Journal of Infrastructure, Policy and Development 2025, 9(2), 9940.
https://doi.org/10.24294/jipd9940

Article

Machine learning predictions for fault detections in solar photovoltaic system: A bibliographic outlook

Oliver O. Apeh*, Nnamdi I. Nwulu

Centre for Cyber-Physical Food, Energy and Water Systems (C-C-P-F-E-W-S), University of Johannesburg, Johannesburg 2006, South Africa
*** Corresponding author:** Oliver O. Apeh, olivera@uj.ac.za

CITATION

Apeh OO, Nwulu NI. (2025). Machine learning predictions for fault detections in solar photovoltaic system: A bibliographic outlook. Journal of Infrastructure, Policy and Development. 9(2): 9940. https://doi.org/10.24294/jipd9940

ARTICLE INFO

Received: 28 October 2024
Accepted: 17 February 2025
Available online: 12 March 2025

Abstract: Photovoltaic systems have shown significant attention in energy systems due to the recent machine learning approach to addressing photovoltaic technical failures and energy crises. A precise power production analysis is utilized for failure identification and detection. Therefore, detecting faults in photovoltaic systems produces a considerable challenge, as it needs to determine the fault type and location rapidly and economically while ensuring continuous system operation. Thus, applying an effective fault detection system becomes necessary to moderate damages caused by faulty photovoltaic devices and protect the system against possible losses. The contribution of this study is in two folds: firstly, the paper presents several categories of photovoltaic systems faults in literature, including line-to-line, degradation, partial shading effect, open/close circuits and bypass diode faults and explores fault discovery approaches with specific importance on detecting intricate faults earlier unexplored to address this issue; secondly, VOSviewer software is presented to assess and review the utilization of machine learning within the solar photovoltaic system sector. To achieve the aims, 2258 articles retrieved from Scopus, Google Scholar, and ScienceDirect were examined across different machine learning and energy-related keywords from 1990 to the most recent research papers on 14 January 2025. The results emphasise the efficiency of the established methods in attaining fault detection with a high accuracy of over 98%. It is also observed that considering their effortlessness and performance accuracy, artificial neural networks are the most promising technique in finding a central photovoltaic system fault detection. In this regard, an extensive application of machine learning to solar photovoltaic systems could thus clinch a quicker route through sustainable energy production.

Keywords: VOSviewer analysis; machine learning; solar photovoltaic system; prediction; bibliometric outlooks

1. Introduction

The energy crisis is becoming more challenging in the global community due to many countries' reliance on fossil fuels for power generation. Unfortunately, the energy from fossil fuels is limited and toxic. It is associated with several negative influences, including environmental hazards and damage to the health of humans and the ecosystem (Naderi et al., 2020). Up till now, at least 80% of power generation is still handled by "unclean" coal globally, unleashing a high percentage of carbon (iv) oxide (CO_2), Sulphur (iv) (SO_2), and Nitrogen oxide (NO_X) into the atmosphere. The amount of electricity produced from coal in 2014 was about 232 Twh, accounting for 2.3% of global coal power production (Jain and Jain, 2017). Coal mining and its application adversely affect all three environmental features, specifically water, air, and land. The yearly emission of CO_2 is about 437.37 Mt or 8.10 t CO_2 per capita, making it the highest greenhouse gas emission globally (Venkatakrishnan et al., 2023).

Due to fossil utilization, these challenges have prompted scientists and researchers to hunt for modern technology that could safely exploit energy sources. Hence, renewable energy resources are growing in popularity as a prospective alternative for future energy supply across the universe (Apeh and Nwulu, 2025).

Renewable energy providers, such as solar energy, have been suggested as options for cleaner power generation (Apeh, et al., 2022; Olorunfemi et al., 2023). The benefits of solar PV systems are enormous (Sohani et al., 2020a). Thus, utilizing this energy for several appliances has rapidly increased in developed and developing countries within the last decade. It is presumed to have more advancement in the near future (Shahverdian et al., 2021). Among the several renewable technologies to be utilized for energy generation, solar PV systems significantly contribute to the global market value (Apeh, et al., 2022). In solar PV systems, the output relies on geographical sites, cell technology, and environmental factors, including temperature and global irradiance (Apeh et al., 2021; Maluta and Mulaudzi, 2018). Several research works have been piloted to assess PV performance factors, where temperature and solar radiation are the most selected variables (Cuce, et al., 2013).

