (PV inverter). Few articles focus on both the DC and AC side faults simultaneously. Therefore, this review integrates the faults and methods used in both parts. The machine learning techniques used in this paper show good results and excellent performance in the field of PV fault diagnosis. However, this performance varies from one model to another depending on the quality and quantity of data used. For example, ANN requires a very large amount of data to achieve good accuracy, while kNN shows poor results as the number of data increases. For these reasons, some models are used more than others, and depending on the results sought, a combination of models is necessary to improve the shortcomings of simple models. ANN algorithms have been widely used in fault diagnosis in recent years compared to other models (kNN, RF, DT, SVM), although they are also regularly used in this context Indeed, the growing development of ANN models is reflected in reviews that show their ability to solve complex problems with large amounts of data. Furthermore, in hybrid models, the ANFIS model is increasingly applied, especially in the prediction of PV systems. This algorithm uses its fuzzy version to improve the shortcomings of the ANN in finding the maximum power point (MPPT). It is more accurate in fault prediction and diagnosis. It is also shown that the integration of fuzzy logic in traditional k-NN improves the accuracy of the k-NN model and gives excellent consistency. However, although this model is used in many areas, it is still lagging behind in the diagnosis of faults in grid-connected PV systems. The diagram in Figure 18 gives an overview of the occurrence of the different algorithms found in the journals used in this article over the period 2016–2023.



Figure 18. Distribution of the implementation rate of the different ML models.

Observing the above graph, we see that the artificial neural network algorithm is widely used in fault diagnosis, with a 22% percentage. Similarly, the hybrid model based on the adaptive neuro-fuzzy inference system (ANFIS) has a higher implementation rate than the others. However, the FKNN and AdaBoost algorithms are poorly implemented in the context of PV system diagnosis, despite their better performance.

5. Proposed Method

In the previous section, several supervised learning algorithms used in fault diagnosis showed the benefit of machine learning in improving diagnostic results in terms of time. This section describes the Extra Trees algorithm as a model capable of detecting and classifying faults in PV systems.

The Extra Trees or Extreme Model Randomized Trees is a supervised learning algorithm from the Ensemble Learning family of algorithms. This algorithm was first proposed and implemented in 2006 in a paper entitled Extremely random tree by Geurts et al. [128]. The model is widely used for both regression and classification problems. The Extra Trees (ET) model uses a set of decision trees to randomly construct a group of unpruned decision trees to reduce the risk of overfitting [130]. The construction of the Extra Trees algorithm is similar to the Random Forest (RF) algorithm which is a tree-like combination of multiple trees [131]. However, the difference between the two algorithms is that the ET model uses the entire training sample to train each tree instead of a Bootstrap sample used by the RF model. Also, the selection of division points is conducted randomly unlike the RF model which uses an optimal distribution. The ET model execution procedure is based on the analysis of the training dataset (S) and three important hyperparameters [130]:

- The number of trees (M) to train based on the number of training samples S
- The number of attributes (K) to be randomly selected and used in each node for each trained ensemble tree
- The minimum number of samples/instances (N_{min}) needed to split a node of each trained ensemble tree

After training the trees, the algorithm makes its final prediction based on test data through majority voting for classification or by calculating the arithmetic mean for regression [38]. The Extra Trees or Extremely Randomized Trees model is widely used in various fields, including health [39,129,132], economy [38], transportation [129], and telecommunications [41–43]. For instance, in medicine, the Extra Trees model has been proposed to detect and classify cancer as malignant or benign tumors. The obtained results indicate an accuracy of 99.27% [40]. Saeed et al. proposed the Extra Trees model to detect faults in a wireless sensor network. The results demonstrate superior performance compared to other machine learning algorithms in the literature, such as artificial neural networks, support vector machines (SVM), Random Forest, and decision trees [43]. Bai et al. utilized the Extra Trees algorithm to predict short-term traffic flow in a non-stationary environment. The obtained result showed higher precision compared to existing methods [133]. To the best of our knowledge, this algorithm has not been proposed before in the context of photovoltaic fault diagnosis. Additionally, the Extra Trees algorithm exhibits high precision, lower computational complexity, and variance compared to other models such as decision trees, support vector machines, artificial neural network (ANN), Random Forest, and decision trees [40]. This is an opportunity to propose the Extra Trees model as an effective algorithm that addresses the inadequacies of other models, such as decision trees, the AdaBoost model, SVM, DT, KNN, and FKNN [133]. The execution algorithm for the Extra Trees model is presented in Figure 19.



Figure 19. ETC model execution algorithm.

