Heliyon 9 (2023) e21491

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Fault diagnosis of photovoltaic systems using artificial intelligence: A bibliometric approach

Edgar Hernando Sepúlveda-Oviedo [a,b,*], Louise Travé-Massuyès [a,c], Audine Subias [a], Marko Pavlov [b], Corinne Alonso [a]

[a] LAAS-CNRS, Université Fédérale de Toulouse, CNRS, UPS, INSA, Toulouse, France
[b] Feedgy, Paris, France
[c] ANITI, Université Fédérale de Toulouse, Toulouse, France

ARTICLE INFO

Keywords:
Photovoltaic (PV)
Fault diagnosis
Machine learning
Artificial intelligence
Research trends
Renewable energy

ABSTRACT

Conventional fault detection methods in photovoltaic systems face limitations when dealing with emerging monitoring systems that produce vast amounts of high-dimensional data across various domains. Accordingly, great interest appears within the international scientific community for the application of artificial intelligence methods, which are seen as a highly promising solution for effectively managing large datasets for detecting faults. In this review, more than 620 papers published since 2010 on artificial intelligence methods for detecting faults in photovoltaic systems are analyzed. To extract major research trends, in particular to detect most promising algorithms and approaches overcoming excessive time calculations, a conventional bibliographic survey would have been extremely difficult to complete. That is why this study proposes to carry out a review with an innovative approach based on a statistical method named Bibliometric and a Expert qualitative content analysis. This methodology consists of three stages. First, a collection of data from databases is carried out with all precautions to achieve a large, robust, high-quality database. Second, multiple bibliometric indicators are chosen based on the objectives to be achieved and analyzed to assess their real impact, such as the quantity and nature of publications, collaborative connections among organizations, researchers, and countries or most cited articles. Finally, the Expert qualitative content analysis carried out by experts identifies the current and emerging research topics that have the greatest impact on fault detection in photovoltaic systems using artificial intelligence.

1. Introduction

The need to reduce and replace fossil fuels has led to a significant increase in alternative energy production, coupled with a serious reduction in the cost of designing, installing, and maintaining photovoltaic (PV) systems. To be effective, these systems must be scalable, making them competitive as a mature energy solution on the market [1]. In addition, in the past few years, harnessing solar energy through photovoltaic electrical production has been acknowledged as the most efficient method for generating power from the environment [2]. Photovoltaic energy is essential to promote social transformation and long-term economic development [3]. However, despite steady progress in emerging technologies regarding efficiency [4], the continuity of production of these various

* Corresponding author.
E-mail address: ehsepulved@laas.fr (E.H. Sepúlveda-Oviedo).

https://doi.org/10.1016/j.heliyon.2023.e21491
Received 8 June 2023; Received in revised form 14 October 2023; Accepted 23 October 2023

Available online 26 October 2023

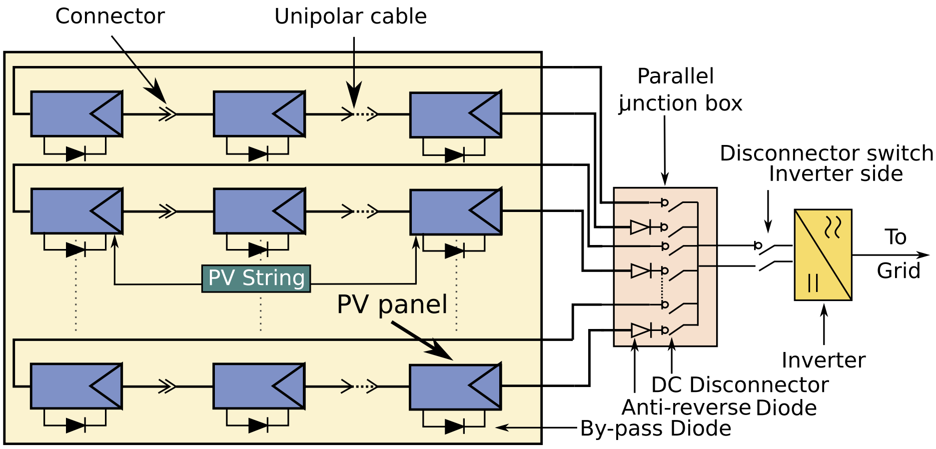

Fig. 1. A photovoltaic power plant consists of photovoltaic modules that are made up of photovoltaic cells and connected sequentially (in series) using unipolar cables to constitute photovoltaic strings. These panels or modules are equipped with secure elements located inside the junction box and power components inside the static converter.

systems encounters a major obstacle because faults are frequently present. Moreover, the time required to detect and resolve these faults can be more than one week, leading to significant production losses [5]. For example, in 2018, the loss of production due to faults and degradation was estimated at 17.5% of the total energy production [6]. In the worst case, some of these faults deteriorate photovoltaic systems at a rate of 0.8% per year, while others can even cause discontinuity in the production of a part or the entire system [7].

Like other power plants, photovoltaic systems are susceptible to faults in any of their components. The fundamental building block of a photovoltaic system is a solar cell, which serves as the active element responsible for converting sunlight directly into electricity. Solar cells have the option to be linked either in a series or in parallel with basic electrical protections such as bypass diodes to form a complete photovoltaic module. To ensure long-term functionality, solar cells are encapsulated in a secure environment with additional protection, known as a photovoltaic module [8]. Furthermore, photovoltaic modules can be interconnected to create a photovoltaic string (where modules are connected in series) or a photovoltaic array (where multiple strings are connected in parallel) [9,10]. Photovoltaic power plants are made up of one or more photovoltaic arrays, equipped with protection diodes called anti-reverse diodes and conventional electrical converters that allow connection to local grids or loads.

Finally, a complete photovoltaic system can be complex, particularly in large-scale power plants. Various architectural approaches exist for assembling arrays in parallel to increase current and reduce the number of components in power converters. Alternatively, additional power electronics can be introduced to ensure continuity of production and achieve the desired power output and reliability for photovoltaic installation [8]. These elements and their respective electrical connections can be clearly observed in Fig. 1.

Fig. 1 illustrates a photovoltaic system that corresponds to one of the most widely used configurations, called the series-parallel configuration. There are other configurations as well, such as Ridge Linked or Total Cross Tied configurations [11].

The selection of a particular configuration is influenced by a range of factors. For instance, despite the fact that the total cross-tied and bridge link configurations exhibit superior power performance compared to the series-parallel configuration, their limited economic feasibility has hindered their widespread adoption [12]. Numerous research studies suggest several comprehensive approaches to design a photovoltaic installation adapted to specific situations, showcasing the depth of knowledge in this scientific domain.

Given the complexity of the structures of the photovoltaic system, faults can be caused by various factors such as deterioration of photovoltaic cells or their environment, including cell cracks, overheating, moisture penetration, degradation of interconnections, and corrosion of the connections between cells [13–34]. Similarly, faults in other parts of photovoltaic systems can be originated by issues such as broken encapsulation, short or open circuits, problems with junction boxes and other secure functions, or faulty inverter modules [14,20,24,27,35–50]. At the array level, faults such as ground, line to line, or arc can occur, especially in photovoltaic plants with a capacity exceeding 100 kWp [51–67]. Faults also often occur in the junction box, which can be induced due to other faults such as short circuits, damage or corrosion of connections, or open circuits [20,27,40,49,68–73]. Furthermore, faults can also be attributed to diodes used as security components, either due to improper implementation (e.g., diode polarity reversal during assembly, poorly connected or omitted diodes) or electrical destruction during various operations [20,49,52,65,74–76].

Detecting faults, regardless of their nature or origin, is crucial for the preventive or corrective maintenance of photovoltaic systems [77]. However, distinguishing between an actual fault and a mere lack of production becomes challenging due to the influence of meteorological conditions [58] and the continuous changes in operating conditions, as well as the presence or absence of maximum power point tracking devices [54]. Indeed, there exist numerous scenarios in this field where the electrical behavior of faulty panels closely resembles that of panels without any faults [3,78,79].

A comprehensive analysis of fault characteristics and the significance of detecting faults at an early stage is presented in reference [80]. Currently, the choice of appropriate methods for identifying faults relies on factors such as the type of the fault (line-line, short circuit, open circuit, hot spot, partial shadow, etc.). The complexity of the problem significantly increases when multiple faults need to be monitored by the same system using the same algorithm, especially when the faults occur intermittently and manifest so rapidly that they cannot be detected [81]. Therefore, there is a need to improve existing strategies to develop more efficient systems with online predictive fault detection capabilities applicable across a broad spectrum of solar or photovoltaic systems.

Taking into account the numerous factors that influence the fault detection processes in photovoltaic (PV) systems, several authors have proposed conventional reviews as a means to understand current fault detection research in photovoltaic systems [1,37,39,45,66,69,82–93]. These reviews highlight the rapid replacement of conventional fault detection methods with artificial intelligence-based approaches. This change is attributed to the limitations of conventional methods, such as infrared/thermography [94–97], ultrasound inspection [98,99], electroluminescence imaging [100–102], and lock-in thermography [103,104]. These conventional methods struggle to handle the substantial amounts of multidomain data collected by new photovoltaic monitoring systems (including electrical, thermal, environmental, historical, and technical data). Additionally, conventional methods were designed to detect specific types of faults in photovoltaic systems, and some even require photovoltaic panels to be disconnected, halting overall photovoltaic production. Furthermore, conventional systems are challenging to reconfigure to detect other types of fault or the same fault under varying test conditions. Even conventional detection methods, such as visual inspection [98,105–107], are highly relying on human observers, making them quite subjective.

Despite offering valuable insights into the research landscape, these state of the art reviews possess notable limitations, which have motivated the objectives of this study. The main objective of this research is to create a comprehensive review that can serve as a fundamental resource for the advancement of future studies and potential policies aimed at enhancing energy efficiency. To achieve this, the study not only explores some of the most representative articles on fault diagnosis in photovoltaic systems using artificial intelligence, but also categorizes them based on the principle of the algorithm used. Additionally, this research provides a synopsis of upcoming studies, with the goal of streamlining the research process by quickly identifying major contributors, publications, journals, research facilities, and countries within the relevant field. All of these aspects accelerate the beginning of research efforts and even aid in the selection of appropriate machine learning algorithms based on specific application conditions.

Academic institutions generally use two research strategies to comprehend the latest developments: quantitative and qualitative research. Quantitative research involves the analysis of numerical data (for example, statistics), while qualitative research delves into non-numerical data, such as language, to study concepts, establish hypotheses, or interpret patterns. However, a comprehensive understanding of the current state of a given field requires the simultaneous use of both research strategies. For this reason, in order to further develop this comprehensive review of the state of the art, this study introduces an innovative hybrid methodology (combining qualitative and quantitative approaches) for analyzing and identifying promising and cutting-edge research topics in artificial intelligence for photovoltaic fault detection, with potential applicability in other domains.

The proposed methodology combines bibliometric analysis (statistical analysis interpreted as a quantitative analysis) and expert qualitative content analysis to assess the latest developments. This approach allows identification of research gaps, trends, collaboration networks, prolific authors, and notable publications in the field. It also aims to detect emerging promising topics and propose new areas of research. Furthermore, this methodology not only identifies documents that use artificial intelligence for photovoltaic fault detection, but also groups them based on similarities in the algorithms used in each research. To our understanding, this approach has never been utilized in the field of artificial intelligence to detect faults in photovoltaic systems before.

This study makes multiple contributions. **Firstly**, it introduces a hybrid methodology capable of analyzing large amounts of existing information in the literature to discover hidden relationships and new research trends objectively. **Secondly**, the proposed methodology can identify the family (analogical, connectionist, evolutionary, symbolic, regression and Bayesian) to which the machine learning algorithms used in each photovoltaic system fault diagnosis article belong. **Thirdly**, this study also presents a compilation of the most used algorithms for each machine learning family, along with the corresponding number of associated articles.

Fourthly, this study not only explores articles addressing the issue of fault detection in photovoltaic systems, but also presents the journals, authors, countries and laboratories with the greatest impact in this area. **Fifthly**, the study outlines a set of key aspects for the selection and development of efficient machine learning algorithms tailored for photovoltaic system fault diagnosis. **Sixth**, it highlights a range of aspects related to data monitoring in photovoltaic systems that significantly improve the performance of fault detection algorithms and facilitate their integration into commercial systems. Lastly, the **seventh** contribution of this study lies in demonstrating the need to construct a formal fault dictionary to solve the challenge of effective fault labeling in photovoltaic systems.

This study follows the following structure: Section 2 introduces the motivation behind this study. Section 3 presents the related work and methodology. Section 4 covers data collection. Section 5 focuses on bibliometric analysis. Section 6 presents the analysis of qualitative content from experts. Section 7 highlights promising research topics. Section 8 discusses challenges. Finally, Section 9 provides the discussion and conclusions drawn from the review analysis.

2. Motivation

As mentioned in this study, numerous review documents are available on various fault detection methods in photovoltaic systems [1,37,39,45,66,69,82–93]. However, these reviews use conventional state of the art review methodologies that have certain limitations.

Firstly, algorithms are mainly compared based on their accuracy value in fault detection. This presents an issue as it must be considered that not all machine learning algorithms are trained under the same conditions. In other words, different datasets, weather conditions, and photovoltaic technologies are used, among others. These conditions vary depending on the specific application for which machine learning algorithms are used.

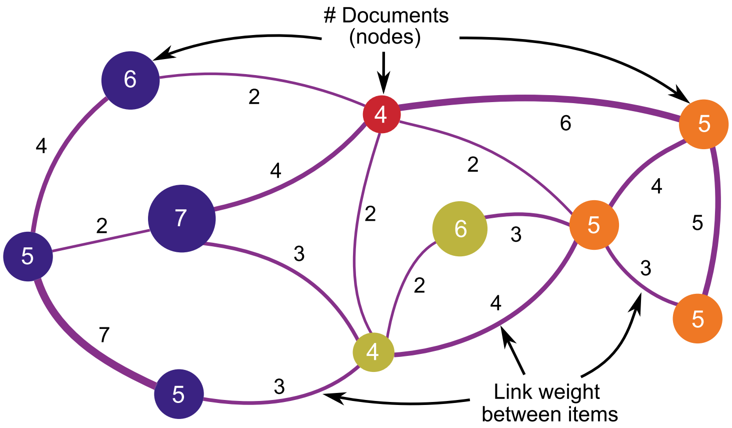

Fig. 2. Example of a bibliometric network represented graphically, consisting of various nodes and lines. Each node is assigned a number and a color, which correspond to different entities based on the type of study conducted. The number represents the size of the node, indicating the proportion of entities sharing similar attributes. Nodes with the same color are grouped together, indicating high affinities and forming clusters. The lines in the graph represent the connections between nodes. The thickness and spacing of the circles on the lines reflect the intensity of the affinity between two nodes. A thicker line and closer circles indicate a stronger connection between entities. A higher number associated with a line confirms that the links between two entities are highly intensive.

Secondly, none of the reviews emphasize evaluating algorithms separately based on the availability of fault labels. A label can be understood as a term that characterizes a specific fault behavior, such as “snail trail”, “broken glass”, “hot spot”, among others. The challenge of assigning labels for accurate fault diagnosis in energy generation systems is one of the main current challenges, as it requires multiple field tests and experts who can accurately, unambiguously, and robustly extract specific fault characteristics. It is this information that allows for accurate fault diagnosis. Not addressing this issue increases the difficulty for researchers in making decisions about the appropriate algorithm for each specific application based on the available data.

Third, subjective factors cannot be completely eliminated when selecting studies for conventional reviews, often leading to the exclusion of important studies [108]. At times, the selection process lacks transparency or authors take a subjective approach when identifying emerging research topics.

Fourth, many times, these reviews fail to provide the necessary discussion and information for academic and industry communities, which hinders the recognition of crucial authors, organizations, countries, documents, sources, and keywords. Furthermore, the timely detection of emerging research domains becomes a challenge, limiting our understanding of prior scientific endeavors and potential academic research. Moreover, the advancement of scientific development and technological innovation is also impeded.

Fifth, these studies fail to harness the full potential of artificial intelligence to handle multivariate data of high dimensions and extract hidden relationships within complex and dynamic environments such as BIG DATA [109]. The motivation behind conducting this study, which offers significant theoretical contributions to the field of fault detection in photovoltaic systems and energy management, is to attempt to provide a novel solution to the five limitations of conventional reviews, along with the observed concerns in the field.

3. Related work and methodology

As mentioned in Section 2, the methodologies have limitations due to the human capacity required to analyze numerous documents accurately and impartially. Consequently, it becomes challenging to provide a comprehensive and unbiased overview of the conceptual and intellectual framework of a scientific field, emerging research areas, or collaboration networks.

Bibliometrics, on the other hand, is a valuable tool that can address the subjectivity issues encountered in conventional state of the art reviews. Additionally, this tool enables the mapping of collaboration networks, emerging research areas, and relevant technological fields. References such as [110–113] highlight the capabilities of bibliometrics in quantitative analysis. Using computer engineering, database management, and statistical techniques, bibliometrics organizes, explores, and analyzes large volumes of information (Big Data) in a structured manner. It can accomplish tasks that are impossible to be analyzed by human capacities alone [114]. A formal definition of bibliometrics is provided in [115]. Moreover, bibliometrics facilitates the decision-making process and the prediction of emerging technologies [116–118].

The treatment of big data using Bibliometric follows a well-defined protocol. It involves translating the analysis into nodes, links between nodes, and networks with varying potential. This graphical representation is particularly suitable for engineering and research applications [119]. Visualizing the results of the analysis using bibliometric, based on the principle of nodes and networks, is called a bibliometric map or network. Fig. 2 presents an example schematic to illustrate how the results of an analysis can be represented using a Bibliometric map.

Fig. 2 illustrates an example of a network of 10 nodes. Each node represents a size and is represented by circles, with the circle size proportional to the number of entities sharing similar attributes, indicated by the value inside the node. Each cluster is denoted by a different color. The thickness of the lines connecting the nodes is proportional to the corresponding value on the line, while the spacing between the circles represents the intensity of collaboration. As mentioned in [120], this type of bibliometric analysis and

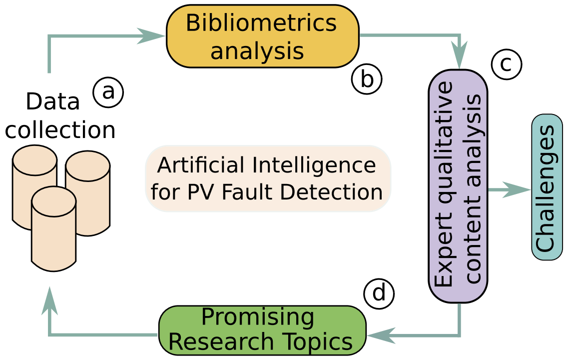

Fig. 3. 3-stage methodology. a) Data collection, ; b) Bibliometrics analysis and c) Expert qualitative content analysis) used for construction of the global review on fault detection using AI in photovoltaic systems.

visualization has become an indispensable tool for measuring scientific progress in various scientific fields, with different applications and degrees of collaboration [121–123].

Specifically, in the field of solar energy, bibliometrics is widely used in areas such as organic solar cells [118,124], energy transition [125], rooftop photovoltaic fields [108], analysis of different Maximum Power Point Tracking methods [126], and more generally [113,118,120,124,127–132]. These studies demonstrate the significant potential of bibliometrics to identify emerging research fields through statistical analysis of article keywords. However, to improve the interpretability of topics derived from bibliometric analysis, it is necessary to conduct expert-supervised qualitative content analysis [133]. Currently, these tools serve as valuable supplements to provide a more accurate and unbiased overview of the current state of research in the field. The main advantage of using these tools lies in their ability to explore and uncover hidden topics within large amounts of textual data [134–137]. Therefore, the content analysis strategy used in this study focuses on topics and involves analyzing the most relevant articles within each topic, as determined by the number of citations received by each article. This approach provides more detailed insights into each topic [138].

One of the main contributions of this research resides in its capacity to tackle the issues discussed. To achieve this, the study proposes a novel hybrid methodology that combines bibliometric analysis and expert qualitative content analysis. This methodology is implemented in several stages, as shown in Fig. 3.

As seen in Fig. 3, the proposed methodology consists of four steps or stages. The first stage is called **Data Collection**, where documents related to the studied areas are retrieved from databases such as *Scopus* and *WoS*. The aim of the initial phase is to construct a data corpus containing all the relevant information for automatically determining hidden relationships. In the second stage, known as **Bibliometric Analysis**, these documents are processed to extract relevant information, including emerging trends, highly cited documents, collaboration networks, keywords, and more. The results of the bibliographic analysis are then analyzed in the third stage, which involves an **Expert Qualitative Content Analysis**. This analysis helps in forming a comprehensive overview of the use of artificial intelligence for fault detection in photovoltaic systems and can be divided into two main parts. The first part of the results corresponds to the output of a methodology called **Promising Research Topics** in this study. This outcome highlights new research trends in fault detection using artificial intelligence. Furthermore, these results of promising research topics can be used to further apply the same proposed methodology. The second part of the **Expert Qualitative Content Analysis** is called **Challenges**. It covers the obstacles that hinder the progress of research in this field and are closely associated with the promising topics mentioned earlier.

It is important to mention that this study was guided by a simplified version of the PRISMA 2020 standard [139], which has been used in works such as [140]. In general, the PRISMA standard comprises four phases: *(i)* identification of the need for review, *(ii)* definition of a search protocol, *(iii)* performance of the review, and*(iv)* analysis of the review. The identification of the need for review was described in the Introduction. Sections *(ii)* – *(iv)* are contained within the sections **Data Collection**, **Bibliometric Analysis** and **Expert Qualitative Content Analysis**. The respective stages of the methodology and their results are presented in detail in the following.

4. Data collection stage

To ensure dependable conclusions during the construction of a wide range of state of the art [141], it is essential to have a data corpus that exhibits adequate accuracy and robustness. With this consideration in mind, the data collection stage has been conducted following the framework presented in Fig. 4.

In the studied area, the data collection stage was applied on May 03, 2022. Initially, a group of documents was retrieved from widely used databases for bibliometric studies, namely Scopus and Web of Science [142,143]. These databases provide data from studies carried out over a longer period compared to other databases [144]. To perform this initial retrieval, a search equation was constructed using the following keywords: “PV”, “Photovoltaic”, “Fault”, “Failure”, “Anomaly”, “Detection”, “Diagnosis”, “Classification”, “Artificial intelligence”, and “Machine learning”. The search equation was configured with Boolean operators, as indicated in part a of Fig. 4. Construction of the search equation using keywords is considered one of the best methods to initiate a systematic

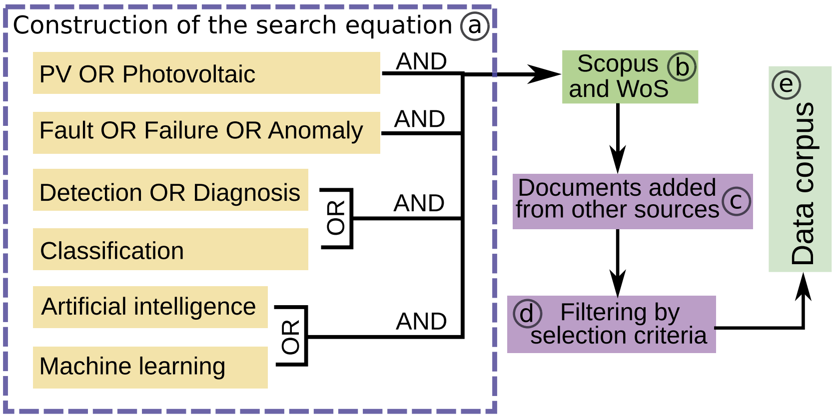

Fig. 4. Detailed proposition scheme of the Data collection stage done for construction of the data corpus. The following steps are involved: a) To build the search equation using keywords and logical operators; b) To use the search equation to retrieve the documents in the *Scopus* and *WoS* databases; c) To extend the number of retrieved documents, adding documents from other sources; d) To filter by selection criteria e) To finish the construction of the data corpus.

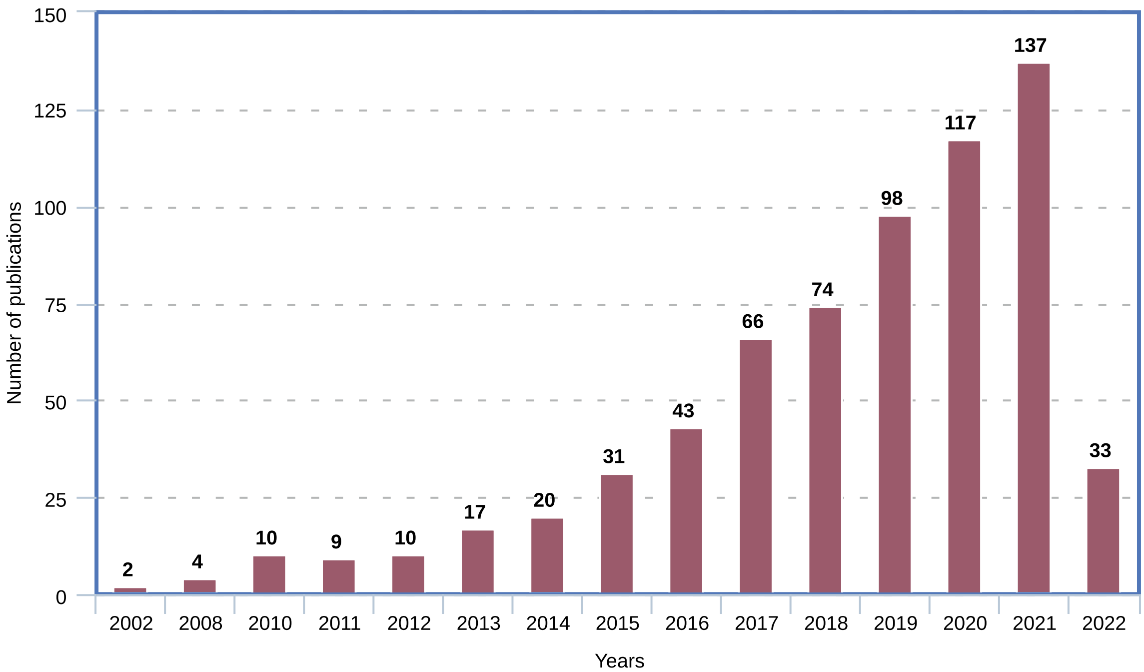

Fig. 5. Number of documents published per year in the Scopus and Web of science databases from the first record found in 2002 to 2022.

search [145]. The second step of the data collection stage involved performing the search for each document title, abstract, and keywords. As a result, an additional group of documents was retrieved from both Scopus (262 documents) and Web of Science (205 documents). Duplicate documents were eliminated using Zotero software [146]. In the third step, a set of documents from other databases was added, peer-reviewed for relevance, to the initial group, resulting in a total of 671 documents. The distribution of these documents retrieved over the years is presented in Fig. 5.