Generally, meteorological stations require high construction and maintenance costs and are not commonly available (Jia et al., 2022). For instance, 1798 meteorological stations in Turkey are available as of 2020, and just 129 are recording solar radiation data as of 2012 (Ağbulut et al., 2021). Also, only 756 stations are installed in China, with just 122 capable of measuring solar radiation data (Sireci, 2006; Zang et al., 2012). Weather parameters were installed to develop a prediction model to forecast an installed PV system’s yearly power generation yield and performance ratio (PR) using three environmental input parameters: ambient air temperature, wind speed, and solar irradiance (Gopi et al., 2022). Similarly, several thermal, optical, and electrical models have been established in the literature to simulate a definite feature of a PV system. The models have been utilized to check the result of a module, assess the electrical yield PV production, suggest new algorithms for MPPT, and examine faults in such systems (Gholami et al., 2023).

Fault detection in solar modules is necessary for ensuring system longevity, reliability and efficiency. Recent developments highlight the progress in intelligent and automated approaches for identifying faults in PV systems. IoT-based solutions, using smart monitoring methods, allow real-time fault detection, considerably improving system output (Merza et al., 2024). Advanced approaches, including statistical methods, hybrid approaches, and predictive maintenance strategies, have been suggested to enhance the accuracy of fault detection and diagnostics (Joshua et al., 2024). These methods efficiently identify several faults, such as DC-side short circuits, which are mainly difficult to detect under low-irradiance conditions. Utilizing effective fault detection and diagnostic techniques needs a deep knowledge of the environmental, electrical, and physical, factors that affect PV systems (Aghaei et al., 2023). Dependable fault detection techniques are essential for optimizing energy performance, inhibiting PV panel injury, and decreasing fire damage. IoT-based and blockchain structures using wireless sensor nodes and machine learning algorithms allow accurate, real-time fault detection and diagnosis, decreasing downtime and reducing maintenance costs (Apeh and Nwulu, 2025).

Nevertheless, any variations in the environmental conditions surrounding the modules may affect the modelling parameters. For example, the current work demonstrated how a withdrawn single-diode model of equivalent electrical circuit parameters should be adjusted as the accrued dust level on a module varies (Gholami et al., 2022; Pierfederici et al., 2022). The study presented over 13,000 various instances that were analyzed, and it was established that if the dust effect is recognized in the parameter extraction of a single-diode model of the ideality factor such as shunt and series resistance, for the diode as well as the ideality factor, photocurrents, and reverse saturation will be measured respectively with 40, 25, 9, 40 and 35% with high accuracy, which then causes 25 to 35% more precision in the last electrical characteristics prediction for the algorithm. Thus, to cut out the cost of installations, forecasting solar radiation centred on frequently available weather data, for example, relative humidity, temperature, and wind speed, becomes imperative and an alternative technique (Hussain and AlAlili, 2017). To this end, different algorithms have been projected to forecast solar radiation and temperature in the past decades, including ML, empirical models, and remote sensing approaches (Fan et al., 2019; Zhang et al., 2018). With the technological improvements, many studies have adopted ML to forecast energy systems, especially in solar radiation in several nations globally (Alzahrani et al., 2014; Ibrahim and Khatib, 2017; Martín et al., 2010; Sharma and Kakkar, 2018). The algorithms are considered artificial intelligence (AI) techniques and can easily solve complex and problematic issues that can be characterized by ordinary models (Li et al., 2019; Liu et al., 2015, 2016).

In the same vein, the installation of an adaptive neuro-fuzzy inference (ANFIS) algorithm to design solar thermal with output forecast of the indicators where membership functions (MFs) used Gaussian as input and Linear function is regarded as output (Li, 2019; Liu et al., 2015, 2016). Similarly, the study to examine solar radiation using the neuro-fuzzy model utilized four indicators: relative humidity, mean sea level, dry-bulb, and wet-bulb temperature (Jović et al., 2016). The results show that dry-bulb temperature and relative humidity were the basic parameters for forecasting solar irradiation. Likewise, for daily global solar radiation predictions, neuro-fuzzy and support vector machine-firefly algorithms were utilized (Mohammadi et al., 2015). The weather data they used are maximum and minimum temperatures as well as sunshine hours as parameters of the network to predict solar radiation. An algorithm was studied to compare artificial neural networks (ANN) and genetic programming (GP), where the developed algorithm showed a better result. Moreover, an ANFIS model was established in Nigeria for solar radiation prediction, where correlation coefficients accounting for 0.8544 and 0.6567 for the train and test data were obtained respectively (Olatomiwa et al., 2015).