6. Conclusions and Future Recommendations

Photovoltaic systems are becoming more prevalent. It is crucial to diagnose faults to maintain their reliability and safety. This article provides an overview of the most common faults in PV systems and the diagnostic methods based on the supervised learning algorithms that are commonly used to resolve them. The most common and dangerous faults are environmental and electrical. It is crucial to promptly identify environmental faults to prevent them from causing electrical faults. This article examines various supervised learning algorithms for fault diagnosis and compares their effectiveness based on diagnostic techniques, measured data, and proposed approaches. In recent years, artificial neural networks (ANN) have gained popularity for fault diagnosis. This paper proposes the Extra Trees model as a highly effective algorithm for fault diagnosis due to its ability to reduce bias and avoid overfitting problems. The article also compares the performance of various models, including decision trees (DT), naive Bayes (NB), support vector machines (SVM), random forests (RF), k-nearest neighbors (KNN), and fuzzy k-nearest neighbors (FKNN). It provides guidance to technicians, engineers, and researchers on the use of supervised learning algorithms for fault diagnosis in photovoltaic systems. Possible future work includes the utilization of the Extremely Randomized Trees for PV system diagnosis, the integration of various diagnostic parameters, the analysis of sensors for data acquisition, and the incorporation of the Internet of Things into diagnostics using the Extra Trees model.

Author Contributions: G.M.T.T. worked on the research paper, J.K., D.K. and F.-D.M. participated in proofreading the paper and S.S.O.-D. carried out supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the World Bank, through the Regional Center of Excellence for Electricity Management (CERME). Source of funding: credit IDA 6512-TG and donation IDA 536 IDA (World Bank).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data can be used upon request.

Acknowledgments: The authors are deeply grateful to the World Bank for funding this research through CERME (Regional Center of Excellence for Electricity Management). We also thank the reviewers for their insightful and positive comments and suggestions, which have greatly helped us improve this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

PV	Photovoltaics
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro Fuzzy Inference System
D.L.	Deep Learning
M.L.	Machine Learning
KNN	K Nearest Neighbor
NB	Naive Bayes
RF	Random Forest
DT.	Decision Tree
SVM	Vector Machine Support
CNN	Convolutional Neural Network
FL	Fuzzy Logic
NLP	Natural Language Processing
PVI	Photovoltaic Inverter
IGBT	Insulate Gate Bipolar Transistor
AdaBoost	Adaptive Boosting
FKNN	Fuzzy K Nearest Neighbor
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LSTM	Long Short Term Memory
YOLOv4	You Only Look Once
BPNN	Backpropagation Neural Network
MLP	Multilayer Propagation
ETC	Extra Trees Classifier
KELM	Kernel Extreme Learning Machine
Μ	Number of trees
N _{min}	The minimum number of samples/instances
Κ	The number of attributes used at each node
S	Training sets

References

- 1. TrendForce. Global Solar Installation May Hit 350.6 GW. Available online: https://www.pv-magazine.com/2023/03/09/pv-product-prices-resume-downward-trend-says-trendforce/ (accessed on 25 April 2023).
- Livera, A.; Theristis, M.; Makrides, G.; Georghiou, G.E. Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems. *Renew. Energy* 2019, 133, 126–143. [CrossRef]
- Madeti, S.R.; Singh, S.N. Modeling of PV system based on experimental data for fault detection using kNN method. *Sol. Energy* 2018, 173, 139–151. [CrossRef]
- 4. Oh, W.-G. A Fault Detection Scheme in Acoustic Sensor Systems Using Multiple Acoustic Sensors. J. Korea Inst. Electron. Commun. Sci. 2016, 11, 203–208. [CrossRef]