The sample concludes in May 2022, so as shown in Fig. 5, the total number of publications for that year may not provide conclusive results. As observed in the same figure, the research development in the field of “fault detection in photovoltaic systems using artificial intelligence” can be divided into two distinct periods. The first period commenced in 2002 with the initial publication focusing on fault detection of high impedance arc faults in transmission lines utilizing the wavelet transform [147]. Although documents published during this period did not use AI techniques, these studies on photovoltaic faults marked the inception of interest in fault detection in electrical generation and transmission systems, as well as the utilization of signal processing for feature extraction in fault detection. This first period can be estimated to extend until 2015 when the number of publications experienced a significant surge, reaching 4.5 times the number of publications in 2015 by 2022. This drastic increase is associated with the global

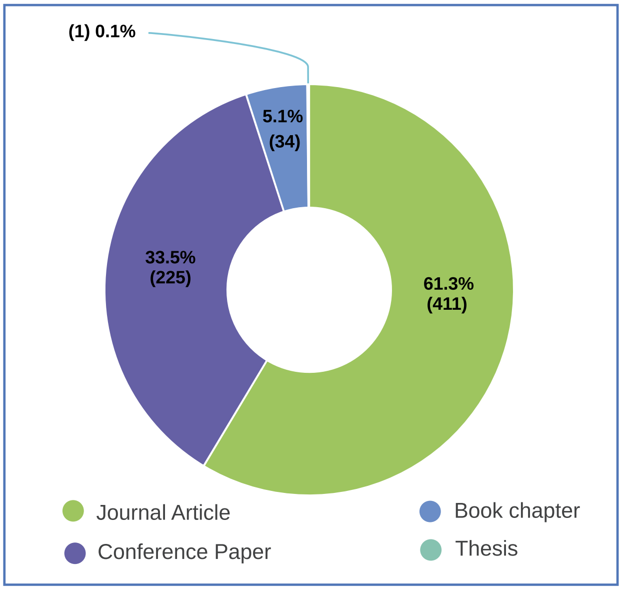

Fig. 6. Distribution of documents retrieved by type: Book, Thesis, Conference paper and journal paper.

Table 1
Criteria for inclusion of documents in the corpus data.

Inclusion Criteria
Article Published in peer Reviewed Journal
English language
Period 2010 – 2022
Conference Papers

interest in a clean energy transition and the necessity to replace fossil fuel sources to mitigate climate change issues and achieve carbon neutrality through more robust and reliable energy production systems.

Therefore, it is expected that this exponential growth will continue in the coming decades, underscoring the importance of conducting studies such as the one presented here. The distribution of the 671 documents retrieved by document type is presented in Fig. 6.

As shown in Fig. 6, most of the documents recovered consist of scientific conference papers and journal articles. While these categories contain valuable documents, it should be noted that some of them are written in languages other than English, which poses challenges in their analysis. To ensure the construction of a relevant and cohesive study corpus, this study proposes the application of a set of inclusion criteria for filtering the retrieved documents, as outlined in Table 1. The objective of this fourth step is to selectively retain documents that meet the following criteria: they must be written in English, originate from reputable scientific journals or conferences, and have been published from 2010 onward.

This filtering step is conducted based on the hypothesis that the use of a set of inclusion criteria will select only relevant, recent, peer-reviewed information with comparable results. The purpose of this additional stage is to improve the consistency and reliability of the results obtained initially through Bibliometric analysis. After applying the filter criteria, a total of 620 peer-reviewed documents are selected. These documents form a comprehensive group, which constitutes the final data corpus for this study (fifth step). The filtering process, applied to the subsequent Expert qualitative content analysis, ensures a structured, replicable, transparent, and iterative study by eliminating documents that are not relevant to the objective of the study [148]. Once the data corpus is complete, a bibliometric analysis process can be conducted.

5. Bibliometric analysis

Bibliometric analysis allows for an objective and replicable review of a data corpus. This method is used to analyze various indicators such as authors, countries, citations, and keywords, enabling the quantitative assessment of different aspects within a specific field of knowledge [149–151]. Consequently, Bibliometrics facilitates the identification of popular topics in the present, past, or future [152]. Therefore, it serves as a crucial tool to review the state of the art in a research field, covering both micro-level analysis involving individual scientists and institutes, as well as macro-level analysis on national and global scales [112,130]. However, not every database of scientific publications can be directly used for this type of bibliometric analysis. It is crucial that the databases are structured and devoid of duplicated publications, variations in author name spellings, and ideally include fields such as article title, authors, journal, publication year, citation count, institutions, countries, keywords, and bibliographic references. The results of Bibliometric analysis can be visualized through maps or networks. In this context, the closeness between two elements signifies the intensity of their connection. A reduced distance signifies a more robust relationship. In these maps, the elements are unevenly distributed in space, which facilitates the assignment of labels. An example of such representation is presented in Fig. 2. There are two main approaches for creating Bibliometric maps: distance-based methods and graphic-based methods.

Different distance-based map construction techniques can be found in the literature. One of the most widely utilized methods within the field of bibliometrics is introduced in [153]. This technique suggests using multidimensional scaling entities to represent proximities between objects within a geometric space of a few dimensions [154]. Alternatively, techniques such as visualization of similarities (VOS) viewer are also proposed [155–157], which have shown better results compared to multidimensional scaling [158]. Another distance-based technique, called $VxOrd$ is proposed in [159,160], and has demonstrated high performance in the construction of maps with large numbers of elements (7000 or more). Finally, a new method implemented in a software called BibTechMon is proposed in [161]. It is difficult to determine the advantages or disadvantages between them, it is necessary to mention that their biggest difference is the visual complexity to interpret the results. Therefore, depending on the number of elements (articles, authors, countries) some of these approaches may be more or less adequate.

Similarly, several works have been published on the elaboration of graph-based maps. In [162], the authors propose constructing graphs using the Euclidean distance between nodes or vertices as the “graph-theoretical” distance. The vertices are then connected by springs of calculated distances, and the arrangement of vertices is determined on the basis of the total elastic energy of the system. Another technique proposed in [163] aims to achieve uniform edge lengths by modifying the spring embedding model using an analogy with forces in natural systems. Some authors [164,165] combine the work presented in [162] with the pathfinder network technique [166]. There are also software versions such as CiteSpace [167], specialized in this type of maps, and software such as Pajek from [168], which combines distance-based and graph-based approaches.

For the bibliometric analysis conducted in the photovoltaic research area, a distance-based bibliometric mapping was performed using VOSviewer, a bibliometric visualization and mapping software developed by Nees Jan Van Eck and Ludo Waltman.[1] VOSviewer was chosen for its interesting distance-based bibliometric results [158] and its widespread use in the field of bibliometrics [169–172]. Bibliometric maps generated using VOSviewer offer a comprehensive means of assessing relationships and their strengths among distinct entities (authors, laboratories, countries, etc.).

To perform the bibliometric analysis, three steps are involved. First, the similarity matrix based on the co-occurrence matrix is computed. Then, a map is created using the VOS mapping technique applied to the similarity matrix. In the end, the map undergoes translation, rotation, and reflection to guarantee uniform outcomes. The similarity matrix is calculated by normalizing the co-occurrence matrix. The choice of normalization metric depends on the total number of occurrences or cooccurrences of elements. To achieve correct normalization, a suitable similarity metric must be selected. The most common normalization metrics include cosine similarity [173] and Jaccard index [174]. However, VOSviewer employs a similarity metric known as the strength of association [156,157], the proximity index [175,176], or the probabilistic affinity index [177], due to its advantages over other similarity measures [178]. In these cases, the similarity α_{mn} between two elements m and n is defined in Equation (1):

$$\alpha_{mn} = \frac{\varphi_{mn}}{\Omega_m \Omega_n} \tag{1}$$

where φ_{mn} denotes the count of occurrences where items m and n appear together. Meanwhile, Ω_m and Ω_n stand for the overall frequency of individual occurrences or the combined occurrences of items m and n, assuming that the occurrences of these items are statistically uncorrelated.

After the creation of the similarity matrix, the VOS mapping technique is implemented. It is worth mentioning that the VOS technique can yield one or multiple solutions of global optimum. In cases where multiple solutions exist, it is important to use translation, rotation, or reflection operations to ensure that consistent results are generated by VOSviewer. The procedure for creating a network entails using input data in the format of bibliographic database files, such as those from sources such as Web of Science, Scopus, Dimensions, Lens, and PubMed [179]. Additionally, reference manager files, which cover formats such as RIS, EndNote, and RefWorks, can also serve as input for the network creation process in VOSviewer [179]. Other formats can be computationally adjusted to be comprehensible by VOSviewer. Further details about this method can be found in [153,179].

In bibliometrics, there are multiple aspects that can be useful for the quantitative and objective evaluation of the impactful input of a researcher or a research team. According to the work presented in [180], these aspects can be categorized into groups such as: scientific activity indicators cover indicators related to scientific production and influence, collaborative efforts, dissemination and production. However, the most commonly used indicators are those of scientific production and collaboration [181].

In the area of scientific production metrics, several key indicators can be identified, including the “total number of publications”, the “Crown indicator” and the “H-index” and their variations. The total number of publications mainly takes into account the large number of articles generated. However, it has the drawback of not taking into account factors such as the number of authors, the inclusion of articles in indexed journals, the number of citations received or the specific nature of the articles.

The Crown indicator, also named field-normalized citation score [182], seeks to address variations in citation numbers across scientific areas and years of publication [183]. However, a major challenge lies in the inappropriate categorization of articles within specific fields, which often occurs when articles published in general journals are misclassified, thus not contributing to the index

[1] See Vosviewer.

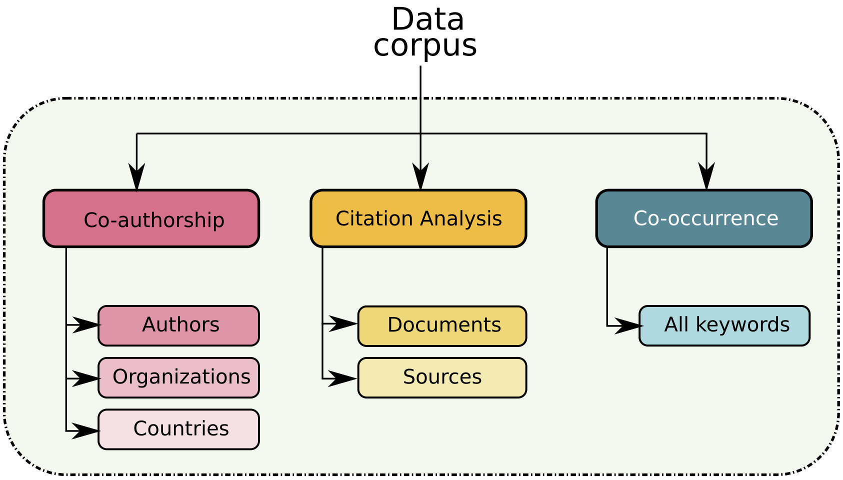

Fig. 7. Bibliometric analysis based on Co-authorship, Citation analysis and Co-occurrence. Each section is analyzed on the units exposed in the diagram (Authors, Organizations, Countries, etc).

calculation. This indicator is also affected by the size of the research group, since the larger the group, the greater the number of group publications. Therefore, some studies recommend its use only when comparing research teams with an equivalent quantity of participants [181].

The "H-index", which is the third index, functions as a measure of productivity for authors, departments, research groups, or universities operating in the same or closely related research areas within a given country [184]. It is worth noting that the "H-index" should be employed in conjunction with other metrics to obtain a comprehensive understanding of research progress in a particular field, due to its two main limitations. First, the "H-index" is time-dependent and does not provide real-time information about the current production of an author or dynamic information about their productivity. Second, when multiple authors participate, it is difficult to accurately assess their individual contributions to each article, since all authors receive the same credit.

As mentioned in [182], various other bibliometric metrics are available for examining scientific output, including the g-index, h(2)-index, contemporary h-index, and more. These metrics or indicators can be employed to assess authors, institutions, or countries. What all of those metrics share is the requirement for complementary metrics to ensure an accurate and comprehensive understanding of the findings.

Collaboration indicators assess researcher interaction, with calculations possible at individual, institutional, or national levels. Despite numerous studies on motivations and barriers to inter institutional collaboration, the current literature lacks comprehensive insights into participation preferences of individual researchers in various collaborative projects [185]. Some hypotheses suggest that the prominence of these collaboration indicators arises from the need to address complex scientific problems or innovative technologies through joint efforts involving researchers, institutions, organizations, and diverse communities. Collaboration occurs across disciplinary, interdisciplinary, multidisciplinary, transdisciplinary, and national/international levels [186]. Collaborative research offers benefits such as cross-disciplinary idea exchange, skill acquisition, funding access, improved result quality, and transformative outcomes. These advantages significantly drive collaborative growth and industry support for academia [187,188].

In more recent times, thanks to the emergence of advanced computational tools, the utilization of keyword-based co-occurrence analysis has become increasingly prominent. Many authors advocate the use of keyword analysis to expose the relative popularity of research topics within a particular field [189–191]. For this study, and without loss of generality, the bibliometric analysis is conducted on three major categories that have been previously used in the literature and have shown high performance in identifying the state of the art in a research area [192–195]. The three categories are: *(i)* Co-authorship, *(ii)* Citation analysis, and *(iii)* Co-occurrence, as illustrated in Fig. 7.

The study of co-authorship, citation analysis, and co-occurrence networks has a rich history of yielding fascinating results in the field of Bibliometric. The origins of these studies can be linked to the 1960s and 1970s [196–198].

In reference to the analysis of citations, it is important to distinguish between self-citations and citations from other studies that are not directly related to the cited document. In this study, a deliberate decision was made to differentiate between these two types of citations and exclude self-citations from the citation analysis.

The bibliographical analysis in this study commences with an exploration of co-authorship.

5.1. Co-authorship

In this part of the analysis, specific choices were made regarding co-authorship. Firstly, self-citations were not taken into account, and secondly, publications with more than 12 authors were excluded. This decision was based on the observation that publications with many authors often represent comprehensive reviews of the state of the art or serve as compilations of previously published

Table 2
Synthesis of co-authoring Bibliometric analysis showing links between authors and their number of published documents. A distinction is made between authors to detect those who have a high level of citation. Second column presents the number of authors with 0 or more citations analyzed and represents all authors identified. Third column presents the results after applying a filter of at least 10 citations per author.

Minimum # of Publications	# Authors (Citations $\geq$ 0)	# Authors (Citations $\geq$ 10)
1	620	127
2	96	39
3	27	21
4	10	10
5	3	3
6	5	5
9	3	3
10	1	1
11	1	1
12	2	2

works. These publications may contain general information rather than specific scientific contributions. Thus, these choices were made to reduce redundancy in information and exclude documents with less scientific value.

Moreover, in the area of bibliometrics, there exists a significant challenge when it comes to attributing co-authored publications to individual authors. Various strategies have been suggested to address this issue [199,200]. Among these strategies, two are methods particularly popular: the full counting [201] and the fractional counting [202,203]. These techniques are commonly employed in the creation of bibliometric networks [201]. To demonstrate the disparity between these two approaches when applied to a bibliometric co-authoring network, an example is provided.

Consider an article that has been published with six authors. The complete counting method gives each connection between the six coauthors a weight of one, indicating an equal contribution from each. In contrast, the fractional counting method assigns a weight of 1/6 to each connection, reflecting the proportional contribution of each coauthor to the publication.

In a co-authoring bibliometric network, the thickness of the connections (links) between two nodes, representing authors, is determined by the co-authoring link value. The size of every node is dictated by the quantity of documents authored by the individual author. In this review, all analyses were conducted using the complete counting method. However, it is important to note that the VOSviewer software provides both complete and fractional counting methods [153]. After analyzing the data corpus, a total of 804 distinct authors who have published at least one article related to the topic of interest were identified, following the extraction of documents involving 12 authors or more. The information presented in Table 2 outlines the correlation between the quantity of authors and the quantity of research papers published in the domain of fault detection employing artificial intelligence in photovoltaic power generation systems.

As illustrated in the co-authoring bibliometric analysis shown in Table 2, there is a notable decline in the number of authors. For instance, in the field of photovoltaics, the author count decreases from 620 individuals with a single publication to 96 authors with two publications, and further diminishes to 27 authors with three publications. Furthermore, the impact of an additional analysis that enforces a minimum requirement of 10 citations is evident in the third column of the same table. This requirement highlights the most scientifically significant papers, resulting in only 10 authors with more than four publications and over 10 citations. It is worth noting that, among these authors with four or more publications, the number of authors remains consistent after filtering, indicating that they also have 10 or more citations within the period between 2010 and 2022.

Importantly, a pretreatment analysis was conducted when creating the data corpus to address instances in which the authors signed their documents with different initials. This analysis is crucial to avoid errors that may arise from considering such cases as separate authors by the software.

In order to create and assess co-authorship networks, we established a minimum criterion of two publications for each author, leading to a foundation of 96 authors. To delve deeper into meaningful connections, partnerships, or clusters of individuals highlighted by the collective dynamics of the networks, we computed the cumulative strength of co-authorship connections with other authors. The top 10 authors, identified by their substantial citation impact or the number of connections, are presented in Table 3.

Another piece of information comes to light with this bibliometric approach, which was able to identify, as shown in Table 3, that the four authors, namely Chen Z, Cheng S, Lin P, and Wu L, have an equal number of publications (6) and also share the highest number of citations (373). This suggests that these four authors have a strong connection and generally collaborate closely.

In order to further analyze this, a bibliometric map was constructed to highlight important collaborative networks among authors. The bibliometric map analysis, as depicted in Fig. 8, demonstrates that only 24 authors have participated in long-term collaborations. Furthermore, the figure illustrates the existence of four separate clusters that are interconnected, signifying four networks and their corresponding collaborative efforts. It is worth noting that the close association observed among the first four authors in Table 3 forms a significant cluster (shown in blue in the figure), validating their strong co-authorship links and their collaborations with the wider scientific community.

The authors excluded from Fig. 8 are likely those who have referenced other authors but have only one publication within the specific field under study. Indeed, each author represented in this map must have more than 2 published articles. This visual data

Table 3
Top 10 of the most influential authors on the topic of interest mentioned with the number of their published documents, citations and/or collaborations. This information is translated in this table through the strength of co-authorship links of each one corresponding to the third column.

Author	Documents	Citations	Link strength
Chen z.	6	373	24
Cheng s.	6	373	24
Lin p.	6	373	24
Wu l.	6	373	24
Zhao y.	3	206	8
Dhimish m.	4	171	6
Spanias a.	12	127	55
Holmes v.	2	125	4
Rao s.	11	125	52
Tepedelenlioglu c.	10	121	47

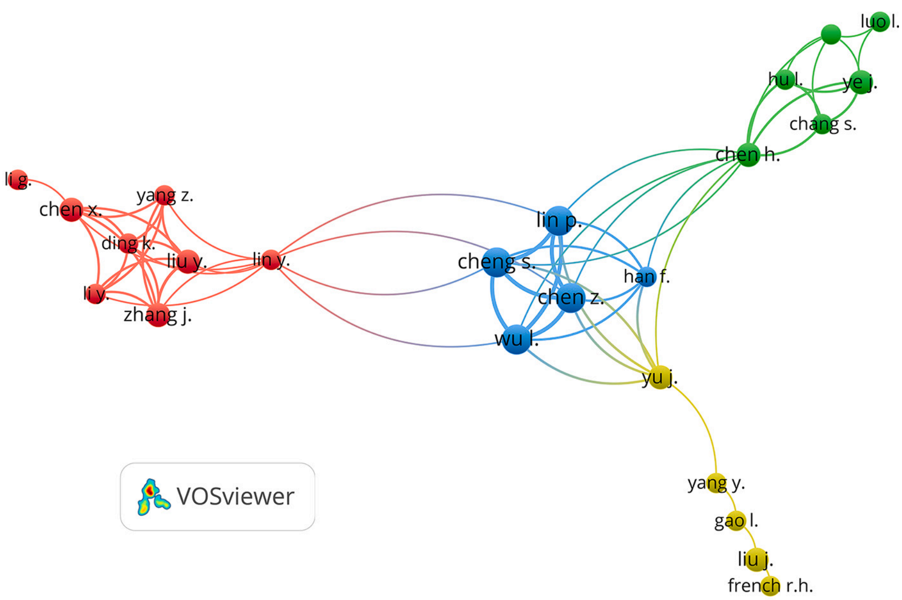

Fig. 8. Network of co-authorship constructed using authors who had contributed to more than two papers in the studied area emphasizing 4 clusters and their links of collaborative works.

is valuable for researchers and academics exploring fault detection in photovoltaic systems with artificial intelligence, offering a distinct overview of key authors in this domain and the interconnections depicted through citations.

Likewise, it is interesting to examine the establishments associated with the most influential authors in the field of interest and the connections among them. Table 4 provides a concise overview of the institutions that receive the highest number of citations, each having published at least two documents on the subject of utilizing artificial intelligence for fault detection in photovoltaic systems.

The Jiangsu Collaborative Innovation Center of Photovoltaic Science and Engineering holds the top position in the rankings with 177 publications, as depicted in Table 4. Arizona State University comes in second with 129 publications, followed by the College of Physics and Information Engineering with 89 publications. These three institutions occupy the highest positions on the list. Notably, the same table reveals that both Germany and the United States each have three institutions actively engaged in research on fault detection, showcasing their significant interest in this field of study. Additionally, it is worth mentioning that two branches of Texas A&M University located in Qatar are also included in the table.

The examination of these data results in the creation of the following question: "How can a country enhance its research and development framework to reinforce its competitiveness?" Is it based on a 'distributed' approach with multiple laboratories conducting research in parallel or on a centralized approach with a small number of research centers focusing on numerous specialized studies in the same research topic?"

As stated in reference [204], the formulation of such inquiries holds important significance, especially in the area of renewable energy, for shaping future research and development policies. Confronting the formidable challenges within this domain is imperative. Table 5 offers an examination of the quantity of research laboratories in relation to the volume of published papers and their

Table 4
21 most productive institutions with 2 or more documents published in research on fault detection in photovoltaic systems using artificial intelligence.

Institutions	Country	Papers	Citations	Link strength
Jiangsu collaborative innovation center of photovoltaic science and engineering	China	2	177	0
Arizona state university, sensip center	United States	5	129	0
College of physics and information engineering	China	2	86	0
Electrical and computer engineering program Texas a&m university	Qatar	3	58	5
Chemical engineering program Texas a&m university	Qatar	2	57	5
Prince Sultan University	Saudi Arabia	3	50	5
University m hamed bougara of boumerdes	Algeria	2	42	4
As-international center of theoretical physics	Italy	2	29	4
Cyprus university of technology	Cyprus	2	29	4
Cyprus academy of sciences, letters and arts	Cyprus	2	29	4
National engineering school of monastir	Tunisia	2	29	3
Pi photovoltaik-institut Berlin ag	Germany	2	18	0
University of Colorado Boulder	United states	2	18	0
Helmholtz-institute erlangen-nuremberg for renewable energies	Germany	2	9	2
Universität erlangen-nürnberg (fau)	Germany	2	9	2
Kuwait college of science and technology	Kuwait	2	5	2
Advance power electronics research lab	India	2	4	2
Computer science department	Saudi Arabia	2	4	2
Jamia millia islamia	India	3	4	0
University of kairouan	Tunisia	2	1	1
Sandia national laboratories	United States	2	1	0

Table 5
Number of Institutions based on the quantity of published documents. First, the number of institutions with 0 or more citations is analyzed, and then a filter of at least 10 citations is applied.

Minimum # of Publications	# institutions (Citations $\geq$ 0)	# institutions (Citations $\geq$ 10)
1	333	102
2	20	11
3	3	1
5	1	1

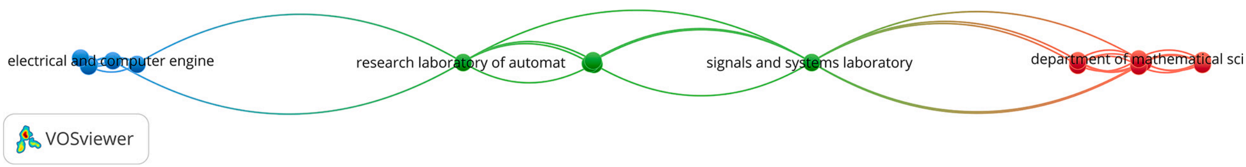

Fig. 9. Visualization of co-authorship networks built from institutions who had produced more than two papers highlighting 3 main clusters and existing links of collaborative work between institutions.

citations. The data reveals a preference for an approach centered on "small, decentralized research units," except for three specific institutions.

In Table 5, discrimination is performed between institutions based on their citation count, distinguishing those with 0 or more citations from those with 10 or more citations. Interestingly, among the total of 333 institutes, only 102 of them have more than 10 citations. It is important to note that the number of institutions drastically decreases from 333 with one publication to 20 with two publications, and further decreases to 3 with three publications. Only one institution is associated with five publications. This observation can be attributed to the fact that the application of machine learning or artificial intelligence techniques to fault detection in renewable energy is a relatively new field, and thus there is limited experience in this area. The results obtained after applying the filter of 10 citations per institution are also presented. The impact of this filtering supports the hypothesis that there is insufficient time between two recent papers for the formation of a collaboration network, especially when the time span is less than three years.

To identify the main interactions between institutions, a bibliographic map was constructed and its result is illustrated in Fig. 9. This map offers a graphical depiction of the connections between 17 institutions with at least two publications, categorizing them into three separate clusters.

As shown in Fig. 9, each node corresponds to an institution from Table 4. However, Fig. 9 indicates that not all institutions are present. Further analysis reveals that this refers to the first three institutions listed in Table 5. The absence of connections with other top-classified institutions may imply their reluctance to collaborate or their citations being considered as self-citations.

Regarding the remaining institutions listed in Table 4, 17 of them can be grouped into 3 clusters based on existing collaborative efforts. This analysis confirms the existence of two types of strategy used by academic institutions. The first type involves conducting studies without collaborations, which is likely the case for larger institutions. The second type, commonly seen in most organizations,

Table 6
Countries linked to international publications on photovoltaic fault detection using AI techniques and associated results obtained by filtering criteria. First corresponds to the number of all countries analyzed, and the second is done after filtering to selected only countries with more than 10 citations.

Minimum # of Publications	# countries (Citations ≥ 0)	# countries (Citations ≥ 10)
1	67	33
2	35	28
3	31	27
4	25	23
5	20	19
6	19	19
7	16	16
8	13	13
9	11	11
11	8	8
12	7	7
15	6	6
16	5	5
17	4	4
20	3	3
40	2	2

leans toward cooperative efforts, perhaps due to the wide range of expertise spanning multiple domains required to detect faults in PV systems through artificial intelligence (as indicated by the size of the node) or the need to unify research initiatives.