In another development, solar radiation prediction was studied in China to compare amended empirical models and ANFIS (Zou et al., 2017). They equally compared the results of daily solar irradiance forecasting obtained by ANFIS with that of the Improved Yang Hybrid Model (IYHM) and the Expanded-Improved Bristow-Campbell Model (E-IBCM). Several soft-computing methods were performed in a warm and sub-humid atmosphere to forecast daily solar radiation (Quej et al., 2017). The algorithms utilized for solar radiation predictions were ANN and support vector machines (SVM) and ANFIS, where SVM showed better results than other methods with minimum and maximum air temperature, extraterrestrial solar radiation, and rainfall. In South Africa, studies on global horizontal irradiance using information from radiometric stations were recently performed (Ranganai and Sigauke, 2020). They applied three techniques, seasonal autoregressive fractionally integrated moving average (SARFIMA), a Regression model with SARFIMA error, and harmonically Coupled SARIMA (HCSAFRIMA), to dialogue the long-term reliance on integral solar irradiance in the country. Another study describes a PV system for solar irradiance with a hybrid model for short forecasting (Cristaldi et al., 2017). They applied a physical model called clear-sky to forecast solar irradiance in South Africa. Moreover, an auto-associative kernel regression (AAKR) method for short-term PV prediction was then executed for fault detection.

Most recently, the development of several ML algorithms has been studied for solar radiation forecasting, including extreme learning machines (Salcedo-Sanz et al., 2018), extreme gradient boosting (Fan et al., 2018), random forest RF (Prasad et al., 2019), neural network-based approach (Badrudeen et al., 2023), determination of electrical and thermodynamic performance parameters through modelling (Cuce et al., 2017) and deep learning (Kaba et al., 2018). However, the ANN models are regularly utilized among the AI algorithms (Voyant et al., 2017). Nevertheless, the SVM algorithm has been newly projected as an inspiring substitute for solar radiation evaluation due to higher forecasting accuracy and calculation efficiency than the ANN algorithm (Ramli et al., 2015).

Machine learning applications to evolving energy technologies

It takes approximately 50 years for a technology to reach its peak time (Venkatasubramanian, 2019); hence, the expectations remain that by 2035–2040, ML, including AI, will attain a reliable commercial diffusion with an extensive effect on human activities. The expansion of this technology in the previous years in some fields, such as interpretation of text, language processing, games, and image recognition, has described AI as a growing power in daily activities (Venkatasubramanian, 2019). Operators' confidence, acceptance, and prospects of this technology have improved steadily; hence, higher reserves in the areas are necessary to extend into further sectors, such as industrial processes and renewable energies. In different industries, almost modern areas were formed (text mining, image recognition, and games), in different conditions (Venkatasubramanian, 2019). The AI distribution in engineering has varied and is mostly regarded as a subsidiary characteristic of current technologies.

The industrial utilisations of algorithms are numerous as they comprise lots of parameters; the automatic and empirical approaches utilized to resolve those difficulties (that is, through modelling and simulation methods) suggest more computational labours that may not essentially bring out anticipated outcomes, such as correct model predictions. It is imperative to state that ML has likewise been utilized by other energy-connected manufacturers that substantially affect energy generation. The application of ML algorithms to productions such as oil, nuclear power, mining coal, and gas has been operative and supportive for decreasing dangers and increasing results and productivity. For example, big data sets in industries including oil and gas are plentiful, allowing ML algorithms to expand through investigation, reservoir management, drilling, and manufacturing (Hajizadeh, 2019). Data produced by devices from oil and gas pipelines could be applied to train ANNs and forecast several system flaws (Mohamed et al., 2015). Similarly, nuclear manufacturing power is mostly centred on growing data-driven algorithms for safety and fault discovery algorithms that are applied in power plants to forecast possible faults coupled with their likely sources.