- 5. Abubakar, A.; Almeida, C.F.M.; Gemignani, M. Review of artificial intelligence-based failure detection and diagnosis methods for solar photovoltaic systems. *Machines* **2021**, *9*, 328. [CrossRef]
- 6. Jiang, Y.; Yin, S.; Kaynak, O. Optimized design of parity relation-based residual generator for fault detection: Data-driven approaches. *IEEE Trans. Industr. Inform.* **2021**, *17*, 1449–1458. [CrossRef]
- Mohammad, S.; Sudhakar, K. Machine Learning-Autonomous Vehicles. Int. J. Manag. 2018, 8. Available online: http://www. ijmra.us (accessed on 29 December 2023).
- Livera, A.; Theristis, M.; Makrides, G.; Georghiou, G.E.; Sutterlueti, J.; Georghiou, G.E. Advanced Diagnostic Approach of Failures for Grid-Connected Photovoltaic (PV) Systems PV-Estia-Enhancing Storage Integration in Buildings with Photovoltaics View Project Modeling and Optimization of Advanced Energy Systems View Project Advanced Diagnostic Approach of Failures for Grid-Connected Photovoltaic (PV) Systems. 2018. Available online: https://userarea.eupvsec.org/proceedings/35th-EU-PVSEC-2018/6BO.6.5/ (accessed on 8 May 2023).
- Karatepe, E.; Syafaruddin; Hiyama, T. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In Proceedings of the 16th International Conference on Intelligent System Applications to Power Systems, Hersonissos, Greece, 25–28 September 2011; pp. 1–6.
- 10. Bendary, A.F.; Abdelaziz, A.Y.; Ismail, M.M.; Mahmoud, K.; Lehtonen, M.; Darwish, M.M.F. Proposed anfis based approach for fault tracking, detection, clearing and rearrangement for photovoltaic system. *Sensors* **2021**, *21*, 2269. [CrossRef] [PubMed]
- Soffiah, K.; Manoharan, P.S.; Deepamangai, P. Fault detection in grid connected pv system using artificial neural network. In Proceedings of the 7th International Conference on Electrical Energy Systems, ICEES, Chennai, India, 11–13 February 2021; pp. 420–424. [CrossRef]
- 12. Gong, S.; Wu, X.; Zhang, Z. Fault diagnosis method of photovoltaic array based on random forest algorithm. In Proceedings of the 2020 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; pp. 4249–4425.
- 13. Hussain, M.; Dhimish, M.; Titarenko, S.; Mather, P. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renew. Energy* **2020**, *155*, 1272–1292. [CrossRef]
- 14. Alimi, O.A.; Meyer, E.L.; Olayiwola, O.I. Solar photovoltaic modules' performance reliability and degradation analysis: A review. *Energies* **2022**, *15*, 5964. [CrossRef]
- 15. Romero, H.F.M.; Rebollo, M.G.; Cardeñoso-Payo, V.; Gómez, V.A.; Plaza, A.R.; Moyo, R.T.; Hernández-Callejo, L. Applications of artificial intelligence to photovoltaic systems: A review. *Appl. Sci.* 2022, 12, 56. [CrossRef]
- 16. Xie, C.; Chen, S.; Guo, F.; Liu, X. A deep residual recurrent neural network model-augmented attention with physical characteristics: Application to turntable servo system. *IEEE Trans. Ind. Electron.* **2022**, *69*, 489. [CrossRef]