The connections between organizations, as indicated by citation links, reveal the interplay of diverse areas of expertise and interests. For example, in *cluster 2*, there are notable collaborations between the chemical engineering department and the electrical and computer engineering department, both affiliated with Texas A&M University in Qatar. They also foster partnerships with other institutions like the Signals and Systems Laboratory of the Institute of Electrical and Electronics Engineering at the University M Hamed Bougara of Boumerdes, Algeria, the Department of Electronic and Communications Engineering at the Kuwait College of Science and Technology, and the research laboratory specializing in automation in signal processing and image at the national engineering school of Monastir, Tunisia. It is worth noting that most collaborations involving these institutions occur with institutions from other countries, indicating that the field of fault detection is of international significance.

In *clusters 1* and *3*, there are also robust collaborations between institutions from countries such as Tunisia, Saudi Arabia, Qatar, and Kuwait. This observation suggests that conducting exploratory bibliometric analyses at the country level could yield interesting insights. Table 6 displays the count of countries determined from published documents.

Table 6 illustrates a significant decrease in the quantity of countries with at least one publication, which stands at 67, as well as the respective figures for countries with 2 publications (35) and 3 publications (11). It becomes evident that, when the requirement of a minimum of 10 citations per country is introduced, the number of countries with at least one publication diminishes to just 33, and for three documents, it drops even further to 27. The impact on the remaining countries is relatively minor; for example, the number of countries with five documents reduces from 20 to 19. Concurrently, an analysis is conducted to identify the most influential countries in the field of research. To complement this analysis, Table 7 provides a list of the top 20 countries with the highest research productivity in the domain of fault detection in photovoltaic systems using artificial intelligence, considering countries that have published five or more documents.

The data presented in Table 7 reveals that China, India, and the United States are the top three countries contributing the highest volume of research in the specified field, with 40, 40, and 39 publications, respectively. In these countries, during the period covered by this review, there is not only an accelerated technological process that has led them to align their government policies towards increasing the use of renewable energies, such as photovoltaics, but also the rapid development of artificial intelligence, promoted in parallel by other industrial sectors.

In Table 7, the first three ranks in citations are held by the United States and China, the latter being the country with the highest photovoltaic solar energy capacity [108]), and France, with 538, 515, and 229 citations, respectively. It can also be observed that despite a high number of documents, India has a lower level of citations. This can be compared to France, which has a relatively high level of citations with a smaller number of publications. France has approximately 6.6 times fewer documents than China or the USA but an excellent ratio of $229/8 = 28.2$ citations per document, compared to $515/40 = 12.9$ for China and $538/39 = 13.8$ for the US.

This difference can be attributed to the fact that in France, many scientific studies are carried out through collaborative approaches involving both academia and industry in long-term projects. It is also linked to the publication culture practiced by French researchers, who often choose to publish in international journals with high impact factors.

For various reasons, countries such as China, the USA, and France can be considered as target countries for creating collaborative networks in the studied area. To understand the collaboration networks involving the 20 countries listed in Table 7, a bibliometric map by countries was created using VOSviewer, as shown in Fig. 10. The overall robustness of a country-to-country relationship depends on the quantity of research papers jointly authored by two or more countries [128]. As mentioned before, these bibliometric maps extract valuable information and assist researchers in positioning their own work and identifying potential collaborators [205].

Table 7
The most 20 productive countries identified with 5 documents or more published in terms of research on fault detection in photovoltaic systems using artificial intelligence.

Country	Documents	Citations	Link strength
China	40	515	13
India	40	165	17
United States	39	538	24
Algeria	19	197	24
Germany	16	78	19
Italy	15	83	13
Saudi Arabia	14	75	22
Iran	11	124	17
Cyprus	10	110	17
Qatar	10	85	18
United Kingdom	10	184	9
Brazil	8	44	3
Tunisia	8	45	15
Australia	7	79	7
Denmark	7	77	11
Spain	7	84	7
France	6	229	8
Malaysia	6	46	5
South Korea	6	90	4
Taiwan	5	8	0

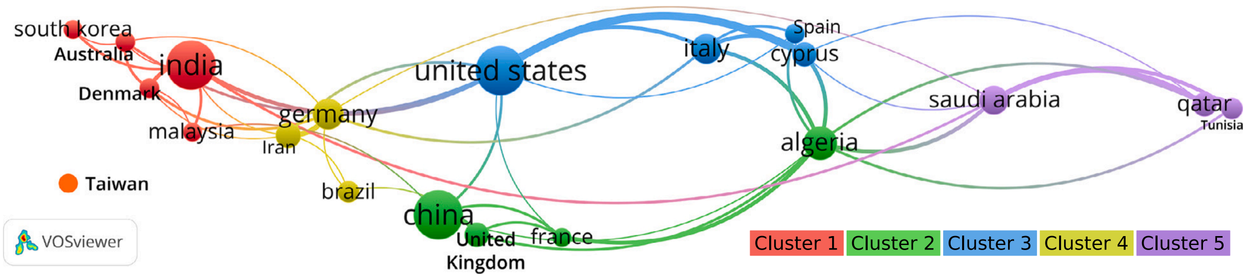

Fig. 10. VOSviewer network visualization map of country co-authorship and international collaboration identified of the countries in Table 7; The countries are grouped into 5 clusters.

In Fig. 10, the five distinct clusters have been established by analyzing the occurrence frequency of common terms within the documents of each country. These clusters are defined by the frequency at which shared terms appear together in documents from two or more interconnected countries. For example, the strength of the link that signifies collaboration between the United States and Cyprus is 6, and this strength is directly proportional to the thickness of the connecting line. In contrast, the link between the United States and France has a value of only 1, indicating a less robust level of collaboration.

Fig. 10 also shows that these interrelationships between countries result in the formation of five clusters, with one country having no collaboration.

Cluster 1, highlighted in red, is led by India, which has 40 documents and a total link force of 17 with other countries. This cluster collaborates with countries such as Australia, Denmark, Malaysia, and South Korea. Recent works conducted by these countries involve fault detection using methods such as locally weighted learning, k-nearest neighbor, nearest neighbor, and K-star algorithm applied to images captured by aerial vehicles [206]. These techniques use advanced technology levels and are relatively recent.

Furthermore, studies utilizing techniques such as categorical boosting (CatBoost), extreme gradient boosting (XGBoost), light gradient boosting method (LGBM) [207], Isolation Forest, Facebook-Prophet, and AutoEncoder Long Short-Term Memory (AE-LSTM) [208] on electrical signals are also explored. Additionally, works based on neural networks during specific seasons of the year [209], or in combination with fuzzy systems [209,210], are studied. Most of these documents are recent, which may explain their lower citation levels despite their scientific originality and relevance.

Cluster 2, visualized in green, is led by China with 40 documents and a total link force of 13. Cluster 2 also includes Algeria, France, and the United Kingdom. These countries are working diligently on the exploration of neural networks as a relevant tool for fault detection. Specifically, they have recently published works that use neural networks with decentralized federated approaches [211], stacked auto-encoder [212], deep learning on luminescence images [213], among others [214].

Cluster 3, represented in blue, is led by the United States with 39 documents and a total link force of 24. This cluster also includes Cyprus, Italy, and Spain. Similar to *Cluster 2*, these countries are focusing their efforts on exploring neural networks, either in the context of images or in analyzing the characteristic electric curves of photovoltaic systems. Their recent publications include works on these topics [215–218].

Table 8
Citation analysis applied on documents issued of the corpus data.

Citations of a document	# of Publications
1	159
5	80
20	32
50	8
100	4
123	3

Cluster 4, shown in yellow, is led by Germany with 16 documents and a total link force of 19. This cluster also includes Brazil and Iran. These countries are exploring methods such as a Fine Tree Classifier [219] and combinations of artificial neural networks with Sugeno Fuzzy Logic [220]. The fuzzy logic models they use allow for the simultaneous handling of numerical data and linguistic knowledge through a unified mathematics [221]. Additionally, they are working on methods involving weather-corrected indexes, temperature-corrected equations, one-class Support Vector Machine (SVM) methods, kernel techniques, and fault detection in photovoltaic plants [222]. Recent works also integrate fault detection and classification with electrical and environmental condition monitoring systems [223], as well as use a ResNet-34 convolutional neural network in conjunction with a supervised contrastive loss function [224].

Cluster 5, depicted in violet, is led by Saudi Arabia, contributing 14 documents and a cumulative link force of 22. This cluster 5 also covers Qatar and Tunisia. These countries have been actively collaborating on the advancement of diverse fault detection and categorization algorithms in photovoltaic systems. Some of the methods they have explored include an Artificial Neural Network (ANN) [225], an interval kernel PCA (IKPCA)-based Ensemble Learning (EL) in micro networks [226], a decision tree-based algorithm [226], probabilistic neural networks in photovoltaic arrays [214], and machine learning techniques centered around kernels, such as Support Vector Regression (SVR) and Gaussian Process Regression (GPR) [227].

Finally, among the countries mentioned, Taiwan is unique in having a minimum of 5 published documents and 8 citations, yet it lacks collaborative ties with other countries. It is important to note, though, that Taiwan shares a common interest with the other countries in the exploration of diverse neural network algorithms.

Some recent studies conducted in Taiwan have focused on comparing the performance of convolutional neural networks (CNN) when utilizing an enhanced gamma correction function and a threshold function on infrared thermographic images of solar modules [228]. Another interesting paper delves into both direct current (DC) and alternating current (AC) signals within the photovoltaic system, transforming them into 3D images through the Gramian Angular Field transform (GAF) for signal preprocessing [229]. The results achieved through these methods outperform those of traditional machine learning techniques like k-nearest neighbor, Random Forest, Decision Tree, and Support Vector Machine.

Furthermore, in Taiwan, research has also been conducted on modifications of conventional methods such as Random Forest and the independent component analysis (ICA) method [230]. These modified methods achieved performances of 99.88% and 99.43% in the proposed scenarios. These situations arise from instances involving imbalanced datasets, where one class contains a significantly larger number of observations than the other classes.

After analyzing the co-authorship patterns in publications in this field, an analysis of citations based on documents and sources (journals) is carried out.

5.2. Citation analysis

Bibliometric analysis has the capability to quantitatively synthesize a research topic and provide information on the main trends in research through the examination of citations [121]. Consequently, a comprehensive citation analysis is conducted to gain a deeper comprehension of the evolution of research in the subject of interest. The mapping of citations via bibliometrics can be performed using various analytical units, such as documents, sources, authors, establishments, or countries. However, this study exclusively concentrates on documents and sources.

Regarding the quantity of documents, there are 159 publications that have received at least one citation. Furthermore, 80 publications have obtained five or more citations, 32 publications have acquired 20 or more citations, and finally, only three documents have accumulated 123 citations or more. These statistics are detailed in Table 8.

The associated bibliographic map is created by considering only the publications with 20 or more citations, as indicated in Table 8, resulting in a selection of 32 publications. Among this collection of documents, just 18 of them exhibit interconnections, indicating that they reference at least one of the other publications on the list. The connections between these documents are represented as links between nodes in Fig. 11.

The analysis described and its graphical representation are relevant because they provide insights into the interconnections among the most significant publications, as evaluated by their total number of citations. Fig. 11 illustrates these connections, where a strong linkage between nodes indicates that highly cited documents reference each other.

In the VOSviewer-generated citation visualization map, each node is equipped with a hyperlink leading to the corresponding online repository of the article. An examination of the data reveals that the articles with the highest number of citations, in decreasing

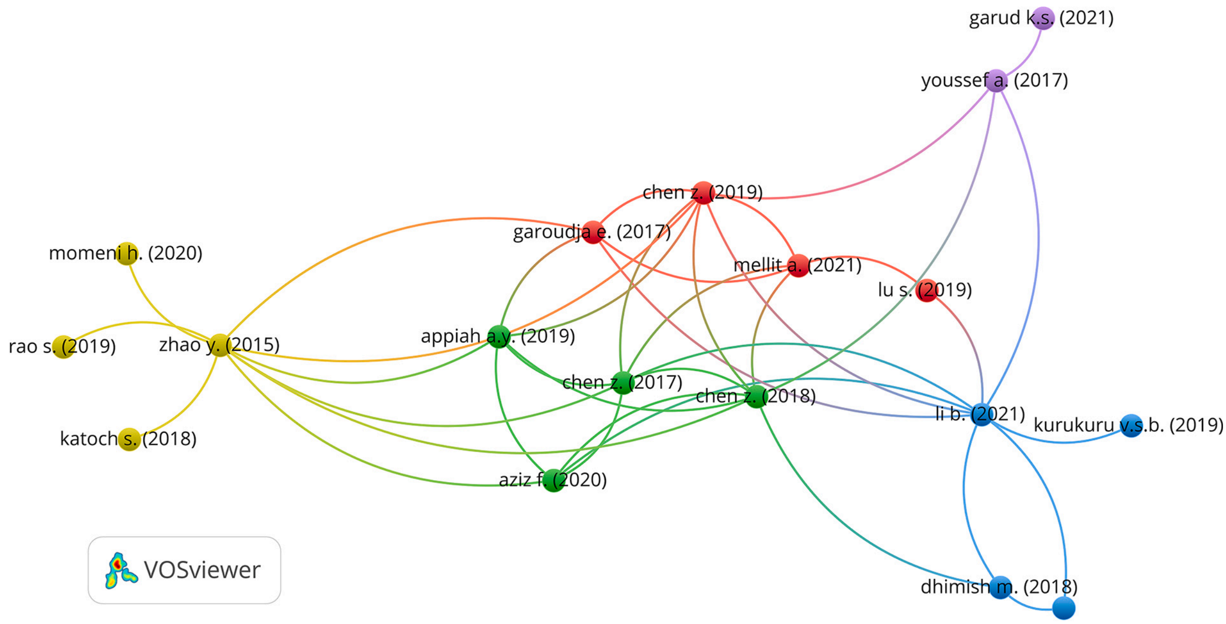

Fig. 11. Citation network applied to publications with more than 20 citations in photovoltaic systems with 6 clusters representing the links of collaborative work between the publications.

Table 9
Synthesis of the 13 most impactful documents retrieved from the corpus data classified since their number of citations.

Title	Year	Journal	Cited by
Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays	2015	IEEE Transactions on Power Electronics	186
Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics	2017	Applied Energy	180
Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection	2018	Renewable Energy	127
Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents	2018	Energy Conversion and Management	110
An enhanced machine learning based approach for failures detection and diagnosis of PV systems	2017	Energy Conversion and Management	78
Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions	2019	Energy Conversion and Management	75
Support Vector Machine-Based Islanding and Grid Fault Detection in Active Distribution Networks	2020	IEEE Trans. Emerg. Sel.	46
DA-DCGAN: An Effective Methodology for DC Series Arc Fault Diagnosis in Photovoltaic Systems	2019	IEEE Access	46
A Novel Convolutional Neural Network-Based Approach for Fault Classification in Photovoltaic Arrays	2020	IEEE Access	46
Long short-Term memory networks based automatic feature extraction for photovoltaic array fault diagnosis	2019	IEEE Access	42
Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters	2020	Renewable Energy	37
Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems	2019	Solar Energy	33
A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography	2020	Solar Energy	28

order, are Zhao (2015) [59] with 186 citations, Chen (2017) [231] with 180 citations, Dhimish (2018) [221] with 127 citations, and Chen (2018) [232] with 110 citations. For a comprehensive list of the top 13 most cited articles derived from the dataset, please consult Table 9.

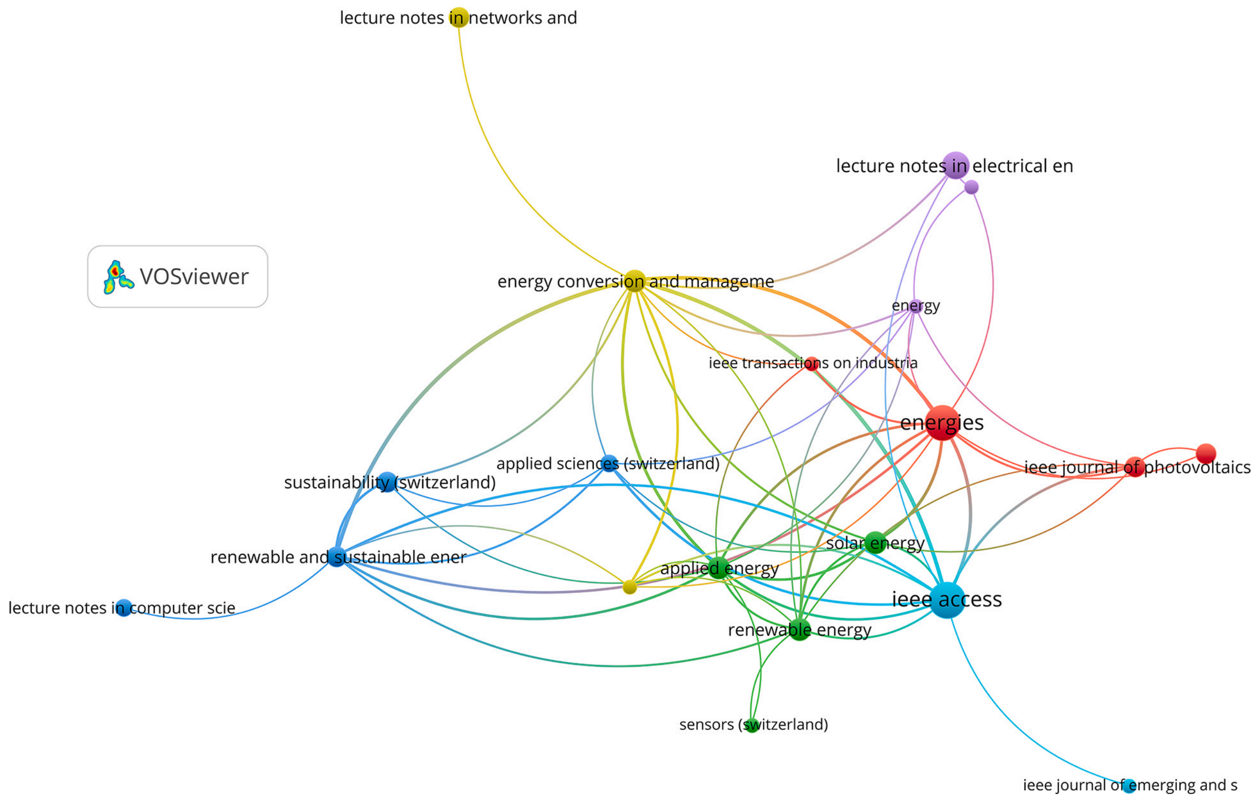

Fig. 12. Bibliometric map for the most frequent sources of scientific journals. 6 clusters are identified through the links of collaborative work between the publications editors.

In a complementary manner, it is also interesting to conduct an analysis based on sources. These sources refer to indexed journals obtained from the corpus of data under study, resulting in a total of 170 different sources. Fig. 12 illustrates the network mapping associated with the 41 sources that share two or more links. However, the largest connected element set represented in the network consists of 21 elements. Like the previous maps, this representation includes information regarding the significance of each node, which represents a scientific journal, along with its corresponding number of publications. The connections between the journals are shown by links.

This type of analysis and visual depiction holds significance as it allows to gain insights into the most frequently referenced sources (research papers or journal publications) and the potential interconnections between these sources via citations in papers published in other journals.

This result enables the academic community to identify the main sources of publications with significant impacts, leading to potentially more citations. It helps to promote the latest innovative approaches, compare and position research with these references, increase visibility, and improve the possibility of disseminating new results in the research field.

This analysis focuses only on sources from documents published in peer-reviewed journals obtained after applying filters. In order to categorize these resources, the commonly acknowledged metric known as the "H-index" can be employed. The "H-index" serves as a measure to assess the scholarly output of an author or journal by considering the quantity and quality of citations their publications receive. This metric takes into account both the volume of publications and the frequency of citations. Essentially, the "H-index" represents the number "h" of documents that have been cited at least "h" times each [233].

However, to provide more relevant information, the H-index can be complemented with other indicators, particularly those dedicated to evaluating the quality of journals in specific fields. For example, it can be supplemented by the well-known "Impact Factor" or a new indicator proposed by Scimago called the "SJR" [234]. The SJR indicator is associated with a quartile (Q) that categorizes journals into four categories. The SJR indicator improves the H-index by incorporating two variables to assess the quality of journal. The first variable is the number of citations, which is modulated based on the origin of each citation. Citations from journals considered as references in the main field of the source are deemed more pertinent than others. A clustering method known as the "authority" principle is used to classify documents within each quartile (Q), illustrating the distribution of themes across different categories. If a journal receives major citations from other journals categorized within the same field, its quartile (Q) will be higher. Q1 represents the highest quality level of a journal, while Q4 represents the lowest. It is important to note that some journals can cover different fields. Thus, in the Scimago platform, these scientific sources have different quartiles associated with each thematic area. The values described in this research pertain specifically to the application of artificial intelligence in the

Table 10
Classification of the scientific sources based on the indicators $H-index$, SJR, quartile, and number of publications in fault detection applied to photovoltaic systems.

Journal	H-Index	SJR	Quartile	Papers
Renewable and Sustainable Energy Reviews	337	3,68	Q1	11
IEEE Transactions on Industrial Electronics	304	3,52	Q1	4
IEEE Transactions on Power Electronics	285	3,34	Q1	9
Applied Energy	235	3,06	Q1	7
Energy	212	2,04	Q1	12
Energy Conversion and Management	210	2,83	Q1	19
Renewable Energy	210	1,88	Q1	17
Sensors	196	0,8	Q2	4
Solar Energy	194	1,42	Q1	41
IEEE Access	158	0,93	Q1	21
International Journal of Electrical Power & Energy Systems	142	1,54	Q1	5
Progress in Photovoltaics: Research and Applications	137	1,98	Q1	4
IEEE Transactions on Sustainable Energy	134	4,16	Q1	4
Energies	111	0,65	Q2	24
Sustainability	109	0,66	Q2	5
IET Renewable Power Generation	84	1,13	Q2	9
IEEE Journal of Photovoltaics	78	1,06	Q1	17
Applied Sciences (Switzerland)	75	0,51	Q2	7
International Journal of Photoenergy	56	0,41	Q3	8
Taiyangneng Xuebao/Acta Energiae Solaris Sinica	27	0,2	Q4	4

identification of faults within photovoltaic systems. In turn, these results are associated with the use and management of renewable energies.

Using these indicators, namely the H-index and the SJR provided by Scimago on May 11, 2022, the quantity of documents released within the field of interest totals 170 out of a grand total of 409 documents. Table 10 presents the 20 sources with four or more published documents on the subject of interest.

Table 10 summarizes the most important journals on fault detection applied to renewable energy based on the amount of publications in the field, it can be observed that Solar Energy stands out with 41 publications, almost twice as many as Energies (24). Following closely are IEEE Access (21) and Energy Conversion and Management (19). Another viewpoint, based on the H-Index, suggests that Renewable and Sustainable Energy Reviews is the most noteworthy journal in this field, with the highest value of 337. This can be compared to the high H-Index values of other internationally recognized peer-reviewed journals in the field, such as IEEE Transactions on Industrial Electronics, IEEE Transactions on Power Electronics, Applied Energy, Energy, Energy Conversion and Management, and Renewable Energy.

Furthermore, considering other indicators, specifically the SJR index associated with the quartiles Q1, Q2, Q3, and Q4, approximately 65% of the sources mentioned in Table 10 belong to the highest category, Q1. Additionally, these sources possess an SJR value exceeding 1, signifying significant citations from influential sources. Taking all these factors into account, it can be deduced that the most valuable articles or references regarding the application of artificial intelligence in detecting faults in photovoltaic systems are determined not just by the volume of publications but also by their impact factor and the sources from which they are cited.

After obtaining a comprehensive understanding through coauthorship and citation analysis, an examination of the existing associations among keywords is carried out using a metric known as cooccurrence. As discussed in the work of Radhakrishnan et al. [235], the goal of analyzing a keyword co-occurrence network is to delve into the components and knowledge framework of a scientific domain by examining the relationships between keywords in prior research. In simpler terms, the cooccurrence metric considers that two or more concepts (keywords) co-occur when they frequently appear together in a set of documents, although it rarely appears independently in other documents. TThis new exploration aims to improve understanding and knowledge of the main research topics pertaining to fault detection in photovoltaic systems and how documents can be categorized into distinct clusters or trends based on their affinities with AI algorithms.

5.3. Co-occurrence

In the field of bibliometrics, an examination of co-occurrence was performed by utilizing a matrix of keywords that appear together. The objective of this analysis was to reveal potential relationships between two elements found within bibliographic works that are present in the same work [236]. The methods for carrying out co-occurrence analysis in bibliometrics have matured over the past two decades [237,238]. Consequently, some researchers have introduced strategies aimed at creating novel approaches for analyzing Medical Subject Heading (MeSH) terms sourced from the PubMed literature database [239]. This approach is commonly referred to as MeSH term co-occurrence analysis. The typical workflow for MeSH term co-occurrence analysis involves: *(i)* extracting MeSH term data from the PubMed database; *(ii)* constructing a co-occurrence matrix; and *(iii)* performing downstream analysis on

Table 11
Occurrence of keywords in the corpus data.

Minimum # of Occurrences	# Keywords
1	1922
5	135
10	56
25	23
50	20
75	4
100	3

Table 12
Top list of the 11 most used keywords retrieved from documents on fault detection in photovoltaic systems using artificial intelligence.

Keyword	Occurrences
Neural networks	48
Support vector machines	47
Deep learning	43
Decision trees	33
Convolutional neural networks	31
Fuzzy logic	30
Anomaly detection	24
Random Forest	22
Feature extraction	22
Nearest neighbor search	16
Internet of things	12

the matrix [236,240–242]. However, despite the widespread utilization of this method, it possesses a notable limitation related to the selection of a critical threshold for eliminating low-frequency elements or reducing the dimensionality of the co-occurrence matrix.

Alternative methods propose: *(i)* employing the Rapid Automatic Keyword Extraction (RAKE) algorithm for keyword extraction from document abstracts; *(ii)* constructing a co-occurrence matrix; *(iii)* calculating the "semantic similarity" between each keyword in matrix form, which is then combined with the co-occurrence matrix of extracted keywords from the document corpus to generate a weighted co-occurrence matrix [243]. The results are then represented graphically through a Keyword Co-occurrence Network (KCN). Alternatively, other authors suggest keyword extraction and matrix construction through a graph-based ranking algorithm called TextRank [244]. These approaches for keyword extraction and visualization have already been automated and standardized in software tools such as Ucinet,[2] Pajek,[3] or VOSviewer. In this study, VOSviewer software was employed because of its extensive adoption for conducting keyword co-occurrence analysis in diverse areas of expertise [245–252]. Co-occurrence analysis was conducted for all keywords using the same tools mentioned earlier (VOSviewer), specifically utilizing a full counting method. Table 11 displays the initial findings of this analysis, highlighting the connection between occurrence count and the number of keywords in photovoltaic systems-related fields.