Consequently, many research articles are available, but it is very scarce to discover a publication devoted to the bibliometric outlooks on ML prediction to detect faults in solar PV systems. Therefore, this study provides a bibliometric outlook on ML prediction for fault detection in PV systems with the following contributions:

(1) The study identifies inconsistencies in solar systems and thus applies ML to detect faults or deviations from normal operating conditions of solar PV systems with possible solutions.

(2) The importance of the published research article is laid, which includes both performance prediction and fault detection, instead of entirely concentrating on either of these objectives. This broad method intends to contribute to a more complete knowledge for stakeholders, researchers, and policymakers, thus facilitating the integration of ML methods into PV systems and bringing a discerning explanation of dominating trends, significant challenges, and prospects and possibilities.

(3) Besides, regarding the type of ML technique, SVM and ANN stand out as the most commonly utilized and showed variable accuracy, which relies on several factors, including the type of fault and the data quality.

(4) Most ML methods present an accuracy exceeding 90%, emphasizing their efficiency in fault diagnosis. Also, among PV array faults, SC, OC, and PS are the most widely researched within PV systems.

The organization of the paper is as follows: In section 2, solar PV systems are presented, which includes the conventional approach and PV performance, as well as the ML approach. Section 3 describes the methodology, followed by section 4, which presents results and discussions. Finally, conclusions and future works are stated in section 5.

2. Solar photovoltaic system

The electricity generation from solar PV utilizes semiconductor materials directly from solar radiation through the photoelectric effect. The theoretical power generation from PV determines the operating temperature and global solar radiation received on the inclined PV surface. The mathematical algorithm of PV power output is presented in Equations (1) and (2) (Pandiyan et al., 2022).

$$P_{PV}(t) = P_n \times \frac{I(t)}{I_{STC}} \times [1 - \beta x(T_{PV}(t) - T_{STC})] \tag{1}$$

$$T_{PV(t)} = T_{amb}(t) + (T_{NOM} - T_{REF}) \times \frac{I(t)}{I_{REF}} \tag{2}$$

where $P_{PV}(t)$ represents output power from solar PV at time, *t*. Also, P_n represents the nominal power of the solar PV while the tilted solar irradiance at the time, *t*, is $I(t)$. Similarly, $T_{amb}(t)$ and $PV(t)$ represent ambient temperature and operating temperature of PV respectively at t time, T_{STC} and I_{STC} signifies the temperature and irradiance at STC, T_{REF} and I_{REF} are the reference temperature and irradiance, respectively, β represent the temperature coefficient and T_{NOM} is the nominal operating cell temperature.

The temperature of a PV cell and the power output are inversely related. This indicates that as the temperature rises, the voltage decreases, bringing about power loss when other factors remain constant. On the other hand, when the temperature decreases, the voltage increases, producing a gain in power output compared to the initial conditions.

One possible sustainable study field for solar energy systems in AI is based on the advancement of design and materials. However, ML models are utilized to improve the existing operational designs and materials of PV cells and solar thermal systems. Hence, in this learning model utilisation, the systematized and accessible data is crucial, especially the observable scarcity in the study field. In view of this, Yıldırım and Odabası applied data to generate long-term consistency information of over four hundred carbon-lead-built halide perovskite cells (Odabaşı and Yıldırım, 2020). Their goal was to detect and define features that generate a decay of efficiency. The data was equally applied to suggest building high, steady perovskite cells. A decision tree (DT) algorithm was implemented to design assumptions and strategies from the obtained information that describe the decay in efficiency as a function of time.

On the other hand, Li et al. studied a fairly short dataset comprising 915 samples with data around the geometric features of solar water heaters to screen conceivable buildings through ANN (Li et al., 2017). The algorithm produced two solar heaters whose features surpassed those from the data set. Similarly, a successful solar water heater was created by employing a computer-assisted method developed with an advanced system based on ANN (Li et al., 2017).

However, the application of ML algorithms into solar energy systems remains at an emerging stage. Renewable energy is inclined to utilize smart grid systems, especially solar, which are vulnerable to rapid environmental variations. So, prediction is one of the subjects of attention in this area. The connections between the variables to be demonstrated are occasionally fairly complicated in the sense that other possibilities, such as hybrid algorithms or deep neural networks (DNNs), are more accepting (such as climate and solar irradiance, the rate of heat collection and extrinsic characteristics in solar collectors (Sharma et al., 2011). Hybrid algorithms involving first-principles and data-focused algorithms are of rising attention. Still, they face difficulties connected to consistency in the performance metrics, set-up parameters, and data sets (Voyant et al., 2017). However, the applications of PV technologies have been widely accepted in fields such as farming systems and stand-alone PV farms. The approaches utilized to evaluate the output of PV systems are grouped in two ways. The initial technique is conventional, which relates to the utilization of governing equations or simple correlations to evaluate the system output, whereas the other approach is machine learning.