- 17. Yau, H.T.; Prior, S.D.; Wang, Y.; Li, Y. IEEE Access Special Section Editorial: Advanced artificial intelligence technologies for smart manufacturing. *IEEE Access* 2021, *9*, 119232–119234. [CrossRef]
- 18. Fei, Z.; Zhang, Z.; Tsui, K.L. Deep learning powered online battery health estimation considering multi-timescale ageing dynamics and partial charging information. *IEEE Trans. Transp. Electrif.* **2023**. [CrossRef]
- Amiri, A.F.; Kichou, S.; Oudira, H.; Chouder, A.; Silvestre, S. Fault detection and diagnosis of a photovoltaic system based on deep learning using the combination of a convolutional neural network (cnn) and bidirectional gated recurrent unit (Bi-GRU). *Sustainability* 2024, *16*, 1012. [CrossRef]
- 20. Rocha, H.R.O.; Fiorotti, R.; Fardin, J.F.; Garcia-Pereira, H.; Bouvier, Y.E.; Rodríguez-Lorente, A.; Yahyaoui, I. Application of AI for short-term pv generation forecast. *Sensors* 2023, 24, 85. [CrossRef]
- 21. Li, B.; Delpha, C.; Diallo, D.; Migan-Dubois, A. Application of artificial neural networks to photovoltaic fault detection and diagnosis: A review. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110512. [CrossRef]
- 22. Berghout, T.; Benbouzid, M.; Bentrcia, T.; Ma, X.; Djurović, S.; Mouss, L.H. Machine learning-based condition monitoring for pv systems: State of the art and future prospects. *Energies* **2021**, *14*, 6316. [CrossRef]
- 23. Al Smadi, T.; Handam, A.; Gaeid, K.S.; Al-Smadi, A.; Al-Husban, Y.; Khalid, A.S. Artificial intelligent control of energy management PV system. *Results Control. Optim.* 2024, 14, 100343. [CrossRef]
- 24. Boubaker, S.; Kamel, S.; Ghazouani, N.; Mellit, A. Assessment of Machine and Deep Learning Approaches for Fault Diagnosis in Photovoltaic Systems Using Infrared Thermography. *Remote Sens.* **2023**, *15*, 1686. [CrossRef]
- 25. Mansouri, M.; Trabelsi, M.; Nounou, H.; Nounou, M. Deep learning-based fault diagnosis of photovoltaic systems: A comprehensive review and enhancement prospects. *IEEE Access* 2021, *9*, 126286–126306. [CrossRef]
- Kuo, W.C.; Chen, C.H.; Hua, S.H.; Wang, C.C. Assessment of different deep learning methods of power generation forecasting for solar pv system. *Appl. Sci.* 2022, 12, 7529. [CrossRef]
- Hichri, A.; Hajji, M.; Mansouri, M.; Nounou, H.; Bouzrara, K. Supervised machine learning-based salp swarm algorithm for fault diagnosis of photovoltaic systems. J. Eng. Appl. Sci. 2024, 71, 12. [CrossRef]
- 28. Chen, Z.; Han, F.; Wu, L.; Yu, J.; Cheng, S.; Lin, P.; Chen, H. Random Forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. *Energy Convers. Manag.* **2018**, *178*, 250–264. [CrossRef]
- 29. Badr, M.M.; Hamad, M.S.; Abdel-Khalik, A.S.; Hamdy, R.A.; Ahmed, S.; Hamdan, E. Fault identification of photovoltaic array based on machine learning classifiers. *IEEE Access* **2021**, *9*, 159113–159132. [CrossRef]
- Dhimish, M.; Badran, G. Photovoltaic hot-spots fault detection algorithm using fuzzy systems. *IEEE Trans. Device Mater. Reliab.* 2019, 19, 671–679. [CrossRef]