As shown in Table 11, there are a total of 1922 distinct keywords that appear in the comprehensive data corpus consisting of 620 peer-reviewed articles. On average, each document contains approximately 3 keywords (1922/620 = 3.1). These initial findings suggest that only a select few keyword groups exhibit a significant frequency of occurrence. Subsequently, an additional study was conducted to identify the most commonly utilized keywords within the field of photovoltaic research. The aim of this study was to accurately position the research and enhance its visibility and dissemination.

Upon analyzing Table 11, it was found that 20 different keywords appeared in at least 50 documents, with only three keywords appearing in 100 corpus publications. Therefore, it was decided to focus additional attention on a detailed analysis highlighting the 11 most commonly used keywords, as shown in Table 12. Before commencing this analysis, it should be noted that the selected words are combined with their synonyms to ensure that two keywords with high relevance are not treated separately.

The main keywords are presented in Table 12 with their corresponding number of occurrences. It was found that some AI techniques are currently being used as scientific approaches with positive results in detecting faults in photovoltaic systems. Among the most frequent keywords we can observe three related to neural networks called "Neural networks", "Convolutional neural networks", and "Deep learning" and three related to conventional methods, namely, "Support Vector Machines", "Decision Trees" and "Random Forest". Two bibliometric maps were created to gain a deeper understanding of key terms, their groupings/associations, and their importance over time. To simplify these maps, only keywords with a minimum occurrence of 5 were included.

[2] UCINET, developed by Steve Borgatti, Martin Everett, and Lin Freeman, is a social network analysis program. The free trial version can be downloaded here.

[3] Pajek is available for noncommercial which can be downloaded here.

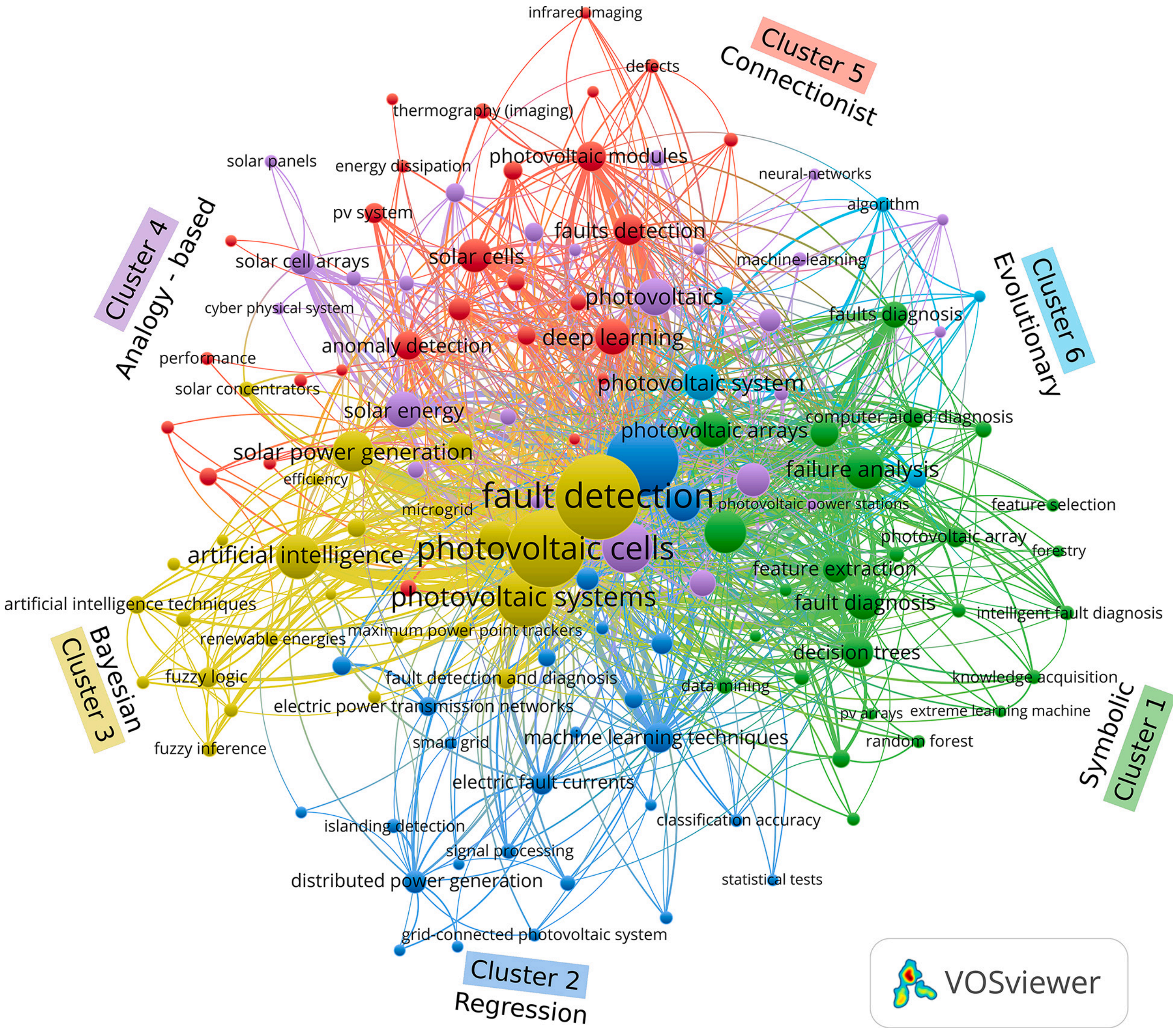

Fig. 13. Visualization Map of co-occurring keywords identifying 6 main clusters. Each cluster corresponds to a type of approach.

The first bibliometric map reveals the associations between the keywords. It identifies around 137 keywords linked to six significant keyword clusters with the purpose of determining global associations between the articles. The degree to which keywords appear together (co-occurrence) is influenced by how close they are to each other and how similar they are [253]. To extract the scientific areas involved, each cluster is classified according to the classification proposed in [254]. Fig. 13 depicts a Cluster Visualization Map that showcases the keywords that occur together within six simultaneous keyword groups. The clusters are categorized based on the respective methods they are associated with.

Cluster 1, labeled the “Symbolic approach” and represented in green in Fig. 13, consists of 26 keywords. A significant portion of these keywords is related to classification techniques, such as “decision trees”, “random forest”, “regression tree”, “C4.5 decision tree”, “gradient boosting”, and “binary tree bagging”, “Adaboost”, “Rotation forest”, “clustering hierarchical” and “fuzzy logic”. This cluster focuses primarily on analyzing the behavior of photovoltaic arrays at various scales with an interest in “feature extraction” and “feature selection”. Some works in this cluster also explore hybrid methodologies, utilizing algorithms such as the “extreme learning machine”. In summary, Cluster 1 can be considered a group of interest for photovoltaic fault detection using symbolic methods based on master algorithms.

Cluster 2, labeled as the “Regression approach” in blue, also consists of 26 keywords. This cluster seems to concentrate on detecting and classifying faults in different systems, including “electric inverters”, “distributed power generation systems”, “electric power transmission systems”, “grid-connected PV systems” and “smart grid”. The detection methods represented in this cluster include terms such as “linear model”, “linear regression”, “Logistic regression”, “principal component analysis”, “islanding detection”, and “regression analysis”. The techniques used in this cluster aim to exploit the richness of electrical signals by performing statistical and signal processing analyses, including "signal processing" based on methods such as time-frequency decomposition using "wavelet transforms" (with 19 occurrences). This cluster can be regarded as photovoltaic fault detection using a regression approach.

Cluster 3, labeled as the "Bayesian approach" in yellow, comprises 25 keywords. Some of the terms found in this cluster include "Gaussian process", "graph-based semi-supervised learning", "naïve Bayes classification", and "Bayesian belief network". Certain hybrid algorithms identified in this cluster also overlap with those found in Cluster 1, and they are linked to keywords such as "fuzzy inference", "fuzzy logic", and "fuzzy neural networks". Many works in this cluster are applied to improve reliability in microgrids. Hence, this cluster can be considered as a group focused on photovoltaic fault detection using a Bayesian approach.

Cluster 4, labeled "Analogy-based approach" in purple, consists of 25 keywords. The documents associated with this cluster place considerable emphasis on the quality and handling of the data. Terms such as "data acquisition", "data handling", and "digital storage" are highlighted in this context. The descriptions for implementing these treatments in physical systems are associated with terms such as "cyber-physical system", "embedded systems", and "internet of things". In terms of algorithms used, the works explore approaches such as

"support vector machine", "local outlier factor", "polynomial SVM", "linear SVM", "support vector regression", "radial basis function", "k-nearest neighbor", and "k-means". Some documents in this cluster combine these algorithms with methods such as "neural networks" or use ensemble learning. The works in this cluster are oriented toward "cost-effectiveness", "monitoring", and "operation and maintenance". The cluster covers a wide range of photovoltaic system scales, including terms such as "photovoltaic panels", "photovoltaic power stations", and "solar cell arrays". Thus, this cluster can be categorized as photovoltaic fault detection using an analogy-based approach.

Cluster 5, labeled as the "Connectionist approach" in red, consists of 28 keywords. Terms such as "deep learning", "convolutional neural network", "backpropagation neural networks", "artificial neural network", "extreme learning machine", "long short-term memory networks", "extension", "recurrent", "bagging", "bootstrap-based", "dynamic recurrent", "deep belief network", "Multilayer neural networks", "Multilayer perceptron", "probabilistic neural networks", "autoencoder", "entropy", and "Online Sequential" are found. Additionally, new approaches linked to terms such as "transfer learning" are discovered, where models are built based on pre-trained models, reducing development time by making a few modifications. These approaches are highly efficient in terms of computational time. Most of the works are related to image analysis, which is why terms such as "image classification", "image processing", "infrared imaging", and "thermography" are found at different scales. This is due to the presence of keywords such as "photovoltaic modules", "photovoltaic plant", and "solar cells". This cluster can be referred to as photovoltaic fault detection using the Connectionist approach.

Lastly, *Cluster 6*, labeled as the "Evolutionary approach" in light blue, consists of only 7 keywords. In this cluster, keywords related to algorithms such as "Ant colony optimization", "Genetic Algorithm", and "Particle Swarm Optimization" are found. It can be observed that this cluster shows an interest in analyzing the "accuracy evaluation" of the algorithm in different elements or, more generally, in a "photovoltaic system". This cluster can be referred to as photovoltaic fault detection using the Evolutionary approach.

As expected, the findings from the analysis of keyword co-occurrence indicate that studies focused on fault detection in photovoltaic systems using artificial intelligence primarily concentrate on analyzing fault detection. It can be inferred that the authors address various scales and technologies of photovoltaic systems and use multiple methods covering both signal processing and artificial intelligence techniques. Six distinct scientific approaches have been identified. However, in order to fully comprehend the recent new objective of research, the insights obtained from the initial co-occurrence analysis can be supplemented by a second analysis conducted using a new bibliometric map, which covers only the last five years. This second analysis is shown in Fig. 14. By incorporating this additional analysis, the visualization network depicting the overlapping the co-occurrence of keywords can be highlighted in Fig. 14.

In addition, Fig. 14 illustrates the evolution of research topics with respect to concurrent keywords during the past five years (2018 - 2022). Several documents indicate that this new analysis emphasizes research into fault detection using artificial intelligence and its recent rapid evolution. This point can be compared to the previous period when many papers focused on fuzzy systems and signal processing until 2018. A transition period (2018 - 2020) can be observed, during which research was supported by issues of signature extraction, monitoring, and some traditional machine learning techniques like random forests, neural networks or decision trees. More recently, since 2020, the authors have primarily focused their efforts on research oriented towards the utilization of neural networks, particularly convolutional neural networks and deep learning.

It is worth mentioning that in-depth studies of specific time periods reveal a notable rise in publications focused on identifying anomalies. This is likely due to the trend toward the high complexity of training systems designed for fault classification in photovoltaic systems. Such systems require previous work to build a robust database and subsequently train artificial intelligence systems under optimal conditions. The effectiveness of the outcomes and their capacity to detect every potential fault in a photovoltaic system is greatly influenced by the quality and amount of accessible data. For these reasons, the main objectives of the authors are aimed at improving the capacity to evaluate the healthy behavior of the photovoltaic power plant and determining when the behavior deviates from the healthy baseline, which is referred to as abnormal or anomaly behavior. Additionally, an important trend has emerged recently, with a shift in research interest from fault detection based only on electrical signals to hybrid detection methods that incorporate image analysis (thermal, infrared, etc.), thereby significantly expanding the spectrum of fault detection and its complexity.

In summary, this approach using keywords and their occurrences can be considered a valuable tool to aid in understanding different trends and why certain machine learning algorithms are more appropriate than others for detection of faults in photovoltaic power plants. However, a more detailed analysis of the content of the documents within each cluster was performed using an expert qualitative content analysis.

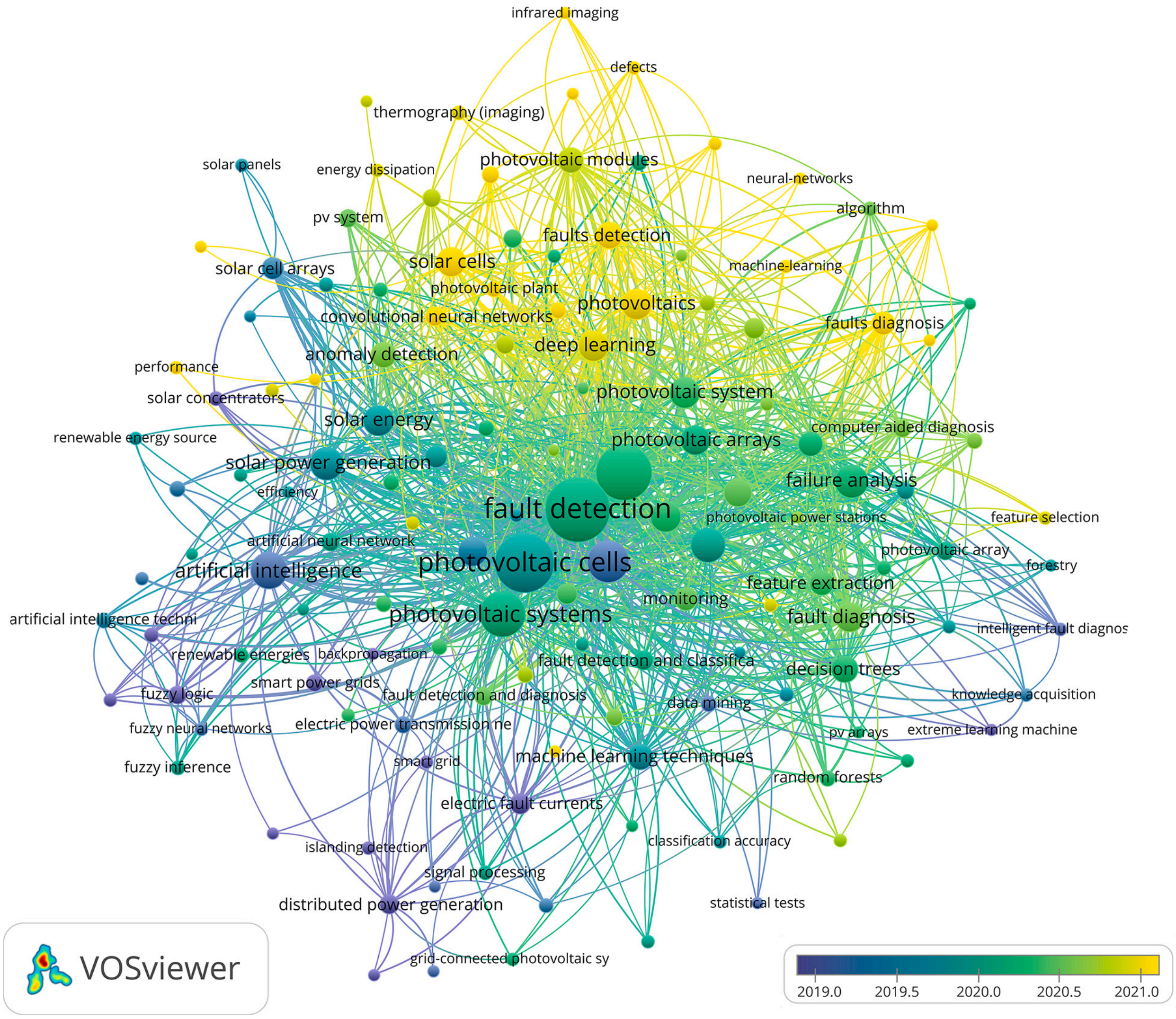

Fig. 14. Overlaying visualization network for co-occurring keywords of the 6 clusters of Fig. 13.

6. Expert qualitative content analysis

To complete the construction of the state of the art, this section exposes a content analysis of each cluster obtained in Section 5.3. In order to streamline this analysis, only the three most relevant articles are considered, with preference given to those that are highly cited, assuming that they best explain and exemplify the purpose of each cluster. The algorithms presented in each article were thoroughly analyzed to ensure that they truly belong to the identified cluster and represent its general trends.

Initially, this analysis revealed that all the identified clusters align with the classification presented by Rodrigues et al. [255]. This classification is further supported by the main contributions of the algorithms found within each group, which were extracted using the “master algorithm” tool. The notion of the “master algorithm” was first presented by Domingos [254] in his book entitled “The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World”. The author regards the “master algorithm” as a unifying factor in machine learning, postulating that it is a hypothetical algorithm capable of comprehending the behavior of any system perfectly.

Considering the perspectives of these two authors, the clusters are assigned the following names: “Symbolic approach”, “Regression approach”, “Bayesian approach”, “Analogy-based approach”, “Connectionist approach”, and “Evolutionary approach”, respectively. This study affirms that each cluster has its own master algorithm with its unique properties and limitations.

6.1. Cluster 1 - symbolic approach

This cluster groups algorithms whose master algorithm is identified as inverse deduction, also known as induction. The approaches in this cluster aim to reach specific conclusions based on preexisting knowledge that has been previously learned [254]. Due to

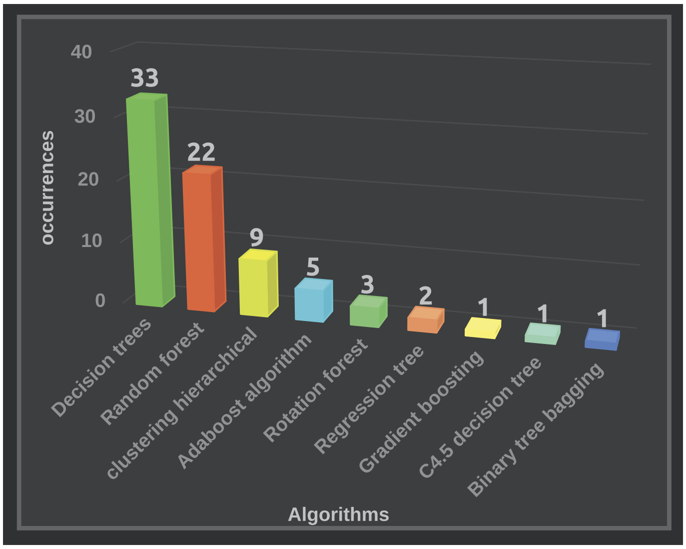

Fig. 15. List of the most used machine learning algorithms in Cluster 1.

the computational expense of inverse deduction, the application of these algorithms on massive datasets has been challenging to generalize. Some well-known algorithms in this cluster include Decision Trees, Random Forests, and Fuzzy Logic [256].

Several works propose the use of these algorithms, also referred to as symbolist algorithms, for fault detection in photovoltaic systems. The most cited article by Chen et al. (2018) [232], with 110 citations, addresses the issues of overfitting and limited performance in supervised fault classification algorithms trained on fault data samples. The authors focus is on identifying and diagnosing faults within photovoltaic panels, such as line-to-line faults, degradation, open circuits, and partial shading. They accomplish this by monitoring the real-time operating voltage and string currents of a laboratory photovoltaic system, and then comparing these values with data from a simulated photovoltaic system in Simulink. In their work, Chen et al. (2018) [232] introduce a classifier known as the Random Forest (RF) ensemble learning algorithm. They employ a grid search method to fine-tune the parameters of the RF model, with the goal of minimizing out-of-bag error estimation. This approach offers several advantages, including enhanced classifier performance and seamless integration with real-time monitoring systems. The method proposed in this study achieves a remarkable level of performance, with a 99.240% accuracy rate for fault detection and a 99.135% accuracy rate for fault classification.

In a study conducted by Dhibi et al. (2020) [257], which has been cited 39 times, two enhanced Random Forest (RF) classifiers were introduced. These classifiers were designed to tackle issues like the lack of correlation between variables in RF. They are known as the Reduced Kernel RF based on Euclidean distance ($RK - RF_{ED}$) and the Reduced Kernel RF based on K-means clustering ($RK - RF_{Kmeans}$). Both of these classifiers utilize real-system data, specifically static information from the process data. These methods follow a two-stage process: first, feature extraction and selection, followed by fault classification. $RK - RF_{ED}$ uses the Euclidean distance as the dissimilarity metric for feature extraction and selection, employing principal component analysis (PCA) for feature selection. This algorithm achieved a 100% precision rate in both the training and testing phases. On the other hand, $RK - RF_{Kmeans}$ reduces the training data volume by applying K-means clustering and PCA. This method also demonstrated a precision rate of 100% in both training and testing phases. The primary focus of these approaches is to reduce the dimensionality of databases while preserving relevant information. These classifiers were assessed using data from an emulated grid-connected photovoltaic system, with the results showcasing high classification accuracy and efficient computation.

Finally, a study by Madani et al. (2012) [258], cited 19 times, introduces an intelligent-based islanding detection algorithm for photovoltaic and Doubly Fed Induction Generator (DFIG) units. Initially, the Decision Tree algorithm is employed for islanding detection, thanks to its low computational cost and interpretability. The authors then introduce a boosting technique known as AdaBoost to enhance the performance of the Decision Tree classifier. AdaBoost iteratively trains weak or base classifiers, giving more importance to previously misclassified data to create a more robust classifier. This method was tested on a distribution system that includes photovoltaic, DFIG, and synchronous generators. For islanding detection features, a set of 16 first- and second-order derivatives of current and voltage signals is used. The results show a precision rate of 100% for AdaBoost and 92.5% for the Decision Tree algorithms.

In Fig. 15, the algorithms detected within this cluster 1 are displayed along with their respective occurrence counts.

As depicted in Fig. 15, the Cluster 1 primarily employs decision tree and random forest algorithms, which are the most commonly utilized methods in this context. However, it is evident that due to the complexity of acquiring reliable fault labels, the usage of unsupervised algorithms, such as hierarchical clustering, is on the rise.

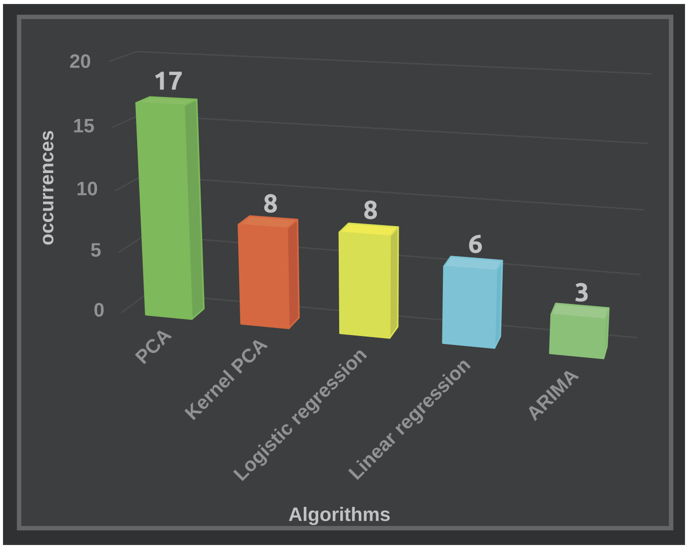

Fig. 16. List of the most used machine learning algorithms within Cluster 2.

6.2. Cluster 2 - regression approach

This paragraph discusses a cluster of algorithms that are categorized according to their data characteristics. These algorithms are not learning algorithms; instead, they are regressive algorithms that depend on past data [254]. Some well-known algorithms in this cluster include ARIMA, Linear Regression models, Principal Component Analysis (PCA), and statistical machine learning approaches [256].

Several studies suggest utilizing regression techniques for fault detection in photovoltaic systems within this particular group. The most cited article [259] in this context, with 33 references, combines machine learning with statistical hypothesis testing to enhance the detection of faults in photovoltaic systems. They employ Gaussian process regression (GPR) in conjunction with a generalized likelihood ratio test diagram. The study assesses the proposed method using both actual and simulated photovoltaic systems, taking into account parameters such as current, voltage, and power. The authors demonstrate the efficacy of their approach through an analysis of computation time, Missed Detection Rate (MDR = 7.14%), and False Alarm Rate (FAR = 6.83%) [259].

The second article, authored by Haiji et al. in 2021 [260], with 21 references, introduces a two-stage fault detection process that begins with feature extraction and selection as the initial step. They employ the PCA algorithm to reduce the dimensionality of the dataset and then input the resulting PCA data into six supervised machine learning classifiers, including Support Vector Machines (SVM), Naive Bayes (NB), Decision Tree, Discriminant Analysis (DA), K-Nearest Neighbors (kNN), and Random Forest (RF). The proposed method is assessed using data from a real grid-connected photovoltaic (GCPV) system [260]. The author provides precision ranges for three algorithms: K-Nearest Neighbors (KNN) with a range of 98.21% to 99.97%, Random Forest (RF) with a range of 99.61% to 100%, and Support Vector Machine (SVM) with a range of 97.73% to 99.97%.

Finally, Hopwood et al. (2020) [261] also propose an article with 3 citations. This article proposes the use of Principal Component Analysis (PCA) to reduce the size of the initial data set and generate uncorrelated variables, thus improving the interpretability. The study focuses on static current-voltage characteristics (IV) and uses a random forest classifier on the reduced data. The authors demonstrate that in the worst case presented, the use of PCA improves classification accuracy up to 99% compared to 98% without PCA. Additionally, the team explores different approaches to extract information from the IV curve, which aids in fault classification, especially for faults with subtle changes in the IV curves [261].

In general, these articles highlight the application of regression approaches, such as ARIMA, Linear Regression, PCA, and statistical machine learning, in fault detection for photovoltaic systems. They demonstrate the effectiveness of these techniques through various evaluation metrics and real-world data.

In Fig. 16, the algorithms detected within this cluster are shown along with their respective occurrence counts.

As seen in Fig. 16, the algorithms found within this cluster can be grouped into two main categories without loss of generality: Dimensionality reducers and regressions.

6.3. Cluster 3 - Bayesian approach

This cluster groups algorithms identified by the master algorithm tool as probabilistic inference algorithms. These algorithms share the common property of reducing uncertainty in new knowledge by utilizing the probabilistic event inference algorithm [255]. They are designed to recognize the inherent uncertainty and incompleteness present in all types of knowledge. In these algorithms, a probability is assigned to each known event. If the data supports a hypothesis, the hypothesis gains more weight. On the contrary, if

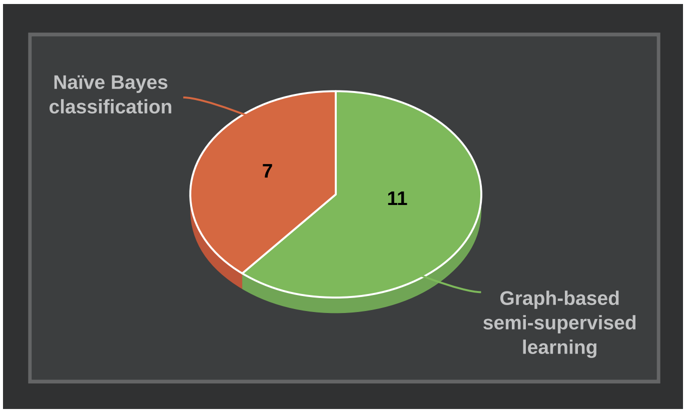

Fig. 17. List of the most used machine learning algorithms in Cluster 3.

the data contradict the hypothesis, less weight is assigned to it [254]. Some well-known methods used in this cluster include Naïve Bayes and Monte Carlo [256].

Multiple works in this cluster propose the detection of faults in photovoltaic systems through the utilization of a Bayesian approach. One of the most frequently referenced articles in this field, authored by Zhao et al. in 2015 [59], and cited 186 times, offers an alternative solution to conventional fault detection methods that typically rely on supervised learning models trained with a substantial amount of labeled data representing various fault types. This study underscores three primary limitations of supervised fault detection techniques: *i)* obtaining labeled photovoltaic data is challenging and costly, especially for high-power photovoltaic installations, *ii)* the trained model is not easily updatable, and *iii)* the model is intricate to visualize and interpret. To tackle these challenges, Zhao et al. [59] puts forth a graph-based semi-supervised learning model that necessitates only a limited number of labeled training data. These training data are normalized beforehand to enhance visualization. This article marks an initial stride toward effective fault identification without the need for extensive data. Notably, the model proposed by Zhao et al. [59] possesses the capability to autonomously learn about photovoltaic systems over time, even under varying weather conditions. The author attests to a 100% accuracy for both line-to-line fault and open-circuit fault.

The second article, presented by Momeni et al. in 2020 [262], with 23 citations, introduces a comprehensive approach to fault identification, classification, localization, and correction employing a graph-based semi-supervised learning algorithm and a substantial number of class labels. The system initially identifies the fault type and location, subsequently isolating it temporarily to ensure uninterrupted system operation until the fault is entirely corrected. Momeni et al. [262] proposes several normalization methods to enhance the differentiation between healthy and impacted panels. The results presented are exceptionally promising and arouse interest in the research community, as the methodology can be applied to photovoltaic systems of varying sizes due to a series of data normalization equations. The author demonstrates a 100% accuracy for both line-to-line fault and open circuit fault.

Lastly, the third article, proposed by Niazi et al. in 2019 [263], with 4 citations, recommends a distributed fault diagnosis approach for photovoltaic arrays that revolves around fine-tuning the Naive Bayes (FTNB) model. This approach addresses faults such as open-circuit, short-circuit, shading, abnormal degradation, and abnormal bypass diode. The Naive Bayes classifier, grounded in Bayesian principles [264], is utilized to compute the probability that each sample belongs to each category based on the attribute value of the sample. The results obtained facilitate the identification of the category with the highest probability as the predicted category for the new sample. After the model is calibrated, a set of fault data is employed to train the system, and fault signatures are then extracted from the maximum power point signals of photovoltaic inverters and meteorological data. However, the results are solely validated using simulated data. The author attests to a 99.2% accuracy for Hotspot fault.

In Fig. 17, the algorithms detected within this cluster are displayed along with their respective occurrence counts.

In Fig. 17, it is evident that there are more articles employing graph-based algorithms. This is because graph-based algorithms enable addressing semi-supervised problems, meaning they partially solve the issue of fault labeling in photovoltaic systems.

6.4. Cluster 4 - analogy-based approach

This cluster comprises algorithms known as kernel machines, which are called by a master algorithm. These algorithms assess resemblances between previous and new data by employing nearest-neighbor kernel machine algorithms to scrutinize their surroundings and create generalizations founded on similarity. This group of algorithms produces results like neural networks [254]. However, a major drawback of these algorithms is their high dependence on the size of the dataset, calculation time, and the complexity of programming, which can quickly become significant. Among the most popular algorithms in this group are K-Nearest Neighbor and support vector machines algorithms [256].

Several studies in this category suggest utilizing analytical-based strategies for detecting faults in photovoltaic systems. The most referenced article in this regard, authored by Baghae et al. in 2020 [265] (with 46 citations), addresses the problem of islanding detection. The proposed technique involves the application of support vector machines (SVM) and acquires input data through

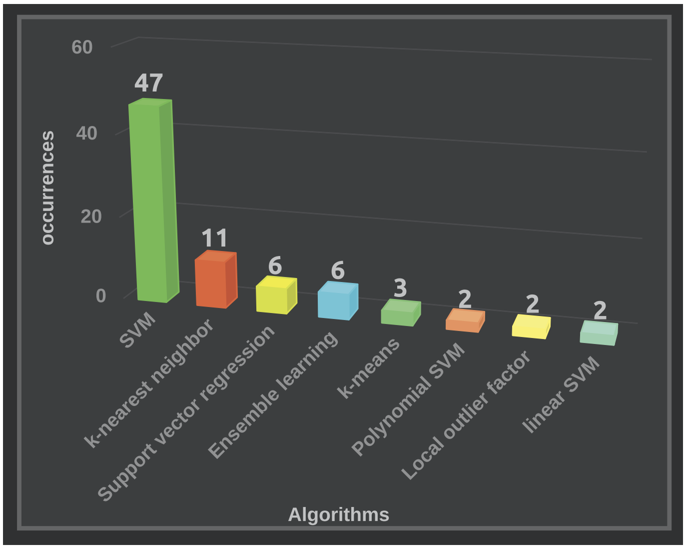

Fig. 18. List of the most used machine learning algorithms in Cluster 4.

simulations in MATLAB/Simulink. These simulations are based on an electrical analogy model of the system, calibrated using experimentally measured parameters from a photovoltaic (PV) plant. The approach assesses nondetection zones (NDZ) and detection time (DT) along two dimensions: the active power generated by the PV plant (ranging from 0% to 110% of the installed power capacity, $P_{installed}$), and the voltage levels of the distribution network (±10% of Nominal distribution network voltage U_n). The author demonstrates a 100% accuracy in detecting islanding and grid fault events.

In another article presented by Harrou et al. in 2019 [266] (with 42 citations), a distinct fault classification approach is introduced, with a focus on anomaly detection. The authors propose a model-based anomaly detection method for monitoring the direct current (DC) side of photovoltaic systems. This method combines the k-nearest neighbors (kNN) technique with univariate monitoring methods. The study introduces kNN-based and exponentially weighted moving average (EWMA) Shewhart schemes with both parametric and nonparametric thresholds to effectively identify faults in photovoltaic systems. The kNN method is particularly highlighted for its strong performance in handling non-linear datasets, such as those found in photovoltaic systems. The EWMA approach is employed due to its sensitivity in detecting minor changes. The kNN model utilizes residuals between actual and simulated signals as input, and parametric thresholds are suggested for fault identification. The final model is validated using a real 9.54 kWp grid-connected photovoltaic plant in Algeria. The author reports precision ranges of 75.6% to 99.1% for detecting open circuits, shading faults and short circuits.

Finally, a third article proposed by Ali et al. in 2020 [267] (with 28 citations) suggested a hybrid feature-based support vector machine (SVM) model that employs infrared thermography for the detection of hot spots in photovoltaic (PV) panels. This method introduces a novel hybrid feature vector that combines RGB, texture, the oriented gradient histogram (HOG), and the local binary pattern (LBP). The suggested approach categorizes panels into three groups: healthy, non-faulty hotspots, and faulty. The authors attain a training accuracy of 96.8% and a test accuracy of 92% using this approach. Moreover, this method exhibits lower computational complexity compared to other machine learning algorithms [267].

Lastly, a third article proposed by Ali et al. in 2020 [267] (28 citations) suggests a hybrid feature-based support vector machine (SVM) model that uses the infrared thermography technique for detecting hot spots in photovoltaic (PV) panels. This method introduces a novel hybrid feature vector that combines RGB, texture, the oriented gradient histogram (HOG), and the local binary pattern (LBP). The proposed approach classifies panels into three types: healthy, non-faulty hotspots, and faulty. The authors achieve a training accuracy of 96.8% and a test accuracy of 92% using this method. Additionally, this approach demonstrates lower computational complexity compared to other artificial intelligence algorithms [267].

In Fig. 18, the algorithms detected within this cluster are shown along with their respective occurrence counts.

As observed in Fig. 18, most of the documents within this cluster use classical support vector machines or their variants, together with K nearest neighbor algorithms. However, it is noteworthy that articles utilizing ensemble learning can also be found. In these cases, the authors connect multiple algorithms sequentially or in parallel to improve diagnostic fault efficiency. Equally interesting is the use of the local outlier factor, focusing primarily on anomaly detection and clustering algorithms (k-means [268]). The hypothesis of this study is that the growing adoption of these types of algorithms is once again linked to the complexity of accurately labeling faults for subsequent use in supervised algorithms.

6.5. Cluster 5 - connectionist approach

This cluster covers algorithms with backpropagation as the master algorithm. These algorithms simulate brain functions by creating artificial neurons and connecting them in a neural network with input, hidden, and output layers. The input neurons

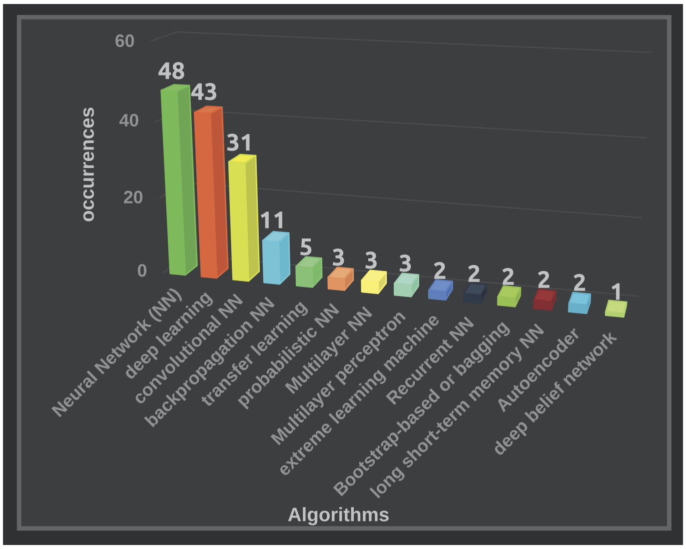

Fig. 19. List of the most used machine learning algorithms in Cluster 5.

transmit information to hidden neurons, which produce output that is readable by other neurons with the same function [254]. The most prominent algorithms in this category include extreme learning machines and artificial neural networks [256].

In this cluster, various studies focus on fault detection in photovoltaic systems using a connectionist approach. The most cited article, authored by Chen et al. in 2017 [231], addresses situations where fault diagnosis is severely limited by limited monitored information, hindering effective model training and updates. To address this, Chen et al. [231] propose extracting seven characteristics from monitored I-V characteristic curves under different environmental conditions. They use the kernel-based emerging extreme learning machine (KELM) for diagnosis due to its rapid learning and strong generalization. The Nelder-Mead Simplex (NMS) optimization method fine-tunes KELM parameters, improving the classification performance. Testing on a comprehensive data set from a photovoltaic modeling approach in Simulink and real data from a photovoltaic plant confirms the success of the optimized KELM-based fault diagnosis model, achieving high accuracy, reliability, and generalization for both real and simulated data, with precision ranging from 97.9% to 100% for short-circuit, temporary shading faults, and open-circuit.

A second article, presented by Garoudja et al. in 2017 [269], with 78 citations, introduces a method to detect and diagnose faults on the direct current (DC) side using the probabilistic neural network (PNN) classifier. This approach involves four stages: *(i)* parameter extraction of the photovoltaic module, *(ii)* simulation and experimental validation of the assembly, *(iii)* creation of a database of healthy and defective operations, and *(iv)* network construction, training, and testing. The unknown electrical parameters of the one-diode model (ODM) are precisely identified in the first stage using the best-so-far ABC algorithm based on the Artificial Bee Colony algorithm. These parameters facilitate system simulation using PSIM/Matlab. The detection and diagnosis of faults are then carried out using the PNN network. Testing in a 9.54 kWp grid-connected photovoltaic system includes scenarios of a healthy system, three modules short-circuited in one string, ten modules short-circuited in one string, and a string disconnected from the array. Garoudja et al. [269] demonstrate improved accuracy of the PNN method compared to feedforward backpropagation Artificial Neural Network (ANN) classifiers, achieving 100% accuracy in detecting string disconnection and short-circuit faults.

Lastly, the third article by Chen et al. in 2019 [270], with 75 citations, explores an alternative approach to fault diagnosis that eliminates the need for manual feature extraction or expert knowledge. This is a significant challenge in artificial intelligence-based fault diagnosis, as it can lead to inefficient or incomplete diagnosis models. The authors propose an intelligent fault detection and diagnosis method for photovoltaic arrays based on a deep residual network model trained with the adaptive moment estimation deep learning algorithm. This approach automatically extracts features from raw current-voltage curves, ambient irradiance, and temperature, enhancing performance with a deeper network. Results are validated with real and simulated data, demonstrating superiority over conventional neural network approaches in terms of accuracy, generalizability, reliability, and training efficiency, with precision ranging from 91.84% to 100% for short circuit, shading, open circuit, and degradation faults.

In Fig. 19, the algorithms detected within this cluster are shown along with their respective occurrence counts.

As seen in Fig. 19, the current computational advancements have led to the testing of multiple neural network-based algorithms. It is noteworthy that deep learning is rapidly gaining ground. Additionally, the utilization of transfer learning, which involves using pre-trained models to decrease the training time for neural network-based algorithms, has started to gain traction. This development is of great significance, as one of the primary drawbacks of neural network-based algorithms used to be their training time; however, transfer learning has significantly mitigated this issue.

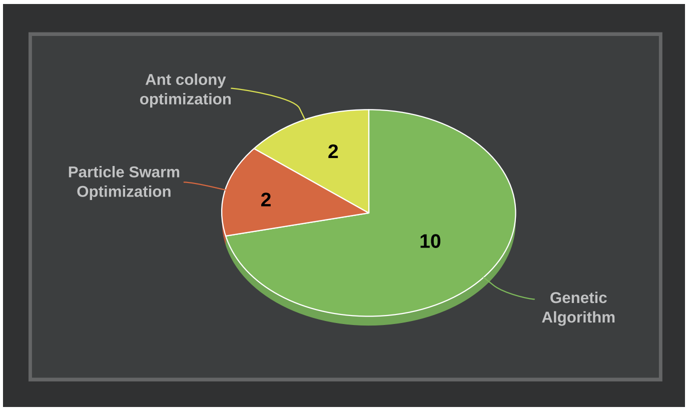

Fig. 20. List of the most used machine learning algorithms in Cluster 6.

6.6. Cluster 6 - evolutionary approach

These algorithms operate by simulating the evolutionary process observed in real genomes and DNA, adhering to Darwinian principles. In this methodology, performance is assessed through the measurement of offspring fitness [255]. These algorithms comprise a group of individuals that compete, undergo recombination and mutation, with only the most fit individuals retained, similar to natural selection in biology [254]. Genetic algorithms and genetic programming are two of the most popular methods falling within this category [256].

Numerous studies suggest applying these evolutionary approaches for detecting faults in photovoltaic systems. An influential work by Mohamed et al. in 2015 [271], with 19 citations, proposes utilizing a genetic algorithm (GA) to optimize artificial neural network (ANN) topology for online diagnosis and repair of photovoltaic energy systems. Comparisons with traditional neural networks and fuzzy-based diagnosis methods reveal substantial enhancements in fault detection. The author reports precision ranging from 97.6% to 98.9% in identifying open circuits, degradation, short circuits, and shading faults.

In another article by Lin et al. in 2015 [272], with 10 citations, a monitoring system is introduced for measuring parameters like output current, voltage, irradiation, ambient temperature, and irradiation per photovoltaic module. The use of a genetic algorithm as an optimizer for fault detection, coupled with a back-propagation neural network (BP) fault diagnosis model, yields improved diagnostic accuracy. The author reports precision levels between 78.69% and 100% for identifying short circuits, shadows, and aging faults.

Lastly, Rodrigues et al. presented a comprehensive approach in 2020 [273], with 3 citations. This approach combines a fuzzy system with five distinct machine learning models: artificial neural networks, multigene genetic programming, regression tree, support vector machines for regression and Gaussian process. The goal was to evaluate at least one algorithm from each machine learning category proposed in [254]. Results indicate that the GA delivered the most favorable outcomes on 5 of the 16 days tested, underscoring its effectiveness in fault detection. The author reports precision levels ranging from 97.5% to 98.7% for string faults within these 5 combinations.

In Fig. 20, the algorithms detected within this cluster are shown along with their respective occurrence counts.

As shown in Fig. 20, the utilization of genetic algorithms to attain optimal hyperparameters for fault detection algorithms in photovoltaic systems was extensively studied. However, due to advancements in computational capabilities of contemporary systems and the proliferation of hyperparameters within machine learning algorithms, the exploration of these optimization algorithms and others will constitute a pivotal aspect in the field of photovoltaics and beyond.

The technical details of the machine learning approaches presented in the Clusters (1-6) are compared in Table 13.

As evident from Table 13, each cluster (1–6) has been explored for diagnosing faults in photovoltaic systems. However, not only the most cited articles set current trends, but it is also important to examine in detail the last few years of publications in the area. Therefore, as a complementary section, this study highlights recent research in this area over the past three years.

6.7. The most recent and significant articles

The articles were chosen based on their popularity, level of innovation, and the number of views and downloads.

In [274], a Convolutional Neural Network (CNN) is suggested for training with historical data, using variables like current, voltage, temperature, and irradiance across five categories. The study reports that the CNN model achieved 97.64% training accuracy and 95.20% testing accuracy. Likewise, in [275], authors propose a deep learning approach using a CNN for photovoltaic systems. This method automates feature extraction, reducing computational load while improving classification. They achieved an accuracy of 98.65% in classifying various issues under different atmospheric conditions using current, voltage, and power as input indicators.

In [276], a CNN and a fine-tuned VGG-16 model are used to diagnose faults with infrared thermographic images of photovoltaic modules. They achieved a fault detection accuracy of 99.91% and a fault diagnosis accuracy of 99.80% using the VGG-16 model. The

Table 13
Comparison of the technical details of the algorithms used in each of the three most cited articles presented in the clusters (1–6).

Cluster	Ref	Target	Accuracy (%)	Validation		Fault	Data Samples	Software	Method	Sampling Time
				Sim	Real					
1	[232]	Detection	99,24	✓	✓	Line to line, open-circuit, shading	182160	Matlab	Random Forest (RF) ensemble learning	5 μs
	[232]	Classification	99,135	✓	✓	Line to line, open-circuit, shading	182160	Matlab	Random Forest (RF) ensemble learning	5 μs
	[257]	Classification	100	–	✓	Open-circuit, protection, poor connection, short-circuit, partial shading	1501	–	$RK - RF_{ED}$	100 μs
	[257]	Classification	100	–	✓	Open-circuit, protection, poor connection, short-circuit, partial shading	1501	–	$RK - RF_{Kmeans}$	100 μs
2	[258]	Detection	92,5	✓	–	Short-circuit	144	PSCAD/ EMTDC	Decision Tree	1 μs
	[258]	Detection	100	✓	–	Short-circuit	144	PSCAD/ EMTDC	AdaBoost	1 μs
	[259]	Detection	Far =7, 1419	✓	✓	Inverter, grid connection, sensor, PV panel, PV panel connection	150	solar Pro	Gaussian process regression	15 min
	[260]	Detection	98,21 – 99,97	✓		Inverter, grid connection, sensor, PV panel, PV panel connection	15775	DSpace 1104 (Matlab)	KNN	100 μs
	[260]	Detection	99,61 – 100	✓		Inverter, grid connection, sensor, PV panel, PV panel connection	15776	DSpace 1104 (Matlab)	Random forest	100 μs
	[260]	Detection	97,73 – 99,97	✓		Inverter, grid connection, sensor, PV panel, PV panel connection	15777	DSpace 1104 (Matlab)	DA	100 μs
	[261]	Classification	98,3 – 99	✓		Partial soiling, cell cracking	–	–	Random forest	30 min
3	[59]	Detection and Classification	100	✓	–	Line to Line, open-circuit	4879	Matlab Simulink	Graph-based semi-supervised learning	200 ms
	[262]	Detection and Classification	100	✓	–	Line to Line, open-circuit	35547	Matlab	Graph-based semi-supervised learning	20 ms
	[263]	Classification	99,2	–	✓	Hotspot	375	Matlab	Naive Bayes Classifier	–

(continued on next page)

Table 13 (*continued*)

Cluster	Ref	Target	Accuracy (%)	Validation		Fault	Data Samples	Software	Method	Sampling Time
				Sim	Real					
4	[265]	Detection	100	–	✓	Islanding and grid fault events	Multiple databases	Matlab Simulink	Support vector machines	5 – 20 ms
	[266]	Detection	75,6 – 99,1	–	✓	Short-circuit, open-circuit, temporary shading	1440	Matlab Simulink and PSIM	kNN	1 min
	[267]	Detection	92	–	✓	Hotspot	75	Matlab	Support vector machines	–
	[231]	Detection and Classification	98,77 – 100	✓	✓	Short-circuit, degradation, open-circuit, partial shading condition	4800 – 2500	Matlab Simulink	kernel-based emerging extreme learning machine	–
	[269]	Detection and Classification	100	✓	✓	String disconnection and short-circuit	740	Matlab Simulink and PSIM	Probabilistic neural network	1 min
5	[270]	Detection and Classification	91,84 – 100	✓	✓	Shading, Degradation, short-circuit, open-circuit	25237	Matlab Simulink	Deep residual network	–
6	[271]	Detection and Classification	97,6 – 98,9	✓	–	Short-circuit, open-circuit, degradation, shading	–	Matlab Simulink	genetic algorithm and artificial neural networks	–
	[272]	Detection and Classification	78,69 – 100	–	✓	Short-circuit, shadow, Aging	1600	Matlab Simulink	genetic algorithm and Backpropagation neural network	–
	[273]	Detection and Classification	97,5 – 98,7	✓	–	String fault	186	Matlab	Regression Tree, multi-gene genetic programming, Artificial Neural Network, Gaussian Process, Support vector machines	–

faults identified in that study are bypass diode malfunction, partially covered photovoltaic module, shading effect, short circuit, and dust deposition on the photovoltaic panel. In [277], a hybrid model combining CNN and machine learning is presented for various photovoltaic module issues. The hybrid model, implemented on a Raspberry Pi 4, achieved an accuracy level of 94% using a database collected from three different climatic regions. This study focuses on various photovoltaic module faults, including accumulated sand faults in photovoltaic modules, covered photovoltaic modules, cracked photovoltaic modules, degradation, dirty photovoltaic modules, short-circuited photovoltaic modules, and overheated bypass diodes.

In [278], authors introduce a deep learning approach for identifying cracks in photovoltaic cells. They evaluate four deep learning models and apply an ensemble learning technique to achieve a precision of 97.93%. Their methodology involves training, evaluation, and comparison of four deep learning models: U-net, LinkNet, FPN, and attention U-net. Model performance is measured using different metrics, including Intersection over Union (IoU) and F1-Score. In [279], machine learning and deep learning techniques are assessed for detecting and diagnosing faults in photovoltaic modules. Deep learning-based methods exhibited a precision of 98.71% for both binary and multiclass detection and classification tasks. The authors compiled a dataset of infrared thermography images from normal and faulty photovoltaic modules such as bypass diode fault, shading effects, photovoltaic module short-circuiting, and accumulated dirt on the photovoltaic module.

In [280], a hotspot detection system is proposed using thermographic images and a CNN. The conventional neural network achieved an accuracy of 95.05%, outperforming a short-term memory-based method. In [281], authors introduce a Probabilistic Neural Network (PNN) with a 100% accuracy rate for analyzing historical datasets related to normal and faulty states of photovoltaic systems during the winter season.

In [282], weightless neural networks (WNN) are used for detecting visual faults in photovoltaic modules. WNN uses random access memory (RAM) devices to simulate neuronal functionality. This approach achieved a 100% classification accuracy within 1.44 seconds of testing. This approach uses a pre-trained deep learning network (DenseNet-201) and to select the most significant features for classification, the authors used a decision tree algorithm (J48).

In [283], Mellit et al. employ machine learning algorithms, decision tree, and random forest to identify and categorize faults based on voltage and current measurement data, achieving 98% accuracy for fault detection and 96% for classification. The faults studied were accumulation of dust, shading, module disconnection, and short-circuited bypass diode using voltage and current measurement data (I-V curves).

In [284], an algorithm using logistic regression with cross-validation for fault detection on the DC side of photovoltaic systems shows an accuracy of 97.11%. This approach was applied to scenarios involving short-circuit, open-circuit, and mismatch faults.

Lastly, in [78], an innovative cascading classifier method is proposed for predicting the health status of photovoltaic systems. The approach accurately predicts health states and quantifies panel degradation, demonstrating cost-effectiveness. This approach initially identifies severely affected photovoltaic panels by only analyzing the series-collected current data sampled at 1-minute intervals. This is one of the few studies that focused on the analysis of the fault called snail trails.

A comprehensive review of these articles reveals the diverse research landscape in photovoltaic system fault detection using artificial intelligence. This comparison offers insights into the objectives pursued by different research teams.

7. Promising research topics

This section provides a summary of the key areas of promising research. Firstly, it has been noted in various parts of the articles, particularly those outlined in Section 6, that utilizing hybrid machine learning methods is increasingly vital to enhance the accuracy of fault detection in photovoltaic systems. The authors suggest the fusion of two or more algorithms from different clusters to achieve synergistic advantages, resulting in improved precision in diagnostic and prognostic systems [78,256]. One highly cited article employing hybrid techniques is [221] with 131 citations, which introduces a novel fault detection algorithm for photovoltaic (PV) systems by combining the ANN radial basis function (RBF) network, Mamdani, and Sugeno fuzzy logic systems through a new interface. Another notable article with 42 citations is [285], focusing on DC arc fault detection, proposing a domain adaptation combined with deep convolutional generative adversarial network (DA-DCGAN) methodology implemented in an embedded system and validated with a real 1.5 kW rooftop photovoltaic grid-connected system following the UL-1699B standard.