- 31. Dhimish, M.; Holmes, V.; Mehrdadi, B.; Dales, M. Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection. *Renew. Energy* 2018, 117, 257–274. [CrossRef]
- 32. Dhimish, M.; Holmes, V. Fault detection algorithm for grid-connected photovoltaic plants. *Sol. Energy* **2016**, *137*, 236–245. [CrossRef]
- Mellit, A.; Tina, G.M.; Kalogirou, S.A. Fault detection and diagnosis methods for photovoltaic systems: A review. *Renew. Sustain.* Energy Rev. 2018, 91, 1–17. [CrossRef]
- Yun, L.; Bofeng, Y.; Dan, Q.; Fengshuo, L. Research on fault diagnosis of photovoltaic array based on random forest algorithm. In Proceedings of the 2021 IEEE International Conference on Power Electronics, Computer Applications, ICPECA, Shenyang, China, 22–24 January 2021; pp. 194–198. [CrossRef]
- 35. Wang, X.; Yang, W.; Qin, B.; Wei, K.; Ma, Y.; Zhang, D. Intelligent monitoring of photovoltaic panels based on infrared detection. *Energy Rep.* **2022**, *8*, 5005–5015. [CrossRef]
- Pillai, D.S.; Rajasekar, N. A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renew. Sustain.* Energy Rev. 2018, 91, 18–40. [CrossRef]
- Islam, M.; Rashel, M.R.; Ahmed, M.T.; Islam, A.K.M.K.; Tlemçani, M. Artificial intelligence in photovoltaic fault identification and diagnosis: A systematic review. *Energies* 2023, 16, 7417. [CrossRef]
- Al Mahkya, D.; Notodiputro, K.A.; Sartono, B. Extra trees method for stock price forecasting with rolling origin accuracy evaluation. *Media Stat.* 2022, 15, 36–47. [CrossRef]
- 39. Mathew, T.E. An optimized extremely randomized tree model for breast cancer classification. *J. Theor. Appl. Inf. Technol.* **2022**, *31*, 5234–5246.
- 40. Aminifar, A.; Shokri, M.; Rabbi, F.; Pun, V.K.I.; Lamo, Y. Extremely randomized trees with privacy preservation for distributed structured health data. *IEEE Access* 2022, *10*, 6010–6027. [CrossRef]
- AlOmar, M.K.; Hameed, M.M.; AlSaadi, M.A. Multi hours ahead prediction of surface ozone gas concentration: Robust artificial intelligence approach. *Atmos. Pollut. Res.* 2020, 11, 1572–1587. [CrossRef]
- Saeed, U.; Jan, S.U.; Lee, Y.D.; Koo, I. Fault diagnosis based on extremely randomized trees in wireless sensor networks. *Reliab.* Eng. Syst. Saf. 2021, 205, 107284. [CrossRef]
- Acosta, M.R.C.; Ahmed, S.; Garcia, C.E.; Koo, I. Extremely randomized trees-based scheme for stealthy cyber-attack detection in smart grid networks. *IEEE Access* 2020, *8*, 19921–19933. [CrossRef]
- 44. Engel, E.; Engel, N. A Review on Machine Learning Applications for Solar Plants. Sensors 2022, 22, 9060. [CrossRef]

45. Gaviria, J.F.; Narváez, G.; Guillen, C.; Giraldo, L.F.; Bressan, M. Machine learning in photovoltaic systems: A review. *Renew.* Energy 2022, 196, 298–318. [CrossRef]

- 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* 2021, 14, 4690. [CrossRef]
- Appiah, A.Y.; Zhang, X.; Ayawli, B.B.K.; Kyeremeh, F. Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques. *Int. J. Photoenergy* 2019, 2019, 6953530. [CrossRef]
- Zenebe, T.M.; Midtgård, O.-M.; Völler, S.; Cali, Ü.; Cali, U. EasyChair Preprint Machine Learning for PV System Operational Fault Analysis: Literature Review Machine Learning for PV System Operational Fault Analysis: Literature Review; Springer: Berlin/Heidelberg, Germany, 2021.
- 49. Hammoudi, Y.; Idrissi, I.; Boukabous, M.; Zerguit, Y.; Bouali, H. Review on maintenance of photovoltaic systems based on deep learning and internet of things. *Indones. J. Electr. Eng. Comput. Sci.* 2022, *26*, 1060–1072. [CrossRef]
- 50. Forootan, M.M.; Larki, I.; Zahedi, R.; Ahmadi, A. Machine Learning and Deep Learning in Energy Systems: A Review. *Sustainability* **2022**, *14*, 4832. [CrossRef]
- 51. Hong, Y.Y.; Pula, R.A. Methods of photovoltaic fault detection and classification: A review. *Energy Rep.* **2022**, *8*, 5898–5929. [CrossRef]
- 52. Osmani, K.; Haddad, A.; Lemenand, T.; Castanier, B.; Alkhedher, M.; Ramadan, M. A critical review of PV systems' faults with the relevant detection methods. *Energy Nexus* 2023, *12*, 100257. [CrossRef]
- Vai, V.; Chhorn, S.; Chhim, R.; Tep, S.; Bun, L. Modeling and Simulation of PV Module for Estimating Energy Production under Uncertainties. In Proceedings of the 2020 8th International Electrical Engineering Congress, IEECON 2020, Chiang Mai, Thailand, 4–6 March 2020. [CrossRef]
- 54. Arani, M.S.; Hejazi, M.A. The comprehensive study of electrical faults in PV arrays. J. Electr. Comput. Eng. 2016, 2016, 8712960. [CrossRef]
- Aouchiche, N. Défauts Liés Aux Systèmes Photovoltaïques Autonomes et Techniques de Diagnostic-Etat de l'art. 2018. Available online: https://www.researchgate.net/publication/328577571 (accessed on 14 June 2023).
- 56. Garoudja, E.; Chouder, A.; Kara, K.; Silvestre, S. An enhanced machine learning based approach for failures detection and diagnosis of PV systems. *Energy Convers. Manag.* **2017**, *151*, 496–513. [CrossRef]
- 57. Madeti, S.R.; Singh, S.N. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Sol. Energy* **2017**, *158*, 161–185. [CrossRef]
- Trejo, D.R.E.; Bárcenas, E.; Díez, J.E.H.; Bossio, G.; Pérez, G.E. Open- and short-circuit fault identification for a boost DC/DC converter in PV MPPT systems. *Energies* 2018, 11, 616. [CrossRef]