In summary, these hybrid techniques significantly enhance the precision and resilience of fault detection algorithms. As a result, researchers are encouraged to continue developing novel hybrid techniques that combine strengths to produce superior detection and classification outcomes. This is especially relevant given the recent advancements in software and hardware technologies, which expand the capabilities of systems for sampled measurements and real-world execution conditions. Furthermore, researchers are encouraged to study hardware approaches for collecting high-quality data specifically designed for fault detection [286].

Upon analyzing the algorithms in Subsection 6.1, it is evident that random forest and decision tree algorithms are the most commonly used. The Symbolic approach generally produces good results. However, there are only a few experiments that use other algorithms less explored from the same cluster. Therefore, it is recommended to increase research on other algorithms such as the C.45 Decision Tree or the binary Tree Bagging.

In Subsection 6.2, the main focus of the work focuses on using the PCA algorithm as a dimensionality reducer, followed by the treatment of the data with a classifier. However, there are other options, such as signal processing using ARIMA, SARIMA and SARIMAX models, that enable fault detection based on prediction. Prediction-based fault detection involves comparing predicted values to actual values and ascertaining fault presence by evaluating predefined limit parameters, which may include system damage levels or fault types.

In Subsection 6.3, the results obtained from these algorithms, particularly the methods based on semi-supervised learning based on graphs, are very interesting. Researchers are encouraged to further research and apply these algorithms in combination with the data normalizations proposed in the referenced articles. Data normalization can expand the size of databases and contribute to the generalization of fault detection models applied to photovoltaic systems.

Analyzing the results of Subsection 6.4, it can be concluded that both kNN and SVM are effectively applied. It is recommended to continue working with these algorithms due to their favorable results. Additionally, exploring other methods within this cluster, such as the local outlier factor or other clustering algorithms that allow anomaly detection such as Dyclee [287] and thus reduce the dependency of fault detection on having fault labels or some kernel modifications for SVM, is also advised.

Subsection 6.5 records the most publications, aligning with the observations presented in Fig. 14, demonstrating the growth of research in neural networks. In general, these algorithms require substantial amounts of data, necessitating high computational power. However, with the rapid technological advancements, this limitation is no longer a concern. We strongly encourage the authors to further research these algorithms and explore alternatives such as the "Extreme Learning Machine", which has shown promising and accurate results. Additionally, they should consider utilizing methods based on pre-trained neural networks to reduce training time such as transfer learning. Furthermore, we strongly encourage researchers to explore various algorithm configurations and methods of connection (such as Bagging, Boosting, Stacking, and Cascading) within the domain of ensemble learning algorithms.

In Subsection 6.6, it becomes evident that only a limited number of articles use evolutionary algorithms. We encourage the authors to explore metaheuristic algorithms more broadly as they allow parameter estimation and model adjustment in fault detection systems. While genetic algorithms are commonly used in fault detection for photovoltaic systems, there are multiple algorithms that offer more accurate results with reduced computation time for the same purposes [288]. We encourage the exploration of alternative algorithms, such as those mentioned in [289].

Moreover, the precision of error detection algorithms correlates directly with the quality of the training data. Therefore, the authors are encouraged to pay more attention to the processes of signature extraction and selection, as discussed in the articles [78, 257]. The current challenge goes beyond proposing new algorithms or modifying/combining existing ones to improve precision results. It involves extracting specific patterns (signatures) that accurately identify faults. Thus, this study strongly advocates the exploration of signal processing techniques such as signal filtering and time/frequency signal decomposition. The ability to extract relevant patterns not only facilitates corrective fault detection but also significantly improves preventive fault detection. Moreover, given the vast amounts of data that modern monitoring systems can accumulate, it is necessary to advance the development of novel algorithms for automated signature selection. These dimensionality reduction algorithms will significantly optimize the execution times of machine learning algorithms, as well as diminish the requirements for data storage, transmission, and processing, which is critical to ensure accurate fault detection. These factors become pivotal in scenarios where fault detection systems are embedded and therefore constrained by limited hardware resources.

Given the complexity of generating labels to train systems for all types of faults, it is worth noting that an interesting approach is to focus on detecting anomalies where fault labels are unnecessary. This method, commonly referred to as outlier detection or novelty detection, focuses on recognizing uncommon items, events, or observations that greatly differ from the majority or seem inconsistent with the standard or baseline data. Several significant and well-cited works on anomaly detection in photovoltaic systems already exist [290–294]. The primary role of outlier detection is to rapidly and unsupervised identify samples that begin to exhibit a significant performance loss. Outlier detection algorithms typically boast shorter computation times. Furthermore, the identification of atypical samples can significantly impact the accuracy of supervised machine learning models. Failing to appropriately detect and manage outlier values can lead to overfitting, inadequate, or biased results. Some well-known methods include Z-Score, Interquartile Range (IQR), and Local Outlier Factor (LOF). Therefore, we strongly advocate for the initial study of anomaly detection and outlier detection within the photovoltaic domain. Furthermore, this study identified a minimum of 26 occurrences of "anomaly detection" and 15 occurrences of "outlier detection", making it significant enough to initiate crucial research processes in this field.

In recent developments, new algorithms have shown promising outcomes in detecting faults in photovoltaic systems. One notable algorithm, discussed in the article titled "Online Fault Diagnosis for Photovoltaic Arrays Based on Fisher Discrimination Dictionary Learning for Sparse Representation" [295], deserves special mention.

8. Challenges and key elements in research and applications

In this review, it is noted that despite the variety of machine learning techniques accessible, most studies use conventional methods. This study seeks to overcome this limitation by conducting a thorough analysis and highlighting crucial aspects and obstacles. The objective is to inspire new research initiatives that can enhance fault diagnosis in photovoltaic systems. The upcoming sections will elaborate on these important aspects and challenges.

8.1. Key elements

The list of key elements identified in this study, necessary to advance the field of interest, is divided into three main aspects based on their affinity: *i)* AI for fault diagnosis in photovoltaic systems; *ii)* Monitoring focused on fault diagnosis; and *iii)* Formal fault dictionary. These three aspects are described in the following.

8.1.1. AI for fault diagnosis in photovoltaic systems

To adequately address a problem of fault diagnosis in photovoltaic systems using artificial intelligence, it is necessary to first build relevant and robust databases. In other words, these databases should include at least the following eight key elements.

First, it is essential to determine the ***data collection level***. This involves specifying whether the algorithm will be applied at the level of individual photovoltaic panels, photovoltaic string aggregates, protection systems, and so on. Second, you need to define the ***nature of the input data for the algorithm***. This entails deciding whether the algorithm will work with data types like characteristic curves, electrical signals including voltage, current, power, time-dependent meteorological data, images, etc. Third, you must establish the ***sampling rate***. The sampling rate is crucial and depends on the type of fault to be detected. Some faults occur rapidly, within a minute or less (such as arc faults), while others might take as long as 15 minutes or even weeks to manifest (like degradation faults). Thus, a balance between sampling frequency and data storage and processing capacity is essential.

Furthermore, a fourth consideration is ***generalizing photovoltaic models***. The accuracy of machine learning diagnostic results is heavily influenced by the quality of the training data. In other words, if the algorithm is trained on a specific dataset A and then used to diagnose data from a completely different photovoltaic plant, even if the data types are similar, the diagnostic results are likely to be inaccurate. Therefore, it is important to develop data preprocessing algorithms that can standardize the data from the two plants for comparison. These algorithms help to generalize the input data for the machine learning algorithm, taking into account factors like the operating time of the photovoltaic plant, inclination, orientation, the number of photovoltaic panels per string, photovoltaic technology, and more.

The fifth aspect is the ***diagnostic window***. It is important to determine the time interval during which the diagnosis will be performed in order to avoid hiding specific behaviors of a fault type. For example, calculating the average signal over an entire day of data versus a minute can significantly change the diagnostic result. This selection should depend on the specific fault to be diagnosed. The sixth aspect is ***multi domain data fusion***. Some faults in photovoltaic systems have a unique thermal signature (hotspot), while others have an electrical signature (premature aging). Therefore, in some cases, it is necessary to fuse data from multiple domains. This data fusion can significantly increase the range of detectable faults. It can even transition the diagnostic approach from corrective maintenance to preventive maintenance as proposed in [296]. The seventh aspect is ***feature engineering***, which had around 48 occurrences within the documents analyzed in this study. As some faults exhibit thermal signatures and other electrical signatures, some faults may not be clearly visible in the time domain, requiring a set of signal processing techniques to extract those hidden behaviors. Treatments such as Fourier transforms, wavelet analysis, etc., can be explored. Finally, the eighth aspect is the ***execution architecture of the diagnostic algorithm***. This aspect allows understanding the framework in which the algorithm will operate, such as limitations of the computational processor, memory, and storage, among others.

8.1.2. Monitoring focused on fault diagnosis

As mentioned earlier in this study, the diagnostic outcome is directly proportional to the quality of training data. For this reason, this study presents a set of nine key elements that can improve photovoltaic system monitoring focused on fault diagnosis.

Firstly, the ***diagnostic algorithm execution*** mechanism plays a crucial role. The way the diagnostic algorithm is executed directly defines the structure and constitution of the monitoring system. If the algorithm is embedded within the monitoring system, it is necessary to consider ***the storage capacity*** and ***processing capability***. On the other hand, if the algorithm is hosted on a server, the monitoring system will only be responsible for data acquisition, and in that case the most important aspect is ***the data transmission protocol***. This protocol must ensure the timely transmission of data without significant delays, discrepancies, or lost/aberrant data during the transfer. Furthermore, it should ensure the continuity of data transmission.

Another crucial consideration regarding monitoring is the identification of **measurable variables**. This involves specifying the sensors to be used, while also considering the potential for combining data. This entails thinking about sensors for measuring electrical and meteorological parameters, and even thermal or electroluminescence imagery. International standards have been established to simplify the planning of photovoltaic (PV) monitoring systems and to assist in the choice of sensor types and quantities. These standards establish criteria for monitoring photovoltaic facilities. For instance, the British standard BS IEC 61724 [297], the recommendations of the European Joint Research Center [298], and the guidelines of the National Renewable Energy Laboratory (NREL) [299] provide comprehensive guidance for assessing the performance of photovoltaic systems.

One more factor to consider is the ***operation environment of the monitoring system***. In cases where the system is placed in a high-capacity solar power facility, it is important to be aware that certain plants might not have wireless networks like WiFi available for data transfer. Additionally, in some cases, there may not be electrical outlets available to power the monitoring system. Therefore, the system must be self-sufficient in terms of energy and adapt to the absence of electrical outlets. Consequently, the monitoring system should have an integrated power source capable of supplying the necessary power to ensure continuous data acquisition.

The fourth aspect is ***noise robustness***. Fault diagnosis algorithms using artificial intelligence are highly sensitive to noise. Hence, monitoring systems must be protected against electromagnetic interference from the environment (such as the presence of parasitic currents) that can disrupt the captured signals.

The fifth aspect to take into account when addressing data management and data privacy is the necessity for a strong ***cybersecurity*** element within the system. This implies that the system should have the capability to encrypt the data collected at the photovoltaic station to safeguard it from being stolen.

Lastly, ***the portability of the monitoring system*** must be defined. Taking into account the aforementioned aspects, it is necessary to determine whether the monitoring system will be continuously connected to the photovoltaic system or if it will only be connected for periodic diagnostic purposes. This aspect directly affects the dimensioning of both the internal elements of the monitoring system and the associated power supply.

8.1.3. Formal fault dictionary

One of the significant challenges in fault diagnosis algorithms employing supervised artificial intelligence involves the need to label training data accurately and to achieve a deep understanding of occurring faults. These two aspects are closely intertwined, as comprehensive knowledge of fault behavior is essential for data labeling. Typically, this process relies on traditional fault identification methods conducted in a lab setting or with input from field experts. Nonetheless, numerous issues are associated with this approach, including disagreements regarding fault classification and disparities between lab-identified fault conditions and those encountered in the field. Using the information extracted from the articles reviewed in this study, this section proposes essential components to consider when establishing a formal fault dictionary for photovoltaic systems, with a primary focus on fault diagnosis and labeling. A label, in this context, refers to the specific designation for the operational state of a photovoltaic panel or string, such as "healthy," "hot spot," or "delamination". The critical aspects that the fault dictionary should cover are described below.

First, the dictionary should include the ***component where the fault occurs***. In other words, it should define whether faults occur at the level of photovoltaic cells, photovoltaic panels, photovoltaic strings, photovoltaic arrays, protective systems, etc. Along the same lines, as a second key aspect, the dictionary should be capable of providing information about the ***origin of the fault***. For example, differentiating between faults of structural, electrical, and thermal origins, among others. In addition, it should define whether the origin is related to external causes, material aging, or other factors. The third key aspect identified in this study that should be part of the dictionary is ***the relationship between faults***. It is becoming increasingly evident that fault diagnosis systems need to be more powerful and computationally lighter. Therefore, the fault dictionary needs to group faults not only based on the component where they occur but also considering the cause-and-effect relationships between them. For example, if there are microcrack faults that can later lead to hot spots. This is crucial for the future of fault diagnosis in photovoltaic systems because it allows focusing efforts only on detecting the cause (microcracks) rather than the symptoms (hot spots), ensuring high diagnostic performance and maximizing the number of diagnosed faults.

A fourth aspect that should be explored is ***the occurrence frequency***. This aspect will help identify diagnostic windows. Defining the effects on fault occurrence based on factors such as the age of the power plant, the meteorological conditions under which the system is installed, or even the technology used. A fifth aspect to consider within the dictionary is an analysis of ***the priority level*** in terms of power loss and human safety. Understanding this priority allows corrective or preventive actions to be taken based on the level of risk for the photovoltaic plant or the personnel within the photovoltaic plant. Lastly, this study is considered a sixth key element to study ***conventional fault detection methods***. The objective is for these methods to be used in a second phase to validate and refine the results of the fault diagnosis.

8.2. Challenges

In this study, it has been noted that the diagnosis of faults in photovoltaic systems with the help of artificial intelligence must shift towards obtaining pertinent real-world data for fault monitoring and detection. This necessitates a departure from the traditional model-based training, which entails creating mathematical representations of systems, and instead embracing data-driven training systems or models built from actual data. This change ensures a more accurate representation of the system, leading to improved and more applicable responses from machine learning systems.

Nonetheless, as previously stated, the key to obtaining reliable results that account for real-world disturbances caused by environmental factors and measurement device signal noise lies in prioritizing the acquisition of high-quality data. The use of pure algorithms is becoming progressively challenging. Hence, researchers are encouraged to explore hybrid approaches that combine the strengths of various algorithms, facilitated by the concept of the 'master algorithm'. By bridging the gap between different algorithmic groups, researchers can extract the optimal outcomes from each approach.

Furthermore, it is essential not only to develop a comprehensive fault detection algorithm capable of functioning in diverse conditions but also to conduct rigorous mathematical work. This mathematical understanding will facilitate a deeper understanding of the relationships between photovoltaic systems and weather conditions, such as temperature, irradiation, humidity, and others. Through this understanding, a set of equations can be derived, enabling comparison of performance and fault detection even when evaluating plants installed under variable conditions.

In addition to these aspects, it is interesting to highlight that the Cluster (1-5) have been widely used for fault detection purposes. Meanwhile, Cluster 6 has been used as a complement to the other clusters, enhancing overall performance by optimizing the algorithm hyperparameters. However, it is important to note that within each cluster, supervised, unsupervised, and semi-supervised approaches can be found. This underscores the necessity of careful attention to data labeling used for system training. The challenge of effectively labeling the data to diagnose faults in photovoltaic systems represents one of the main contemporary challenges.

9. Discussion and conclusions

The proposed methodology in this study introduces, for the first time, a way to extract a set of conditions and trends from a considerable number of documents regarding fault detection in photovoltaic systems. This aspect had never been clearly established before using a methodology based on numerous bibliometric analyses. Consequently, the methodology applied to photovoltaic systems could serve as a valuable tool for researchers approaching scientific topics from a Big Data perspective. In this study, over 620 research papers on fault detection in photovoltaic systems that use artificial intelligence were identified and analyzed. Articles that only presented frameworks without including experimental research were excluded from consideration.

This research introduces a method for performing a literature review that integrates bibliometric content analysis with expert qualitative content analysis. This method attempts to identify essential parameters, obstacles, and research perspectives to support and guide upcoming studies in the domain of fault detection in photovoltaic systems. Furthermore, it offers a comprehensive collection of algorithms and relevant papers grouped together to enhance understanding of the various categories of algorithms to be explored in future research.

Multiple relevant aspects for future research were obtained. First, this study recommends the research of hybrid machine learning and comparative studies of these methods in fault detection in photovoltaic systems. Second, research on photovoltaic system forecasting should be encouraged to enable condition-based preventive maintenance and reduce investment recovery times. While industrial demands tend to focus on complete photovoltaic arrays, monitoring at the photovoltaic panel level still requires significant development and should progress alongside machine learning algorithms. Third, the current challenge is not only the construction of more robust photovoltaic systems but also ensuring their optimal production. In this regard, this study recommends diligent research on the construction of robust databases and specialized signal processing for early fault diagnosis.

Additionally, this study, as demonstrated in Section 5.1, identifies highly interesting countries or target countries for establishing collaborative networks. For example, China, India, and the United States are countries that dominate most of the research in the field of study. Interestingly, this methodology also highlights France as a country of interest. With approximately 15% publications compared to China, India, or the United States, France surpasses India in terms of citations and achieves roughly half the citations of China or the United States. Therefore, this study encourages close monitoring of efforts on artificial intelligence and renewable energy topics within these four countries.

This study also demonstrated the need for the design of photovoltaic monitoring systems to be adapted to the new requirements of Industry 4.0 and Big Data. This means constructing new monitoring systems capable of hosting artificial intelligence for forecasting, tracking, anomaly detection, and health state diagnostics, among other functions. Additionally, these systems should demonstrate high performance in terms of data transmission, processing speeds, storage, multi domain data, and other aspects outlined in the key elements section.

The detailed analysis conducted in this study made it evident that there is no consensus on a general classification of faults in photovoltaic systems. Therefore, this study proposed a set of key aspects that would enable future fault diagnosis research to follow a unified path, accelerating progress and facilitating collaborative work and direct algorithm extrapolation.

Finally, the research findings validate the importance of ongoing efforts by researchers to create novel algorithms and tools for fault detection in photovoltaic systems with the aid of artificial intelligence. This will help offer a comprehensive and objective perspective on the current state of sustainable AI research in the energy sector.

CRediT authorship contribution statement

Edgar Hernando Sepúlveda-Oviedo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Louise Travé-Massuyès:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Audine Subias:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marko Pavlov:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Corinne Alonso:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

This project is related to ANITI within the framework of the French project called ”Investing for the Future – PIA3”, with financing reference number ANR-19-PI3A-0004.

Data availability

Data will be made available on request.

Acknowledgements

This project is related to ANITI within the framework of the French project called “Investing for the Future – PIA3”, with financing reference number ANR-19-PI3A-0004.

References

[1] Q. Navid, A. Hassan, A.A. Fardoun, R. Ramzan, A. Alraeesi, Fault diagnostic methodologies for utility-scale photovoltaic power plants: a state of the art review, Sustainability 13 (4) (2021) 1629.

[2] S.R. Madeti, S.N. Singh, Monitoring system for photovoltaic plants: a review, Renew. Sustain. Energy Rev. 67 (2017) 1180–1207.
[3] R. Hariharan, M. Chakkarapani, G.S. Ilango, Challenges in the detection of line-line faults in PV arrays due to partial shading, in: 2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS), 2016, pp. 23–27.
[4] (NREL) NREL. Best Research-Cell Efficiency Chart, 2022.
[5] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, et al., Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents, Energy Convers. Manag. 178 (2018) 250–264.
[6] N.G. Dhere, N.S. Shiradkar, Fire hazard and other safety concerns of photovoltaic systems, J. Photon. Energy 2 (1) (2012) 1–14.
[7] C.E. Chamberlin, M.A. Rocheleau, M.W. Marshall, A.M. Reis, N.T. Coleman, P.A. Lehman, Comparison of PV module performance before and after 11 and 20 years of field exposure, in: 2011 37th IEEE Photovoltaic Specialists Conference, 2011, 000101-5.
[8] A. Berasategi Arostegi, New Optimized Electrical Architectures of a Photovoltaic Generators with High Conversion Efficiency, Theses, Université Paul Sabatier - Toulouse III, 2013.
[9] S.N. Deshkar, S.B. Dhale, J.S. Mukherjee, T.S. Babu, Rajasekar N. Solar, PV array reconfiguration under partial shading conditions for maximum power extraction using genetic algorithm, Renew. Sustain. Energy Rev. 43 (2015) 102–110.
[10] B.I. Rani, G.S. Ilango, Nagamani C. Enhanced, Power generation from PV array under partial shading conditions by shade dispersion using Su Do Ku configuration, IEEE Trans. Sustain. Energy 4 (3) (2013) 594–601.
[11] N.D. Kaushika, N.K. Gautam, Energy yield simulations of interconnected solar PV arrays, IEEE Trans. Energy Convers. 18 (1) (2003) 127–134.
[12] D. Picault, Reduction of mismatch losses in grid-connected photovoltaic systems using alternative topologies, Theses, Institut National Polytechnique de Grenoble - INPG, 2010.
[13] M. Sun, S. Lv, X. Zhao, R. Li, W. Zhang, X. Zhang, 1. Defect detection of photovoltaic modules based on convolutional neural network, in: Third International Conference, MLICOM 2018, 2018, pp. 122–132.
[14] L. Cristaldi, M. Faifer, M. Lazzaroni, M.M.A.F. Khalil, M. Catelani, L. Ciani, Diagnostic architecture: a procedure based on the analysis of the failure causes applied to photovoltaic plants, Measurement 67 (2015) 99–107.
[15] P. Manganiello, M. Balato, M. Vitelli, A survey on mismatching and aging of PV modules: the closed loop, IEEE Trans. Ind. Electron. 62 (11) (2015) 7276–7286.
[16] E.L. Meyer, E.E. van Dyk, Assessing the reliability and degradation of photovoltaic module performance parameters, IEEE Trans. Reliab. 53 (1) (2004) 83–92.
[17] M. Köntges, I. Kunze, S. Kajari-Schröder, X. Breitenmoser, B. Bjørneklett, The risk of power loss in crystalline silicon based photovoltaic modules due to micro-cracks, Sol. Energy Mater. Sol. Cells 95 (4) (2011) 1131–1137.
[18] C.E. Packard, J.H. Wohlgemuth, S.R. Kurtz, Development of a Visual Inspection Data Collection Tool for Evaluation of Fielded PV Module Condition, National Renewable Energy Lab. (NREL), Golden, CO, USA, 2012.
[19] B. Adothu, P. Bhatt, S. Chattopadhyay, S. Zele, J. Oderkerk, H.P. Sagar, et al., Newly developed thermoplastic polyolefin encapsulant–a potential candidate for crystalline silicon photovoltaic modules encapsulation, Sol. Energy 194 (2019) 581–588.
[20] M. Koentges, S. Kurtz, C.E. Packard, U. Jahn, K.A. Berger, K. Kato, et al., Review of Failures of Photovoltaic Modules, International Energy Agency IEA, 2014.
[21] A. Parretta, M. Bombace, G. Graditi, R. Schioppo, Optical degradation of long-term, field-aged c-Si photovoltaic modules, Sol. Energy Mater. Sol. Cells 86 (3) (2005) 349–364.
[22] A.B. Rabii, M. Jraidi, A. Bouazzi, Investigation of the degradation in field-aged photovoltaic modules, in: 3rd World Conference on Photovoltaic Energy Conversion, 2003. Proceedings of, 2003, pp. 2004–2006, Vol. 2.
[23] D.C. Jordan, T.J. Silverman, J.H. Wohlgemuth, S.R. Kurtz, K.T. VanSant, Photovoltaic failure and degradation modes, Prog. Photovolt. 25 (4) (2017) 318–326.
[24] B. Li, Health monitoring of photovoltaic modules using electrical measurements, Theses, Université Paris-Saclay, 2021.
[25] N. Kim, K.J. Hwang, D. Kim, J.H. Lee, S. Jeong, D.H. Jeong, Analysis and reproduction of snail trails on silver grid lines in crystalline silicon photovoltaic modules, Sol. Energy 124 (2016) 153–162.
[26] T. Felder, H. Hu, W. Gambogi, K.R. Choudhury, S. MacMaster, L.G. Lles, et al., Field Study and Analysis of Backsheet Degradation in 450MW+ PV Installations, Dupont, 2017.
[27] M.A. Munoz, M.C. Alonso-García, N. Vela, F. Chenlo, Early degradation of silicon PV modules and guaranty conditions, Sol. Energy 85 (9) (2011) 2264–2274.
[28] L.N. Dumas, Shumka A. Photovoltaic, Module reliability improvement through application testing and failure analysis, IEEE Trans. Reliab. R-31 (3) (1982) 228–234.
[29] G. Oreski, G.M. Wallner, Damp heat induced physical aging of PV encapsulation materials, in: 2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 2010, pp. 1–6.
[30] C.G. Zimmermann, C. Nömayr, M. Kolb, A. Rucki, A mechanism of solar cell degradation in high intensity, high temperature space missions, Prog. Photovolt. 21 (4) (2013) 420–435.
[31] N. Park, C. Han, W. Hong, D. Kim, The effect of encapsulant delamination on electrical performance of PV module, in: 2011 37th IEEE Photovoltaic Specialists Conference, 2011, 001113-5.
[32] M.D. Kempe, Modeling of rates of moisture ingress into photovoltaic modules, Sol. Energy Mater. Sol. Cells 90 (16) (2006) 2720–2738.
[33] E.D. Dunlop, D. Halton, The performance of crystalline silicon photovoltaic solar modules after 22 years of continuous outdoor exposure, Prog. Photovolt. 14 (1) (2006) 53–64.
[34] S. Kaplanis, E. Kaplani, Energy performance and degradation over 20 years performance of BP c-Si PV modules, in: Sustainable Energy and Environmental Protection "SEEP2009", Simul. Model. Pract. Theory 19 (4) (2011) 1201–1211.
[35] XH. Nguyen, Matlab/Simulink based modeling to study effect of partial shadow on solar photovoltaic array, Environ. Syst. Res. 4 (2015) 1–10.
[36] H. Patel, V. Agarwal, MATLAB-Based modeling to study the effects of partial shading on PV array characteristics, IEEE Trans. Energy Convers. 23 (1) (2008) 302–310.
[37] A. Triki-Lahiani, A. Bennani-Ben Abdelghani, I. Slama-Belkhodja, Fault detection and monitoring systems for photovoltaic installations: a review, Renew. Sustain. Energy Rev. 82 (2018) 2680–2692.
[38] M.R. Maghami, H. Hizam, C. Gomes, M.A. Radzi, M.I. Rezadad, S. Hajighorbani, Power loss due to soiling on solar panel: a review, Renew. Sustain. Energy Rev. 59 (2016) 1307–1316.
[39] Rajasekar N. Pillai DS, A comprehensive review on protection challenges and fault diagnosis in PV systems, Renew. Sustain. Energy Rev. 91 (2018) 18–40.
[40] J. Solórzano, M.A. Egido, Automatic fault diagnosis in PV systems with distributed MPPT, Energy Convers. Manag. 76 (2013) 925–934.
[41] D. Ji, C. Zhang, M. Lv, Y. Ma, Guan N. Photovoltaic, Array fault detection by automatic reconfiguration, Energies 10 (5) (2017).
[42] S. Stettler, P. Toggweiler, E. Wiemken, W. Heidenreich, A.C. de Keizer, W. van Sark, et al., Failure detection routine for grid connected Pv systems as part of the Pvsat-2 project, in: EUROSUN2006, 2005, pp. 1–6.
[43] M. Dreidy, B. Alsayid, S. Alsadi, J. Jallad, Partial shading of PV system simulation with experimental results, Smart Grid Renew. Energy (04) (2013 09).
[44] E. Koutroulis, F. Blaabjerg, A new technique for tracking the global maximum power point of PV arrays operating under partial-shading conditions, IEEE J. Photovolt. 2 (2) (2012) 184–190.
[45] D.S. Pillai, F. Blaabjerg, N. Rajasekar, A comparative evaluation of advanced fault detection approaches for PV systems, IEEE J. Photovolt. 9 (2) (2019) 513–527.
[46] F.J. Pern, Ethylene-vinyl acetate (EVA) encapsulants for photovoltaic modules: degradation and discoloration mechanisms and formulation modifications for improved photostability, Angew. Makromol. Chem. 252 (1) (1997) 195–216.

[47] A. Ndiaye, A. Charki, A. Kobi, C.M.F. Kébé, P.A. Ndiaye, V. Sambou, Degradations of silicon photovoltaic modules: a literature review, Sol. Energy 96 (2013) 140–151.
[48] E. Kaplani, Degradation effects in sc-Si PV modules subjected to natural and induced ageing after several years of field operation, J. Eng. Sci. Technol. Rev. 5 (2012 12) 18–23.
[49] G. Schirripa Spagnolo, P. Del Vecchio, G. Makary, D. Papalillo, A. Martocchia, A review of IR thermography applied to PV systems, in: 2012 11th International Conference on Environment and Electrical Engineering, 2012, pp. 879–884.
[50] V. Sharma, S.S. Chandel, Performance and degradation analysis for long term reliability of solar photovoltaic systems: a review, Renew. Sustain. Energy Rev. 27 (2013) 753–767.
[51] L. Chen, S. Li, Wang X. Quickest, Fault detection in photovoltaic systems, IEEE Trans. Smart Grid 9 (3) (2018) 1835–1847.
[52] Y. Zhao, B. Lehman, J. de Palma JF Mosesian, R. Lyons, Challenges to overcurrent protection devices under line-line faults in solar photovoltaic arrays, in: 2011 IEEE Energy Conversion Congress and Exposition, 2011, pp. 20–27.
[53] A. Appiah, X. Zhang, B. Ayawli, Kyeremeh F. Review, Performance evaluation of photovoltaic array fault detection and diagnosis techniques, Int. J. Photoenergy 2019 (02) (2019) 1–19.
[54] Y. Zhao, J. de Palma JF Mosesian, R. Lyons, Lehman B. Line–Line, Fault analysis and protection challenges in solar photovoltaic arrays, IEEE Trans. Ind. Electron. 60 (9) (2013) 3784–3795.
[55] Y. Zhao, B. Lehman, J. de Palma JF Mosesian, R. Lyons, Fault analysis in solar PV arrays under: low irradiance conditions and reverse connections, in: 2011 37th IEEE Photovoltaic Specialists Conference, 2011, 002000-5.
[56] M.C. Falvo, S. Capparella, Safety issues in PV systems: design choices for a secure fault detection and for preventing fire risk, Case Stud. Fire Saf. 3 (2015) 1–16.
[57] Y. Zhao, Fault detection, classification and protection in solar photovoltaic arrays, Ph.D. thesis, Northeastern University, 2015.
[58] Z. Yi, Etemadi AH. Line-to-Line, Fault detection for photovoltaic arrays based on multiresolution signal decomposition and two-stage support vector machine, IEEE Trans. Ind. Electron. 64 (11) (2017) 8546–8556.
[59] Y. Zhao, R. Ball, J. Mosesian, J.F. de Palma, B. Lehman, Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays, IEEE Trans. Power Electron. 30 (5) (2015) 2848–2858.
[60] A. Mellit, S. Kalogirou, Assessment of machine learning and ensemble methods for fault diagnosis of photovoltaic systems, Renew. Energy 184 (2022) 1074–1090.
[61] E. Spooner, N. Wilmot, Safety issues, arcing and fusing in PV arrays, in: ISES-AP - 3rd International Solar Energy Society Conference – Asia Pacific Region (ISES-AP-08) Incorporating the 46th ANZSES Conference, 2008, pp. 1–9.
[62] K. Xia, Z. He, Y. Yuan, Y. Wang, P. Xu, An arc fault detection system for the household photovoltaic inverter according to the DC bus currents, in: 2015 18th International Conference on Electrical Machines and Systems (ICEMS), 2015, pp. 1687–1690.
[63] J. Johnson, M. Montoya, S. McCalmont, G. Katzir, F. Fuks, J. Earle, et al., Differentiating series and parallel photovoltaic arc-faults, in: 2012 38th IEEE Photovoltaic Specialists Conference, 2012, 000720-6.
[64] J. Johnson, Electrical and Thermal Finite Element Modeling of Arc Faults in Photovoltaic Bypass Diodes, SANDIA, 2012.
[65] K. AbdulMawjood, S.S. Refaat, W.G. Morsi, Detection and prediction of faults in photovoltaic arrays: a review, in: 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), IEEE, Doha, 2018, pp. 1–8.
[66] M.K. Alam, F.H. Khan, J. Johnson, J. Flicker, PV faults: overview, modeling, prevention and detection techniques, in: COMPEL Salt (Ed.), 2013 IEEE 14th Workshop on Control and Modeling for Power Electronics, IEEE, Lake City, UT, USA, 2013, pp. 1–7.
[67] A. Mellit, S. Kalogirou, 6 - Machine learning and deep learning algorithms for fault diagnosis of photovoltaic systems, in: A. Mellit, S. Kalogirou (Eds.), Handbook of Artificial Intelligence Techniques in Photovoltaic Systems, Academic Press, 2022, pp. 183–226.
[68] B. Verhoeven, Utility Aspects of Grid Connected Photovoltaic Power Systems, International Energy Agency IEA, 1998.
[69] A. Mellit, G.M. Tina, S.A. Kalogirou, Fault detection and diagnosis methods for photovoltaic systems: a review, Renew. Sustain. Energy Rev. 91 (2018) 1–17.
[70] M. Chang, C. Chen, C.H. Hsueh, W.J. Hsieh, E. Yen, K.L. Ho, et al., The reliability investigation of PV junction box based on 1GW worldwide field database, in: 2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC), 2015, pp. 1–4.
[71] P. Sánchez-Friera, M. Piliougine, J. Peláez, J. Carretero, M. Sidrach de Cardona, Analysis of degradation mechanisms of crystalline silicon PV modules after 12 years of operation in Southern Europe, Prog. Photovolt. 19 (6) (2011) 658–666.
[72] D.L. King, M.A. Quintana, J.A. Kratochvil, D.E. Ellibee, B.R. Hansen, Photovoltaic module performance and durability following long-term field exposure, AIP Conf. Proc. 462 (1) (1999) 565–571.
[73] N.D. Kaushika, A.K. Rai, An investigation of mismatch losses in solar photovoltaic cell networks, Energy 32 (5) (2007) 755–759.
[74] L. Kato, PV module failures observed in the field- solder bond and bypass diode failures, International Energy Agency IEA (2015).
[75] W. Rezgui, N.K. Mouss, L.H. Mouss, M.D. Mouss, Y. Amirat, M. Benbouzid, Faults modeling of the impedance and reversed polarity types within the PV generator operation, in: 3rd International Symposium on Environmental Friendly Energies and Applications (EFEA), 2014, pp. 1–6.
[76] S. Duman, J. Li, L. Wu, N. Yorukeren, Symbiotic organisms search algorithm-based security-constrained AC–DC OPF regarding uncertainty of wind, PV and PEV systems, Soft Comput. 25 (14) (2021 jul) 9389–9426.
[77] C. Huang, L. Wang, R.S.C. Yeung, Z. Zhang, H.S.H. Chung, A. Bensoussan, A prediction model-guided Jaya algorithm for the PV system maximum power point tracking, IEEE Trans. Sustain. Energy 9 (1) (2018) 45–55.
[78] Sepúlveda Oviedo EH, L. Travé-Massuyès, A. Subias, C. Alonso, M. Pavlov, Feature extraction and health status prediction in PV systems, Adv. Eng. Inform. 53 (2022) 101696.
[79] Sepúlveda Oviedo EH, L. Travé-Massuyès, A. Subias, C. Alonso, M. Pavlov, Hierarchical clustering and dynamic time warping for fault detection in photovoltaic systems, in: X Congreso Internacional Ingeniería Mecánica, Mecatrónica y Automatización (XCIMM), Bogotá, Colombia, 2021, pp. 1–2.
[80] A. Colli, Failure mode and effect analysis for photovoltaic systems, Renew. Sustain. Energy Rev. 50 (2015) 804–809.
[81] H. Zhu, L. Lu, J. Yao, S. Dai, Y. Hu, Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model, Sol. Energy 176 (2018) 395–405.
[82] A. Livera, M. Theristis, G. Makrides, G.E. Georghiou, Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems, Renew. Energy 133 (2019) 126–143.
[83] T.P. Carvalho, F.A.A.M.N. Soares, R. Vita, R.d.P. Francisco, J.P. Basto, S.G.S. Alcalá, A systematic literature review of machine learning methods applied to predictive maintenance, Comput. Ind. Eng. 137 (2019) 106024.
[84] S.R. Madeti, S.N. Singh, A comprehensive study on different types of faults and detection techniques for solar photovoltaic system, Sol. Energy 158 (2017) 161–185.
[85] M.K. Alam, F. Khan, J. Johnson, J. Flicker, A comprehensive review of catastrophic faults in PV arrays: types, detection, and mitigation techniques, IEEE J. Photovolt. 5 (3) (2015) 982–997.
[86] A. Okere, M.T. Iqbal, A review of conventional fault detection techniques in solar PV systems and a proposal of long range (LoRa) wireless sensor network for module level monitoring and fault diagnosis in large solar PV farms, Eur. J. Electr. Eng. Comput. Sci. 4 (6) (2020).
[87] S. Jadidi, H. Badihi, Zhang Y. Fault, Diagnosis in microgrids with integration of solar photovoltaic systems: a review, IFAC-PapersOnLine 53 (2) (2020) 12091–12096.

[88] S. Lu, B.T. Phung, D. Zhang, A comprehensive review on DC arc faults and their diagnosis methods in photovoltaic systems, Renew. Sustain. Energy Rev. 89 (2018) 88–98.
[89] J. Hare, X. Shi, S. Gupta, A. Bazzi, Fault diagnostics in smart micro-grids: a survey, Renew. Sustain. Energy Rev. 60 (2016) 1114–1124.
[90] J.A. Tsanakas, L. Ha, C. Buerhop, Faults and infrared thermographic diagnosis in operating c-Si photovoltaic modules: a review of research and future challenges, Renew. Sustain. Energy Rev. 62 (2016) 695–709.
[91] A. Mellit, S. Kalogirou, Artificial intelligence and Internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: challenges, recommendations and future directions, Renew. Sustain. Energy Rev. (2021) 143.
[92] B. Li, C. Delpha, D. Diallo, A. Migan-Dubois, Application of artificial neural networks to photovoltaic fault detection and diagnosis: a review, Renew. Sustain. Energy Rev. 138 (2021) 110512.
[93] A. Abubakar, C.F.M. Almeida, M. Gemignani, Review of artificial intelligence-based failure detection and diagnosis methods for solar photovoltaic systems, Machines 9 (12) (2021) 328.
[94] M. Cubukcu, A. Akanalci, Real-time inspection and determination methods of faults on photovoltaic power systems by thermal imaging in Turkey, Renew. Energy 147 (2020) 1231–1238.
[95] Y.Y. Hong, R.A. Pula, Methods of photovoltaic fault detection and classification: a review, Energy Rep. 8 (2022) 5898–5929.
[96] S. Vergura, F. Marino, M. Carpentieri, Processing infrared image of PV modules for defects classification, in: 2015 International Conference on Renewable Energy Research and Applications (ICRERA), 2015, pp. 1337–1341.
[97] P. Guerriero, G. Cuozzo, S. Daliento, Health diagnostics of PV panels by means of single cell analysis of thermographic images, in: 2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC), 2016, pp. 1–6.
[98] K. AbdulMawjood, S.S. Refaat, W.G. Morsi, Detection and prediction of faults in photovoltaic arrays: a review, in: 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), 2018, pp. 1–8.
[99] T.D. Hund, D.L. King, Analysis Techniques Used on Field Degraded Photovoltaic Modules, Sandia National Laboratory, 1995, p. 1.
[100] G. Alves dos Reis Benatto, C. Mantel, S. Spataru, N. Santamaria Lancia AA Riedel, S. Thorsteinsson, et al., Drone-based daylight electroluminescence imaging of PV modules, IEEE J. Photovolt. 10 (3) (2020) 872–877.
[101] S. Koch, T. Weber, C. Sobottka, A. Fladung, P. Clemens, Berghold J. Outdoor, Electroluminescence imaging of crystalline photovoltaic modules: comparative study between manual ground-level inspections and drone-based aerial surveys, in: 32nd European Photovoltaic Solar Energy Conference and Exhibition, 2016, pp. 1736–1740.
[102] R. Ebner, S. Zamini, Újvári G. Defect, Analysis in different photovoltaic modules using electroluminescence (EL) and infrared (IR)-thermography, in: 25th European Photovoltaic Solar Energy Conference and Exhibition, 2010, pp. 333–336.
[103] R. Asadpour, D.B. Sulas-Kern, S. Johnston, J. Meydbray, Alam MA. Dark, Lock-in thermography identifies solder bond failure as the root cause of series resistance increase in fielded solar modules, IEEE J. Photovolt. 10 (5) (2020) 1409–1416.
[104] Y. Cao, Y. Dong, Y. Cao, J. Yang, M.Y. Yang, Two-stream convolutional neural network for non-destructive subsurface defect detection via similarity comparison of lock-in thermography signals, NDT Int. 112 (2020) 102246.
[105] S. Fadhel, Efficacité énergétique et surveillance d'un microgrid à courant continu alimenté par des panneaux photovoltaïques, Ph.D. thesis, Université Paris-Sud, 2019.
[106] S.R. Madeti, S.N. Singh, A comprehensive study on different types of faults and detection techniques for solar photovoltaic system, Sol. Energy 158 (2017) 161–185.
[107] A. Mellit, G.M. Tina, S.A. Kalogirou, Fault detection and diagnosis methods for photovoltaic systems: a review, Renew. Sustain. Energy Rev. 91 (2018) 1–17.
[108] Y. Shen, L. Ji, Y. Xie, G. Huang, X. Li, L. Huang, Research landscape and hot topics of rooftop PV: a bibliometric and network analysis, Energy Build. 251 (2021) 111333.
[109] T. Wuest, D. Weimer, C. Irgens, K.D. Thoben, Machine learning in manufacturing: advantages, challenges, and applications, Prod. Manuf. Res. 4 (1) (2016) 23–45.
[110] T. Saheb, T. Saheb, Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works, J. Big Data 7 (1) (2020) 12.
[111] T. Saheb, L. Izadi, Paradigm of IoT big data analytics in the healthcare industry: a review of scientific literature and mapping of research trends, Telemat. Inform. 41 (2019) 70–85.
[112] G. Mao, X. Liu, H. Du, J. Zuo, L. Wang, Way forward for alternative energy research: a bibliometric analysis during 1994–2013, Renew. Sustain. Energy Rev. 48 (2015) 276–286.
[113] D.F. Dominković, J.M. Weinand, F. Scheller, M. D'Andrea, R. McKenna, Reviewing two decades of energy system analysis with bibliometrics, Renew. Sustain. Energy Rev. 153 (2022-01) 111749.
[114] T. Dereli, A. Baykasoglu, K. Altun, A. Durmusoglu, I.B. Türksen, Industrial applications of type-2 fuzzy sets and systems: a concise review, Comput. Ind. 62 (2) (2011 Feb) 125–137.
[115] A. Pritchard, Statistical bibliography or bibliometrics, J. Doc. 25 (4) (1969) 348–349.
[116] S. Kumar, Co-authorship networks: a review of the literature, Aslib J. Inf. Manag. 67 (1) (2015) 55–73.
[117] Y. Kajikawa, J. Yoshikawa, Y. Takeda, K. Matsushima, Tracking emerging technologies in energy research: toward a roadmap for sustainable energy, Technol. Forecast. Soc. Change 75 (6) (2008) 771–782.
[118] B. Dong, G. Xu, X. Luo, Y. Cai, W. Gao, A bibliometric analysis of solar power research from 1991 to 2010, Scientometrics 93 (3) (2012) 1101–1117.
[119] M.A. Reyes-Belmonte, A bibliometric study on integrated solar combined cycles (ISCC), trends and future based on data analytics tools, Sustainability 12 (19) (2020) 8217.
[120] H. Du, N. Li, M.A. Brown, Y. Peng, Y. Shuai, A bibliographic analysis of recent solar energy literatures: the expansion and evolution of a research field, Renew. Energy 66 (C) (2014) 696–706.
[121] I. Zupic, T. Čater, Bibliometric methods in management and organization, Organ. Res. Methods 18 (3) (2015) 429–472.
[122] Al Mamun MA, M.A.K. Azad, M.A. Al Mamun, M. Boyle, Review of flipped learning in engineering education: scientific mapping and research horizon, Educ. Inf. Technol. 27 (1) (2022) 1261–1286.
[123] M. Akinlolu, T.C. Haupt, A bibliometric review of trends in construction safety technology research, in: Proceedings of International Structural Engineering and Construction, vol. 7(2), 2020.
[124] K.W. Qadir, Q. Zafar, N.A. Ebrahim, Z. Ahmad, K. Sulaiman, R. Akram, et al., Methodical review of the literature referred to the dye-sensitized solar cells: bibliometrics analysis and road mapping, Chin. Phys. B 28 (11) (2019 Oct) 118401.
[125] W. Zhang, B. Li, R. Xue, C. Wang, W. Cao, A systematic bibliometric review of clean energy transition: implications for low-carbon development, PLoS ONE 16 (12) (2021) e0261091.
[126] S. Hadke, R. Mishra, A. Werulkar, A bibliometric analysis of different maximum power point tracking methods for photovoltaic systems, Int. J. Trend Sci. Res. Dev. 5 (4) (2021) 1353–1357.
[127] K.C. Garg, P. Sharma, L. Sharma, Bradford's law in relation to the evolution of a field. A case study of solar power research, Scientometrics 27 (2) (1993) 145–156.
[128] A.K. Azad, S. Parvin, Bibliometric analysis of photovoltaic thermal (PV/T) system: from citation mapping to research agenda, Energy Rep. 8 (2022) 2699–2711.

[129] P. Zhang, F. Yan, C. Du, A comprehensive analysis of energy management strategies for hybrid electric vehicles based on bibliometrics, Renew. Sustain. Energy Rev. 48 (2015) 88–104.
[130] A. Calderón, C. Barreneche, K. Hernández-Valle, E. Galindo, M. Segarra, A.I. Fernández, Where is thermal energy storage (TES) research going? – a bibliometric analysis, Sol. Energy 200 (2020) 37–50.
[131] D. Yu, X. He, A bibliometric study for DEA applied to energy efficiency: trends and future challenges, Appl. Energy 268 (C) (2020).
[132] G. Mao, N. Huang, L. Chen, H. Wang, Research on biomass energy and environment from the past to the future: a bibliometric analysis, Sci. Total Environ. 635 (2018) 1081–1090.
[133] H. Kim, H. Choi, H. Kang, J. An, S. Yeom, T. Hong, A systematic review of the smart energy conservation system: from smart homes to sustainable smart cities, Renew. Sustain. Energy Rev. 140 (2021) 110755.
[134] M.W. Nielsen, L. Börjesonb, Gender diversity in the management field: does it matter for research outcomes?, Res. Policy 48 (7) (2019) 1617–1632.
[135] H. Jiang, M. Qiang, P. Lin, A topic modeling based bibliometric exploration of hydropower research, Renew. Sustain. Energy Rev. 57 (2016) 226–237.
[136] D.A. McFarland, D. Ramage, J. Chuang, J. Heer, C.D. Manning, D. Jurafsky, Differentiating language usage through topic models, Poetics 41 (6) (2013) 607–625.
[137] L.G. Nichols, A topic model approach to measuring interdisciplinarity at the national science foundation, Scientometrics 100 (3) (2014) 741–754.
[138] T. Saheb, T. Saheb, D.O. Carpenter, Mapping research strands of ethics of artificial intelligence in healthcare: a bibliometric and content analysis, Comput. Biol. Med. 135 (2021) 104660.
[139] M.J. Page, J.E. McKenzie, P. Bossuyt, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, et al., The prisma 2020 statement: an updated guideline for reporting systematic reviews, Syst Rev. 10 (89) (2021).
[140] M. Madhavan, S. Wangtueai, M.A. Sharafuddin, T. Chaichana, The precipitative effects of pandemic on open innovation of SMEs: a scientometrics and systematic review of industry 4.0 and industry 5.0, J. Open Innov.: Technol. Mark. Complex. 8 (3) (2022) 152.
[141] M.K. Lim, Y. Li, C. Wang, M.L. Tseng, A literature review of blockchain technology applications in supply chains: a comprehensive analysis of themes, methodologies and industries, Comput. Ind. Eng. 154 (2021) 107133.
[142] Y. Hong, D.W.M. Chan, A.P.C. Chan, J.FY. Yeung, Critical analysis of partnering research trend in construction journals, J. Manag. Eng. 28 (2) (2012) 82–95.
[143] M.R. Hosseini, M. Maghrebi, A. Akbarnezhad, I. Martek, M. Arashpour, Analysis of citation networks in building information modeling research, J. Constr. Eng. Manage. 144 (8) (2018) 04018064.
[144] X. Zhao, S. Wang, Wang X. Characteristics, Trends of research on new energy vehicle reliability based on the web of science, Sustainability 10 (10) (2018) 3560.
[145] A.K. Kaushik, Rahman Z. Perspectives, Dimensions of consumer innovativeness: a literature review and future agenda, J. Int. Consum. Mark. 26 (3) (2014) 239–263.
[146] N. Rakshikar, Zotero: an ultimate citation management tool for researchers and academicians, in: UGC Sponsored Two Day's National Seminar on Emerging Trends in Library Technology, 2015, pp. 1–6.
[147] C.H. Kim, H. Kim, Y.H. Ko, S.H. Byun, R.K. Aggarwal, A.T. Johns, A novel fault-detection technique of high-impedance arcing faults in transmission lines using the wavelet transform, IEEE Trans. Power Deliv. 17 (4) (2002) 921–929.
[148] D. Tranfield, D. Denyer, P. Smart, Towards a methodology for developing evidence-informed management knowledge by means of systematic review, Br. J. Manag. 14 (3) (2003) 207–222.
[149] J. Keiser, J. Utzinger, Trends in the core literature on tropical medicine: a bibliometric analysis from 1952-2002, Scientometrics 62 (3) (2005 March) 351–365.
[150] G. Zhang, S. Xie, Y.S. Ho, A bibliometric analysis of world volatile organic compounds research trends, Scientometrics 83 (2010 05) 477–492.
[151] A. Bjurström, M. Polk, Climate change and interdisciplinarity: a co-citation analysis of IPCC third assessment report, Scientometrics 87 (2011 06) 525–550.
[152] F. Zhou, H.C. Guo, Y.S. Ho, C.Z. Wu, Scientometric analysis of geostatistics using multivariate methods, Scientometrics 73 (2007 12) 265–279.
[153] L. Van Eck NJ Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics 84 (2010) 523–538.
[154] I. Borg, P. Groenen, Modern Multidimensional Scaling, 2nd ed., Springer, Berlin, 2005.
[155] N.J. Van Eck, L. Waltman, VOS: A new method for visualizing similarities between objects, in: R. Decker, H.J. Lenz (Eds.), Advances in Data Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 299–306.
[156] N.J. Van Eck, L. Waltman, Bibliometric mapping of the computational intelligence field, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 15 (05) (2007) 625–645.
[157] N.J. Van Eck, L. Waltman, J. van den Berg, U. Kaymak, Visualizing the computational intelligence field, Application Notes, IEEE Comput. Intell. Mag. 1 (4) (2006) 6–10.
[158] L. Van Eck NJ Waltman, R. Dekker, J. van den Berg, A comparison of two techniques for bibliometric mapping: multidimensional scaling and VOS, J. Am. Soc. Inf. Sci. Technol. 61 (12) (2010) 2405–2416.
[159] G.S. Davidson, B.N. Wylie, K.W. Boyack, Cluster stability and the use of noise in interpretation of clustering, in: IEEE Symposium on Information Visualization, INFOVIS 2001, 2001, pp. 23–30.
[160] R. Klavans, K.W. Boyack, Quantitative evaluation of large maps of science, Scientometrics 68 (2006) 475–499.
[161] A. Kopcsa, E. Schiebel, Science and technology mapping: a new iteration model for representing multidimensional relationships, J. Am. Soc. Inf. Sci. 49 (1) (1998) 7–17.
[162] T. Kamada, S. Kawai, An algorithm for drawing general undirected graphs, Inf. Process. Lett. 31 (1) (1989) 7–15.
[163] T.M.J. Fruchterman, E.M. Reingold, Graph drawing by force-directed placement, Softw. Pract. Exp. 21 (11) (1991) 1129–1164.
[164] S.G.F. de Moya-Anegón, B. Vargas-Quesada, Z. Chinchilla-Rodríguez, E. Corera-Álvarez, F.J. Munoz-Fernández, V. Herrero-Solana, Visualizing the marrow of science, J. Am. Soc. Inf. Sci. Technol. 58 (14) (2007) 2167–2179.
[165] B. Vargas-Quesada, F. Moya-Anegon, Visualizing the Structure of Science, Springer, 2007.
[166] R.W. Schvaneveldt, D.W. Dearholt, F.T. Durso, Graph theoretic foundations of pathfinder networks, Comput. Math. Appl. 15 (4) (1988) 337–345.
[167] C. Chen, Mapping Scientific Frontiers, Springer, Berlin, 2003.
[168] W. De Nooy, A. Mrvar, Batagelj V. Exploratory, Social Network Analysis with Pajek: Revised and Expanded Edition for Updated Software, 3rd ed., Structural Analysis in the Social Sciences, Cambridge University Press, 2018.
[169] J. Wang, X. Zhao, X. Guo, B. Li, Analyzing the research subjects and hot topics of power system reliability through the Web of Science from 1991 to 2015, Renew. Sustain. Energy Rev. 82 (2018) 700–713.
[170] J. Wang, M.K. Lim, C. Wang, M.L. Tseng, The evolution of the Internet of Things (IoT) over the past 20 years, Comput. Ind. Eng. 155 (2021) 107174.
[171] W. Pan, L. Jian, T. Liu, Grey system theory trends from 1991 to 2018: a bibliometric analysis and visualization, Scientometrics 121 (3) (2019 December) 1407–1434.
[172] Y. Bai, H. Li, Y. Liu, Visualizing research trends and research theme evolution in E-learning field: 1999-2018, Scientometrics 126 (2020 11).
[173] A. Huang, Similarity measures for text document clustering, in: Proceedings of the 6th New Zealand Computer Science Research Student Conference, 2008 01.
[174] S. Shtovba, M. Petrychko, Jaccard index-based assessing the similarity of research fields in dimensions, in: 1st International Workshop on Digital Content & Smart Multimedia, 2020, pp. 1–12.
[175] H.P.F. Peters, A.F.J. van Raan, Co-word-based science maps of chemical engineering. Part I: representations by direct multidimensional scaling, Res. Policy 22 (1) (1993) 23–45.
[176] A. Rip, J.P. Courtial, Co-word maps of biotechnology: an example of cognitive scientometrics, Scientometrics 6 (1984) 381–400.