- 59. Guerriero, P.; Piegari, L.; Rizzo, R.; Daliento, S. Mismatch based diagnosis of pv fields relying on monitored string currents. *Int. J. Photoenergy* **2017**, 2017, 2834685. [CrossRef]
- Abdulmawjood, K.; Refaat, S.S.; Morsi, W.G. Detection and prediction of faults in photovoltaic arrays: A review. In Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering, CPE-POWERENG 2018, Doha, Qatar, 10–12 April 2018; pp. 1–8. [CrossRef]
- 61. Roger, P.Y.; Emilio, C.C.J.; Rubén, R.H. Fault diagnostic methodology for grid-connected photovoltaic systems. *J. Multiapp.* **2021**, 2, 10–30. [CrossRef]
- 62. Maghami, M.R.; Mutambara, A.G.O. Challenges associated with hybrid energy systems: An artificial intelligence solution. *Energy Rep.* **2023**, *9*, 924–940. [CrossRef]
- Khalil, I.U.; Ul-Haq, A.; Mahmoud, Y.; Jalal, M.; Aamir, M.; Ahsan, M.U.; Mehmood, K. Comparative analysis of photovoltaic faults and performance evaluation of its detection techniques. *IEEE Access* 2020, *8*, 26676–26700. [CrossRef]
- Bhimrao, B.; Vishwakarma, S. Study of partial shading effect on solar module using MATLAB development of a MATLAB. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. 2017, 6, 5303–5308.
- 65. Malvoni, M.; Chaibi, Y. Machine learning based approaches for modeling the output power of photovoltaic array in real outdoor conditions. *Electronics* **2020**, *9*, 315. [CrossRef]
- Dhakshinamoorthy, M.; Sundaram, K.; Murugesan, P.; David, P.W. Bypass diode and photovoltaic module failure analysis of 1.5kW solar PV array. *Energy Sources Part A Recovery Util. Environ. Eff.* 2022, 44, 4000–4015. [CrossRef]
- 67. Platon, R.; Martel, J.; Woodruff, N.; Chau, T.Y. Online fault detection in pv systems. *IEEE Trans. Sustain. Energy* 2015, *6*, 1200–1207. [CrossRef]
- 68. Kim, S.; Kim, S. Performance estimation modeling via machine learning of an agrophotovoltaic system in South Korea. *Energies* **2021**, *14*, 6724. [CrossRef]
- Im, W.S.; Kim, J.S.; Kim, J.M.; Lee, D.C.; Lee, K.B. Diagnosis methods for IGBT open switch fault applied to 3-phase AC/DC PWM converter. J. Power Electron. 2012, 12, 120–127. [CrossRef]
- Puthiyapurayil, M.R.M.K.; Nasirudeen, M.N.; Saywan, Y.A.; Ahmad, M.W.; Malik, H. A Review of Open-Circuit Switch Fault Diagnostic Methods for Neutral Point Clamped Inverter. *Electronics* 2022, 11, 3169. [CrossRef]
- Gunda, T.; Hacket, S.; Kraus, L.; Downs, C.; Jones, R.; Mcnalley, C.; Bolen, M.; Walker, A. A Machine learning evaluation of maintenance records for common failure modes in PV inverters. *IEEE Access* 2020, *8*, 211610. [CrossRef]
- 72. Zouinar, M. Developments in artificial intelligence: What are the challenges for human activity and the human-machine relationship at work? *Activities* **2020**. [CrossRef]
- Vodapally, S.N.; Ali, M.H. Overview of intelligent inverters and associated cybersecurity issues for a grid-connected solar photovoltaic system. *Energies* 2023, 16, 5904. [CrossRef]
- 74. Li, Y.; Ding, K.; Zhang, J.; Chen, F.; Chen, X.; Wu, J. A fault diagnosis method for photovoltaic arrays based on fault parameters identification. *Renew. Energy* **2019**, *143*, 52–63. [CrossRef]
- Sarikh, S.; Raoufi, M.; Bennouna, A.; Benlarabi, A.; Ikken, B. Fault diagnosis in a photovoltaic system through I–V characteristics analysis. In Proceedings of the 2018 9th International Renewable Energy Congress, IREC 2018, Hammamet, Tunisia, 26–28 March 2018; pp. 1–6. [CrossRef]
- 76. Rahmoune, M.B.; Iratni, A.; Amari, A.S.; Hafaifa, A.; Colak, I. Fault detection and diagnosis of photovoltaic system based on neural networks approach. *Diagnostyka* **2023**, *24*, 166428. [CrossRef]
- Wang, Z.; Yao, L.; Cai, Y.; Zhang, J. Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis. *Renew. Energy* 2020, 155, 1312–1327. [CrossRef]
- Zhao, Y.; Yang, L.; Lehman, B.; De Palma, J.F.; Mosesian, J.; Lyons, R. Decision tree-based fault detection and classification in solar photovoltaic arrays. In Proceedings of the Conference Proceedings—IEEE Applied Power Electronics Conference and Exposition—APEC, Orlando, FL, USA, 5–9 February 2012; pp. 93–99. [CrossRef]
- 79. Ashok, V.; Yadav, A.; Naik, V.K. Fault detection and classification of multi-location and evolving faults in double-circuit transmission line using ANN. In *Advances in Intelligent Systems and Computing*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 758, pp. 307–317. [CrossRef]
- Yuan, Z.; Xiong, G.; Fu, X. Artificial neural network for fault diagnosis of solar photovoltaic systems: A Survey. *Energies* 2022, 15, 8693. [CrossRef]
- Mittal, M.; Bora, B.; Saxena, S.; Gaur, A.M. Performance prediction of PV module using electrical equivalent model and artificial neural network. Sol. Energy 2018, 176, 104–117. [CrossRef]
- 82. Chine, W.; Mellit, A.; Lughi, V.; Malek, A.; Sulligoi, G.; Pavan, A.M. A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. *Renew. Energy* **2016**, *90*, 501–512. [CrossRef]
- Khelil, C.K.M.; Amrouche, B.; Benyoucef, A.S.; Kara, K.; Chouder, A. New intelligent fault diagnosis (IFD) approach for grid-connected photovoltaic systems. *Energy* 2020, 211, 118591. [CrossRef]
- 84. Asghar, F.; Talha, M.; Kim, S.H. Neural network-based fault detection and diagnosis system for three-phase inverter in variable speed drive with induction motor. *J. Control. Sci. Eng.* **2016**, 2016, 1–12. [CrossRef]
- Barakate, A.A.; Rida, A.; Wahbi, A.; Maddi, M.; Hlou, L.; Hadjoudja, A. Modeling, development and analysis performance of an intelligent control of photovoltaic system by fuzzy logic approach for maximum power point tracking. *Int. J. Commun. Netw. Inf. Secur.* 2021, 13, 42–47. [CrossRef]

- Abid, M.; Laribi, S.S.; Al-Asgar, Z.S.; Larbi, M. Artificial neural network approach assessment of short-circuit fault detection in a three-phase inverter. In Proceedings of the 2021 International Congress of Advanced Technology and Engineering, ICOTEN, Taiz, Yemen, 4–5 July 2021. [CrossRef]
- 87. Parimalasundar, E.; Kumar, R.S.; Chandrika, V.S.; Suresh, K. Fault diagnosis in a five-level multilevel inverter using an artificial neural network approach. *Electr. Eng. Electromechanics* **2023**, 2023, 31–39. [CrossRef]
- 88. Salem, F.; Awadallah, M.A. Detection and assessment of partial shading in photovoltaic arrays. *J. Electr. Syst. Inf. Technol.* **2016**, *3*, 23–32. [CrossRef]
- 89. Dhimish, M.; Tyrrell, A.M. This Is a Repository Copy of Photovoltaic Bypass Diode Fault Detection Using Artificial Neural Networks. Photovoltaic Bypass Diode Fault Detection Using Artificial Neural Networks. Available online: https://eprints.whiterose.ac.uk/ (accessed on 21 December 2023).
- 90. Thakur, P.; Tripathi, D.R.N.; Mishra, H. Performance analysis of ANN Based DC To DC Converter. *Int. J. Eng. Res. Appl.* **2023**, *13*, 24–30.
- Zendehboudi, A.; Baseer, M.A.; Saidur, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. J. Clean. Prod. 2018, 199, 272–285. [CrossRef]
- Yin, Z.; Hou, J. Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. *Neurocomputing* 2016, 174, 643–650. [CrossRef]
- Natarajan, K.; Kumar, B.P.; Kumar, V.S.; Kumar, P. Fault detection of solar PV system using SVM and thermal image processing. Int. J. Renew. Energy Res. 2020, 10, 967–977.
- Kuraku, N.V.P.; He, Y.; Ali, M. Fault diagnosis of open circuit multiple igbt's using ppca-svm in single phase five-level voltagecontrolled h-bridge MLI. *IEEJ J. Ind. Appl.* 2020, 9, 61–72. [CrossRef]
- Harrou, F.; Dairi, A.; Taghezouit, B.; Sun, Y. An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class support vector machine. *Sol. Energy* 2019, 179, 48–58. [CrossRef]
- 96. Fatima, M.; Pasha, M. Survey of machine learning algorithms for disease diagnostic. J. Intell. Learn. Syst. Appl. 2017, 9, 1–16. [CrossRef]
- Eskandari, A.; Milimonfared, J.; Aghaei, M. Optimization of SVM classifier using Grid Search Method for Line-Line Fault detection of photovoltaic systems. In Proceedings of the Conference Record of the IEEE Photovoltaic Specialists Conference, Calgary, AB, Canada, 15 June–21 August 2020; pp. 1134–1137. [CrossRef]
- Harrou, F.; Taghezouit, B.; Sun, Y. Improved KNN-Based monitoring schemes for detecting faults in PV systems. *IEEE J. Photovolt.* 2019, 9, 811–821. [CrossRef]
- 99. Qin, J.; Wang, L.; Huang, R. Research on Fault Diagnosis Method of Spacecraft Solar Array Based on f-KNN Algorithm. In Proceedings of the Prognostics and System Health Management Conference (PHM Harbin), Harbin, China, 9–12 July 2017.
- 100. Janarthanan, R.; Maheshwari, R.U.; Shukla, P.K.; Shukla, P.K.; Mirjalili, S.; Kumar, M. Intelligent detection of the PV faults based on artificial neural network and type 2 fuzzy systems. *Energies* **2021**, *14*, 6584. [CrossRef]
- Majd, A.A.; Samet, H.; Ghanbari, T. k-NN based fault detection and classification methods for power transmission systems. *Prot. Control. Mod. Power Syst.* 2017, 2, 32. [CrossRef]
- Nguyen, X.H.; Nguyen, M.P. Mathematical modeling of photovoltaic cell/module/arrays with tags in Matlab/Simulink. *Environ.* Syst. Res. 2015, 4, 24. [CrossRef]
- 103. Xu, H.; Peng, Y.; Su, L. Research on open circuit fault diagnosis of inverter circuit switching tube based on machine learning algorithm. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, 452, 042015. [CrossRef]
- 104. Galeano, A.G.; Bressan, M.; Vargas, F.J.; Alonso, C. Shading ratio impact on photovoltaic modules and correlation with shading patterns. *Ratio Impact Photovolt. Modul. Correl. Shading Patterns. Energ.* **2018**, *11*, 852. [CrossRef]
- 105. Kadri, F.; Charif, F.; Tamissa, Y.; Benchabane, A.; Hamida, M.A. Multiple Fuzzy Diagnosis for Voltage Source Inverter Open Circuit Fault in Direct Torque Control Induction Motor Drive Biometrics Identification and Authentication View Project Fault Diagnosis for Voltage Source Inverter View Project Younes Tamissa Multiple Fuzzy Diagnosis for Voltage Source Inverter Open Circuit Fault in Torque Direct Control Induction Motor Drive. Available online: https://www.researchgate.net/publication/358573643 (accessed on 14 June 2023).
- Mehta, P.; Sahoo, S.; Dhiman, H. Open circuit fault diagnosis in five-level cascaded h-bridge inverter. *Int. Trans. Electr. Energy* Syst. 2022, 2022, 1–13. [CrossRef]
- 107. Zaki, S.A.; Zhu, H.; Yao, J. Fault detection and diagnosis of photovoltaic system using fuzzy logic control. *E3S Web Conf.* 2019, 107, 02001. [CrossRef]
- Abbas, M.; Zhang, D. A smart fault detection approach for PV modules using Adaptive Neuro-Fuzzy Inference framework. Energy Rep. 2021, 7, 2962–2975. [CrossRef]
- Tojeiro, D.O.; Olazabal, D. ScienceDirect fault detection based on detection based on detection based on and residual evaluation with fuzzy fault detection based on and residual evaluation with fuzzy models and residual evaluation with models and residual evaluation. *IFAC Pap.* 2021, 54, 717–722. [CrossRef]
- 110. Yi, Z.; Etemadi, A.H. Line-to-line fault detection for photovoltaic arrays based on multi-resolution signal decomposition and two-stage support vector machine. *IEEE Trans. Ind. Electron.* **2017**, *64*, 8546–8556. [CrossRef]
- 111. You, L.; Ling, Z.; Cui, Y.; Cai, W.; He, S. Open circuit fault detection of t-type grid connected inverters using fast s transform and random forest. *Entropy* **2023**, *25*, 778. [CrossRef] [PubMed]