[177] M. Zitt, E. Bassecoulard, Y. Okubo, Shadows of the past in international cooperation: collaboration profiles of the top five producers of science, Scientometrics 47 (2000) 627–657.
[178] L. Van Eck NJ Waltman, How to normalize cooccurrence data? An analysis of some well-known similarity measures, J. Am. Soc. Inf. Sci. Technol. 60 (8) (2009) 1635–1651.
[179] N.J. Van Eck, L. Waltman, VOSviewer Manual, Universiteit Leiden, 2022.
[180] F. Thonon, R. Boulkedid, T. Delory, S. Rousseau, M. Saghatchian, W. van Harten, et al., Measuring the outcome of biomedical research: a systematic literature review, PLoS ONE 10 (4) (2015 04) 1–14.
[181] C. García-Villar, J.M. García-Santos, Bibliometric indicators to evaluate scientific activity, Radiología (English Ed.) 63 (3) (2021) 228–235.
[182] Joshi M. Bibliometric, Indicators for evaluating the quality of scientific publications, J. Contemp. Dent. Pract. 15 (2014 08) 258–262.
[183] L. Waltman, N.J. Van Eck, T.N. van Leeuwen, MS Visser, A.F.J. van Raan, Towards a new crown indicator: some theoretical considerations, J. Informetr. 5 (1) (2011) 37–47.
[184] P.N. Tyrrell, A.R. Moody, J.O.C. Moody, N. Ghiam, Departmental h-index: evidence for publishing less?, Can. Assoc. Radiol. J. 68 (1) (2017) 10–15.
[185] F. Van Rijnsoever, L. Hessels, How academic researchers select collaborative research projects: a choice experiment, J. Technol. Transf. 46 (2021 12) 1–32.
[186] S. Bansal, S. Mahendiratta, D.S. Kumar, A. Singh, A. Prakash, B. Medhi, Collaborative research in modern era: need and challenges, Indian J. Pharmacol. 51 (2019 05) 137.
[187] F. Ahmed, M.T. Fattani, S.R. Ali, R.N. Enam, Strengthening the bridge between academic and the industry through the academia-industry collaboration plan design model, Front. Psychol. (2022) 13.
[188] S. Ankrah, O. AL-Tabbaa, Universities–industry collaboration: a systematic review, Scand. J. Manag. 31 (3) (2015) 387–408.
[189] B.J. Pesta, Bibliometric analysis across eight years 2008–2015 of intelligence articles: an updating of Wicherts (2009), Intelligence 67 (2018) 26–32.
[190] G. Chen, L. Xiao, Selecting publication keywords for domain analysis in bibliometrics: a comparison of three methods, J. Informetr. 10 (1) (2016) 212–223.
[191] H. Yin, F. Zhang, X. Yang, X. Meng, Y. Miao, M.S. Noor Hussain, et al., Research trends of artificial intelligence in pancreatic cancer: a bibliometric analysis, Front. Oncol. 12 (2022) 1–13.
[192] S. Jalal, Co-authorship and co-occurrences analysis using BibliometrixR package: a casestudy of India and Bangladesh, Ann. Libr. Inf. Stud. 06 (2019) 57–64.
[193] A.Z.M. Sayeed, A bibliometric and co-occurrence analysis of COVID-19–related literature published between December 2019 and June 2020, Sci. Ed. 8 (1) (2021) 57–63.
[194] I. Ali, M. Balta, T. Papadopoulos, Social media platforms and social enterprise: bibliometric analysis and systematic review, Int. J. Inf. Manag. 69 (2023) 102510.
[195] S. Noor, Y. Guo, S.H.H. Shah, Butt A. Bibliometric, Analysis of social media as a platform for knowledge management, Int. J. Knowl. Manag. 16 (2020 07) 33–51.
[196] D.J. de Solla Price, Networks of scientific papers, Science 149 (3683) (1965) 510–515.
[197] M.M. Kessler, Bibliographic coupling between scientific papers, Am. Doc. 14 (1) (1963) 10–25.
[198] H. Small, Co-citation in the scientific literature: a new measure of the relationship between two documents, J. Am. Soc. Inf. Sci. 24 (1973 07) 265–269.
[199] L. Waltman, A review of the literature on citation impact indicators, J. Informetr. 10 (2) (2016) 365–391.
[200] M. Gauffriau, P. Larsen, I. Maye, A. Roulin-Perriard, Ins M. Publication, Cooperation and productivity measures in scientific research, Scientometrics 73 (2007 11) 175–214.
[201] A. Perianes-Rodriguez, L. Waltman, N.J. Van Eck, Constructing bibliometric networks: a comparison between full and fractional counting, J. Informetr. 10 (4) (2016) 1178–1195.
[202] D.W. Aksnes, J.W. Schneider, M. Gunnarsson, Ranking national research systems by citation indicators. A comparative analysis using whole and fractionalised counting methods, J. Informetr. 6 (1) (2012) 36–43.
[203] L. Waltman, N.J. Van Eck, Field-normalized citation impact indicators and the choice of an appropriate counting method, J. Informetr. 9 (4) (2015) 872–894.
[204] A. Pouris, A bibliometric assessment of energy research in South Africa, South Afr. J. Sci. 112 (2016).
[205] L. Romero, E. Portillo-Salido, Trends in Sigma-1 receptor research: a 25-year bibliometric analysis, Front. Pharmacol. 10 (2019 05).
[206] S. Naveen Venkatesh, V. Sugumaran, Machine vision based fault diagnosis of photovoltaic modules using lazy learning approach, Measurement 191 (2022) 110786.
[207] D. Adhya, S. Chatterjee, A.K. Chakraborty, Performance assessment of selective machine learning techniques for improved PV array fault diagnosis, Sustain. Energy, Grids Netw. 29 (2022) 100582.
[208] M. Ibrahim, A. Alsheikh, F.M. Awaysheh, Alshehri MD. Machine, Learning schemes for anomaly detection in solar power plants, Energies 15 (3) (2022).
[209] V.S.P. Machina, K.S. Suprabhath, Madichetty S. Fault, Detection in solar photovoltaic systems during winter season- a deep learning approach, in: 2022 IEEE Texas Power and Energy Conference (TPEC), 2022, pp. 1–6.
[210] R. Janarthanan, R.U. Maheshwari, P.K. Shukla, P.K. Shukla, S. Mirjalili, M. Kumar, Intelligent detection of the PV faults based on artificial neural network and type 2 fuzzy systems, Energies 14 (20) (2021).
[211] X. Wang, B. Yang, Q. Liu, T. Jin, C. Chen, Collaboratively diagnosing IGBT open-circuit faults in photovoltaic inverters: a decentralized federated learning-based method, in: IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, 2021, pp. 1–6.
[212] Y. Liu, K. Ding, J. Zhang, Y. Li, Z. Yang, W. Zheng, et al., Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with I-V curves, Energy Convers. Manag. 245 (2021) 114603.
[213] Z. Abdullah-Vetter, Y. Buratti, P. Dwivedi, A. Sowmya, T. Trupke, Z. Hameiri, Localization of defects in solar cells using luminescence images and deep learning, in: 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), 2021, 0745-9.
[214] S. Tchoketch Kebir, N. Cheggaga, A. Ilinca, Boulouma S. An, Efficient neural network-based method for diagnosing faults of PV array, Sustainability 13 (11) (2021).
[215] P. Haidari, A. Hajiahmad, A. Jafari, A. Nasiri, Deep learning-based model for fault classification in solar modules using infrared images, Sustain. Energy Technol. Assess. 52 (2022) 102110.
[216] A. Jarndal, M.S. Mahmoud, Mohammad Abbas O. Fault, Detection and identification based on image processing and deep learning, in: 2022 Advances in Science and Engineering Technology International Conferences (ASET), 2022, pp. 1–4.
[217] J. Fioresi, D.J. Colvin, R. Frota, R. Gupta, M. Li, H.P. Seigneur, et al., Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images, IEEE J. Photovolt. 12 (1) (2022) 53–61.
[218] M. Bindi, F. Corti, I. Aizenberg, F. Grasso, G.M. Lozito, A. Luchetta, et al., Machine learning-based monitoring of DC-DC converters in photovoltaic applications, Algorithms 15 (3) (2022).
[219] C. Barker, S. Cipkar, T. Lavigne, C. Watson, Azzouz M. Real-Time, Nuisance fault detection in photovoltaic generation systems using a fine tree classifier, Sustainability 13 (4) (2021).
[220] R.G. Vieira, M. Dhimish, F.M.U. de Araújo, M.I.S. Guerra, PV module fault detection using combined artificial neural network and Sugeno fuzzy logic, Electronics 9 (12) (2020).
[221] M. Dhimish, V. Holmes, B. Mehrdadi, M. Dales, Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection, Renew. Energy 117 (2018) 257–274.
[222] J.H. Shin, J.O. Kim, On-line diagnosis and fault state classification method of photovoltaic plant, Energies 13 (17) (2020).

[223] A.E. Lazzaretti, C.H. da Costa, M.P. Rodrigues, G.D. Yamada, G. Lexinoski, G.L. Moritz, et al., A monitoring system for online fault detection and classification in photovoltaic plants, Sensors 20 (17) (2020).
[224] L. Bommes, M. Hoffmann, C. Buerhop-Lutz, T. Pickel, J. Hauch, C. Brabec, et al., Anomaly Detection in IR Images of PV Modules using Supervised Contrastive Learning, 2021.
[225] A.A. Al-katheri, E.A. Al-Ammar, M. Alotaibi, Ghazi GA. Artificial, Neural network application for faults detection in PV systems, in: 2022 IEEE Delhi Section Conference (DELCON), 2022, pp. 1–6.
[226] K. Dhibi, M. Mansouri, K. Abodayeh, K. Bouzrara, H. Nounou, Nounou M. Interval-Valued, Reduced ensemble learning based fault detection and diagnosis techniques for uncertain grid-connected PV systems, IEEE Access 10 (2022) 47673–47686.
[227] F. Harrou, A. Saidi, Y. Sun, S. Khadraoui, Monitoring of photovoltaic systems using improved kernel-based learning schemes, IEEE J. Photovolt. 11 (3) (2021) 806–818.
[228] H.P.C. Hwang, C.C.Y. Ku, J.CC. Chan, Detection of malfunctioning photovoltaic modules based on machine learning algorithms, IEEE Access 9 (2021) 37210–37219.
[229] Y.Y. Hong, R.A. Pula, Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network, Energy 246 (2022) 123391.
[230] N.C. Yang, Ismail H. Robust, Intelligent learning algorithm using random forest and modified-independent component analysis for PV fault detection: in case of imbalanced data, IEEE Access 10 (2022) 41119–41130.
[231] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, W. Lin, Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics, Appl. Energy 204 (2017) 912–931.
[232] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, et al., Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents, Energy Convers. Manag. 178 (2018) 250–264.
[233] J.E. Hirsch, An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. 102 (46) (2005) 16569–16572.
[234] V.P. Guerrero-Bote, F. Moya-Anegón, A further step forward in measuring journals' scientific prestige: the SJR2 indicator, J. Informetr. 6 (4) (2012) 674–688.
[235] S. Radhakrishnan, S. Erbis, J.A. Isaacs, S. Kamarthi, Novel keyword co-occurrence network-based methods to Foster systematic reviews of scientific literature, PLoS ONE 12 (12) (2017 03) 1–16.
[236] X. Zhou, M. Zhou, D. Huang, L. Cui, A probabilistic model for co-occurrence analysis in bibliometrics, J. Biomed. Inform. 128 (2022) 104047.
[237] L. Leydesdorff, L. Vaughan, Co-occurrence matrices and their applications in information science: extending ACA to the Web environment, J. Am. Soc. Inf. Sci. Technol. 57 (12) (2006) 1616–1628.
[238] P. Ahlgren, B. Jarneving, R. Rousseau, Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient, J. Am. Soc. Inf. Sci. Technol. 54 (6) (2003) 550–560.
[239] C. Lei, L. Wei, Y. Lei, Z. Han, H. Yuefang, H. Yingna, et al., Development of a text mining system based on the co-occurrence of bibliographic items in literature databases, Data Anal. Knowl. Discov. 24 (8) (2008) 70–75.
[240] K. Yang, L. Wang, G. Yang, X. Jiang, Research hotspots and trends in nursing education from 2014 to 2020: a co-word analysis based on keywords, J. Adv. Nurs. 78 (3) (2022) 787–798.
[241] T. Millington, S. Luz, Analysis and classification of word co-occurrence networks from Alzheimer's patients and controls, Front. Comput. Sci. 3 (2021).
[242] J. Luo, S. Leng, Y. Bai, Food supply chain safety research trends from 1997 to 2020: a bibliometric analysis, Front. Public Health 9 (2022).
[243] A. Bhuyan, K. Sanguri, H. Sharma, Improving the keyword co-occurrence analysis: an integrated semantic similarity approach, in: 2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2021, pp. 482–487.
[244] R. Mihalcea, Tarau P. TextRank, Bringing order into text, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pp. 404–411.
[245] L. Feng, Q. Chen, Bibliometric analysis of the synthesis of nanocatalyst (1999–2018), IOP Conf. Ser. Earth Environ. Sci. 558 (2020 09) 042042.
[246] A. Kirby, Exploratory bibliometrics: using VOSviewer as a preliminary research tool, Publications 11 (1) (2023) 1–14.
[247] F. Turatto, E. Mazzalai, F. Pagano, G. Migliara, P. Villari, C. De Vito, A systematic review and bibliometric analysis of the scientific literature on the early phase of COVID-19 in Italy, Front. Public Health 9 (2021 06).
[248] D. Kalibatiene, J. Miliauskaitė, A systematic mapping with bibliometric analysis on information systems using ontology and fuzzy logic, Appl. Sci. 11 (7) (2021) 1–20.
[249] Y. Yu, Y. Li, Z. Zhang, Z. Gu, H. Zhong, Q. Zha, et al., A bibliometric analysis using VOSviewer of publications on COVID-19, Ann. Transl. Med. 8 (13) (2020) 1–11.
[250] J.K. Tamala, E.I. Maramag, K.A. Simeon, J.J. Ignacio, A bibliometric analysis of sustainable oil and gas production research using VOSviewer, Clean. Eng. Technol. 7 (2022) 100437.
[251] S.I. Abdelwahab, M.M.E. Taha, S.S. Moni, A.A. Alsayegh, Bibliometric mapping of solid lipid nanoparticles research (2012–2022) using VOSviewer, Med. Nov. Technol. Devices 17 (2023) 100217.
[252] Y. Yan, Z. Yang, H. Li, C. Jiang, A bibliometric analysis of research on acoustic emission for nondestructive testing, IOP Conf. Ser., Mater. Sci. Eng. 1167 (1) (2021 jul) 012009.
[253] H. Liang, S. Zhang, Y. Su, The structure and emerging trends of construction safety management research: a bibliometric review, Int. J. Occup. Saf. Ergon. 26 (3) (2020) 469–488.
[254] P. Domingos, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. the Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World, Basic Books, New York, NY, US, 2015, Pages: xxi, 329.
[255] S. Rodrigues, H.G. Ramos, Morgado-Dias F. Machine, Learning in PV fault detection, diagnostics and prognostics: a review, in: 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), IEEE, Washington, DC, 2017, pp. 3178–3183.
[256] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, Antonanzas-Torres F. de Pison FJM, Review of photovoltaic power forecasting, Sol. Energy 136 (2016) 78–111.
[257] K. Dhibi, R. Fezai, M. Mansouri, M. Trabelsi, A. Kouadri, K. Bouzara, et al., Reduced kernel random forest technique for fault detection and classification in grid-tied PV systems, IEEE J. Photovolt. 10 (6) (2020) 1864–1871.
[258] S.S. Madani, A. Abbaspour, M. Beiraghi, P.Z. Dehkordi, A.M. Ranjbar, Islanding detection for PV and DFIG using decision tree and AdaBoost algorithm, in: 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), 2012, pp. 1–8.
[259] R. Fazai, K. Abodayeh, M. Mansouri, M. Trabelsi, H. Nounou, M. Nounou, et al., Machine learning-based statistical testing hypothesis for fault detection in photovoltaic systems, Sol. Energy 190 (2019) 405–413.
[260] M. Hajji, M.F. Harkat, A. Kouadri, K. Abodayeh, M. Mansouri, H. Nounou, et al., Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems, Eur. J. Control 59 (2021) 313–321.
[261] M. Hopwood, T. Gunda, H. Seigneur, J. Walters, An assessment of the value of principal component analysis for photovoltaic IV trace classification of physically-induced failures, in: 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), 2020, 0798-802.
[262] H. Momeni, N. Sadoogi, M. Farrokhifar, Gharibeh HF. Fault, Diagnosis in photovoltaic arrays using GBSSL method and proposing a fault correction system, IEEE Trans. Ind. Inform. 16 (8) (2020) 5300–5308.
[263] K.A.K. Niazi, W. Akhtar, H.A. Khan, Y. Yang, S. Athar, Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier, Sol. Energy 190 (2019) 34–43.
[264] W. He, D. Yin, K. Zhang, X. Zhang, Zheng J. Fault, Detection and diagnosis method of distributed photovoltaic array based on fine-tuning Naive Bayesian model, Energies 14 (14) (2021).

[265] H.R. Baghaee, D. Mlakić, S. Nikolovski, Dragicević T. Support, Vector machine-based islanding and grid fault detection in active distribution networks, IEEE J. Emerg. Sel. Top. Power Electron. 8 (3) (2020) 2385–2403.
[266] F. Harrou, B. Taghezouit, Y. Sun, Improved *k*NN-based monitoring schemes for detecting faults in PV systems, IEEE J. Photovolt. 9 (3) (2019) 811–821.
[267] M.U. Ali, H.F. Khan, M. Masud, K.D. Kallu, A. Zafar, A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography, Sol. Energy 208 (2020) 643–651.
[268] Sepúlveda Oviedo EH, L. Travé-Massuyès, A. Subias, M. Pavlov, C. Alonso, DTW k-means clustering for fault detection in photovoltaic modules, in: XI Congreso Internacional de Ingeniería Mecánica, Mecatrónica y Automatización 2023, Cartagena, Colombia, 2023, pp. 1–2.
[269] E. Garoudja, A. Chouder, K. Kara, S. Silvestre, An enhanced machine learning based approach for failures detection and diagnosis of PV systems, Energy Convers. Manag. 151 (2017) 496–513.
[270] Z. Chen, Y. Chen, L. Wu, S. Cheng, P. Lin, Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions, Energy Convers. Manag. 198 (2019) 111793.
[271] A.H. Mohamed, A.M. Nassar, New algorithm for fault diagnosis of photovoltaic energy systems, Int. J. Comput. Appl. 114 (9) (2015 March) 26–31.
[272] H. Lin, Z. Chen, L. Wu, P. Lin, S. Cheng, On-line monitoring and fault diagnosis of PV array based on BP neural network optimized by genetic algorithm, in: A. Bikakis, X. Zheng (Eds.), Multi-Disciplinary Trends in Artificial Intelligence, Springer International Publishing, Cham, 2015, pp. 102–112.
[273] S. Rodrigues, G. Mütter, H.G. Ramos, Morgado-Dias F. Machine, Learning photovoltaic string analyzer, Entropy 22 (2) (2020).
[274] S.A. Memon, Q. Javed, W.G. Kim, Z. Mahmood, U. Khan, M. Shahzad, A machine-learning-based robust classification method for PV panel faults, Sensors 22 (21) (2022) 1–14.
[275] S.A. Zaki, H. Zhu, M.A. Fakih, A.R. Sayed, J. Yao, Deep-learning–based method for faults classification of PV system, IET Renew. Power Gener. 15 (1) (2021) 193–205.
[276] N. Kellil, A. Aissat, A. Mellit, Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions, Energy 263 (2023) 125902.
[277] M. Benghanem, A. Mellit, Moussaoui C. Embedded, Hybrid model (CNN and ML) for fault diagnosis of photovoltaic modules using thermographic images, Sustainability 15 (10) (2023) 1–20.
[278] A. Sohail, N. Ul Islam, A. Ul Haq, S. Ul Islam, I. Shafi, J. Park, Fault detection and computation of power in PV cells under faulty conditions using deep-learning, Energy Rep. 9 (2023) 4325–4336.
[279] S. Boubaker, S. Kamel, N. Ghazouani, A. Mellit, Assessment of machine and deep learning approaches for fault diagnosis in photovoltaic systems using infrared thermography, Remote Sens. 15 (6) (2023) 1–21.
[280] H. Bakır, F.A. Kuzhippallil, A. Merabet, Automatic detection of deteriorated photovoltaic modules using IRT images and deep learning (CNN, LSTM) strategies, Eng. Fail. Anal. 146 (2023) 107132.
[281] B. Basnet, H. Chun, Bang J. An, Intelligent fault detection model for fault detection in photovoltaic systems, J. Sens. 1 (6960328) (2020) 1–11.
[282] N.V. Sridharan, J.V. Joseph, S. Vaithiyanathan, Aghaei M. Weightless, Neural network-based detection and diagnosis of visual faults in photovoltaic modules, Energies 16 (15) (2023) 1–17.
[283] A. Mellit, O. Herrak, C. Rus Casas, A. Massi Pavan, A machine learning and Internet of things-based online fault diagnosis method for photovoltaic arrays, Sustainability 13 (23) (2021) 1–14.
[284] S. Voutsinas, D. Karolidis, I. Voyiatzis, M. Samarakou, Development of a machine-learning-based method for early fault detection in photovoltaic systems, J. Eng. Appl. Sci. 70 (2023 04) 1–17.
[285] S. Lu, T. Sirojan, B.T. Phung, D. Zhang, E. Ambikairajah, DA-DCGAN: an effective methodology for DC series arc fault diagnosis in photovoltaic systems, IEEE Access 7 (2019) 45831–45840.
[286] A.S. Spanias, Solar energy management as an Internet of Things (IoT) application, in: 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), 2017, pp. 1–4.
[287] N. Barbosa Roa, L. Travé-Massuyès, Grisales-Palacio VH. DyClee, Dynamic clustering for tracking evolving environments, Pattern Recognit. 94 (2019) 162–186.
[288] A. Banerjee, D. Singh, S. Sahana, Nath I. Chapter, 3 - Impacts of metaheuristic and swarm intelligence approach in optimization, in: S. Mishra, H.K. Tripathy, P.K. Mallick, A.K. Sangaiah, G.S. Chae (Eds.), Cognitive Big Data Intelligence with a Metaheuristic Approach. Cognitive Data Science in Sustainable Computing, Academic Press, 2022, pp. 71–99.
[289] G. Devika, A.G. Karegowda, Bio-inspired optimization: algorithm, analysis and scope of application, in: M.A. Aceves-Fernández (Ed.), Swarm Intelligence, IntechOpen, Rijeka, 2023, pp. 1–38.
[290] J. Pereira, Silveira M. Unsupervised, Anomaly detection in energy time series data using variational recurrent autoencoders with attention, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1275–1282.
[291] Y. Hu, V.Y. Gunapati, P. Zhao, D. Gordon, N.R. Wheeler, M.A. Hossain, et al., A nonrelational data warehouse for the analysis of field and laboratory data from multiple heterogeneous photovoltaic test sites, IEEE J. Photovolt. 7 (1) (2017) 230–236.
[292] Y. Zhao, Q. Liu, D. Li, D. Kang, Q. Lv, Shang L. Hierarchical, Anomaly detection and multimodal classification in large-scale photovoltaic systems, IEEE Trans. Sustain. Energy 10 (3) (2019) 1351–1361.
[293] Q. Liu, Y. Zhao, Y. Zhang, D. Kang, Q. Lv, L. Shang, Hierarchical context-aware anomaly diagnosis in large-scale PV systems using SCADA data, in: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), 2017, pp. 1025–1030.
[294] M. Benninger, M. Hofmann, M. Liebschner, Anomaly detection by comparing photovoltaic systems with machine learning methods, in: NEIS 2020; Conference on Sustainable Energy Supply and Energy Storage Systems, 2020, pp. 1–6.
[295] P. Xi, P. Lin, Y. Lin, H. Zhou, S. Cheng, Z. Chen, et al., Online fault diagnosis for photovoltaic arrays based on Fisher discrimination dictionary learning for sparse representation, IEEE Access 9 (2021) 30180–30192.
[296] Sepúlveda Oviedo EH, L. Travé-Massuyès, A. Subias, M. Pavlov, C. Alonso, Detection and classification of faults aimed at preventive maintenance of PV systems, in: XI Congreso Internacional de Ingeniería Mecánica, Mecatrónica y Automatización 2023, Universidad Nacional de Colombia, Cartagena, Colombia, 2023, pp. 1–2.
[297] IEC, Photovoltaic System Performance Monitoring—Guidelines for Measurement, Data Exchange and Analysis, International Standard IEC 61724, 1998.
[298] G. Blaesser, D. Munro, Guidelines for the Assessment of Photovoltaic Plants Document a Photovoltaic System Monitoring, Research Centre Joint, 1995.
[299] S. Kurtz, J. Newmiller, A. Kimber, R. Flottemesch, E. Riley, T. Dierauf, et al., Analysis of Photovoltaic System Energy Performance Evaluation Method, National Renewable Energy Lab. (NREL), Golden, CO, USA, 2013.