- 112. Xia, K.; He, S.; Tan, Y.; Jiang, Q.; Xu, J.; Yu, W. Wavelet packet and support vector machine analysis of series DC ARC fault detection in photovoltaic system. *IEEJ Trans. Electr. Electron. Eng.* **2019**, *14*, 192–200. [CrossRef]
- 113. Wang, T.; Bi, T.; Wang, H.; Liu, J. Decision tree based online stability assessment scheme for power systems with renewable generations. *CSEE J. Power Energy Syst.* **2015**, *1*, 53–61. [CrossRef]
- 114. Benkercha, R.; Moulahoum, S. Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system. *Sol. Energy* **2018**, *173*, 610–634. [CrossRef]
- 115. Lakshmanaprabu, S.K.; Shankar, K.; Ilayaraja, M.; Nasir, A.W.; Vijayakumar, V.; Chilamkurti, N. Random Forest for big data classification in the internet of things using optimal features. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2609–2618. [CrossRef]
- Xia, J.; Zhang, S.; Cai, G.; Li, L.; Pan, Q.; Yan, J.; Ning, G. Adjusted weight voting algorithm for random forests in handling missing values. *Pattern Recognit.* 2017, 69, 52–60. [CrossRef]
- Liu, S.; Qian, X.; Wan, H.; Ye, Z.; Wu, S.; Ren, X. NPC Three-level inverter open-circuit fault diagnosis based on adaptive electrical period partition and random forest. J. Sens. 2020, 2020, 1–18. [CrossRef]
- 118. Shin, J.H.; Kim, J.O. On line diagnosis and fault state classification method of photovoltaic plant. Energies 2020, 13, 4584. [CrossRef]
- Wang, L.; Liu, J.; Guo, X.; Yang, Q.; Yan, W. Online Fault Diagnosis of Photovoltaic Modules Based on Multi-Class Support Vector Machine. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017.
- Edwards, P.K.; Duhon, D.; Scotiabank, S.S. Real AdaBoost: Boosting for Credit Scorecards and Similarity to WOE Logistic Regression. 2017. Available online: https://www.semanticscholar.org/paper/Real-AdaBoost-:-boosting-for-credit-scorecardsand-Edwards-Duhon/36638aff184754db62547b75bade8fa2076b1b19 (accessed on 12 December 2023).
- 121. Schapire, R.E. Explaining AdaBoost. Available online: https://www.semanticscholar.org/paper/Explaining-AdaBoost-Schapire/ e2682f2a2752cba7a05fd3db1cb43731c1afb002 (accessed on 12 December 2023). [CrossRef]
- 122. Lodhi, E.; Wang, F.-Y.; Xiong, G.; Zhu, L.; Tamir, T.S.; Rehman, W.U.; Khan, M.A. A Novel Deep Stack-Based Ensemble Learning Approach for Fault Detection and Classification in Photovoltaic Arrays. *Remote Sens.* **2023**, *15*, 1277. [CrossRef]
- Ghoneim, S.S.M.; Rashed, A.E.; Elkalashy, N.I. Fault detection algorithms for achieving service continuity in photovoltaic farms. *Intell. Autom. Soft Comput.* 2021, 30, 467–479. [CrossRef]
- 124. Feng, D.-C.; Liu, Z.-T.; Wang, X.-D.; Chen, Y.; Chang, J.-Q.; Wei, D.-F.; Jiang, Z.-M. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Constr. Build Mater.* **2020**, 230, 117000. [CrossRef]
- 125. Ghaffarzadeh, N.; Azadian, A. A Comprehensive review and performance evaluation in solar (pv) systems fault classification and fault detection techniques. J. Sol. Energy Res. 2018, 4, 252–272.
- 126. Naveen Venkatesh, S.; Sugumaran, V. Fault diagnosis of visual faults in photovoltaic modules: A Review. *Int. J. Green Energy* **2021**, *18*, 37–50. [CrossRef]
- 127. Youssef, A.; El-Telbany, M.; Zekry, A. The role of artificial intelligence in photo-voltaic systems design and control: A review. *Renew. Sustain. Energy Rev.* **2017**, *78*, 72–79. [CrossRef]
- 128. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. Mach. Learn. 2006, 63, 3–42. [CrossRef]
- 129. Rodrigues, S.; Ramos, H.G.; Morgado-Dias, F. Machine Learning in PV Fault Detection, Diagnostics and Prognostics: A Review. June 2017. Available online: https://www.researchgate.net/publication/320353848_Machine_Learning_in_PV_Fault_Detection_ Diagnostics_and_Prognostics_A_Review (accessed on 21 December 2023).
- 130. Mishra, G.; Sehgal, D.; Valadi, J.K. Open Access Volume 13(3) Hypothesis Quantitative Structure Activity Relationship Study of the Anti-Hepatitis Peptides Employing Random Forests and Extra-Trees Regressors. 2017. Available online: https://www.bioinformation.net (accessed on 21 December 2023).
- 131. Prastiwi, D. Online shopping website analysis for marketing strategy using clickstream data and extra trees classifier algorithm. *J. Actuar. Financ. Risk Manag.* **2022**, *1*, 36–41.
- Moruff, O.A.; Bolaji, O.G.; Olufadi, H.I.; Aolat, R.G.; Buraimah, O.H.; Rilwan, D.M. A Study on Lung Cancer Identification Using Extra Trees-Based Model. Available online: https://www.researchgate.net/publication/363771547 (accessed on 21 December 2023).
- 133. Zhang, F.; Bai, J.; Li, X.; Pei, C.; Havyarimana, V. An ensemble cascading extremely randomized trees framework for short-term traffic flow prediction. *KSII Trans. Internet Inf. Syst.* **2019**, *13*, 1975–1988. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.