2.1. Conventional approach of photovoltaic performance

Solar PV modules can either be installed as a rooftop or have the prospect of being incorporated with the walls of buildings. **Figure 1** shows the outline of the conventional approach to solar PV systems.

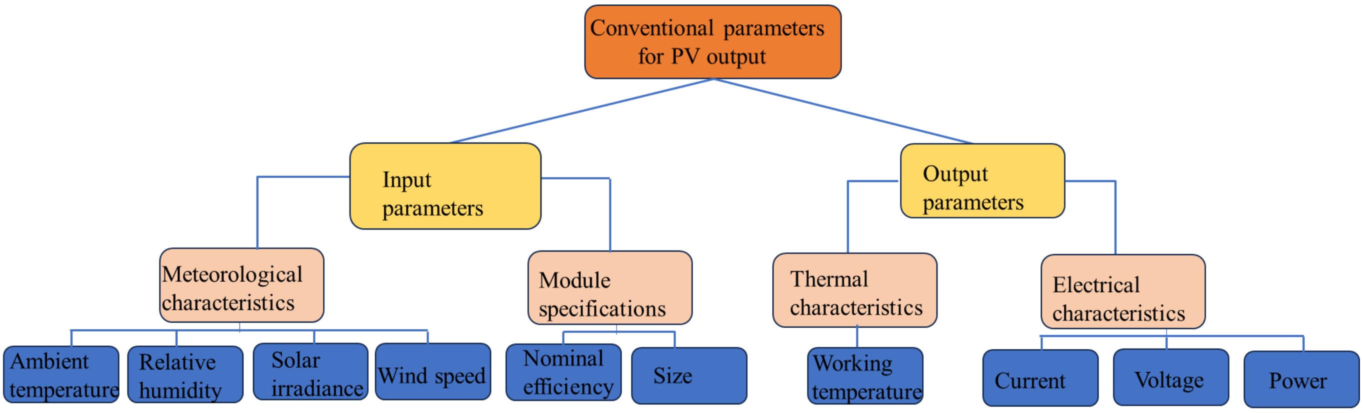

Figure 1. Conventional methods showing input and output parameters for PV systems performance prediction.

2.1.1. Thermal modeling of PV systems

Thermal modelling aims to ascertain the operational solar module temperature and other associated metric performances.

Nominal operating cell temperature (NOCT)

The NOCT stands as the most straightforward technique for projecting the operational temperature of a solar PV module. It requires ambient temperature (T_{amb}) and incident solar radiation (G). In this approach, the module temperature (T_{mod}) is forecasted using Equation (3) (Sohani et al., 2022).

$$T_{mod,NOCT} = T_{amb} + \frac{G}{G_{ref}}(T_{NOCT} - T_{ref}) \quad (3)$$

The NOCT temperature (T_{NOCT}) is an input variable associated with each module, with reference condition 'ref' written in the subscript. Under reference conditions, the temperature and irradiance are set at 20 °C and 800 W/m^2, respectively. It is important to note that the reference condition is distinct from the STC, where the irradiance and temperature are 1000 W/m^2 and 25 °C, respectively.

Correlation techniques

Although Equation (1) is readily employed in individual solar modules, it has certain limitations:

- PV modules with similar output member families typically share the same TNOCT values. Consequently, the NOCT method forecasts a uniform value of every solar module output. However, due to disparities in dimensions and heat exchange rate, T_{mod} differ among several sizes of modules within an output member family, even when T_{amb} and G are constant.
- Apart from T_{amb} and G factors such as wind velocity (V_w) and relative humidity (φ) can influence T_{mod}. Nevertheless, the NOCT method does not account for these variables.

Lately, alternative research works on correlations have been based on experimental findings, with a notable range of functions being developed. Among them, the nominal module operating temperature (NMOT) has gained extensive adoption and has been designated with IEC number 61853–2 for PV rating. This work revealed that only a few of the numerous proposed correlations consider the impact of relative humidity. However, in certain recent works, such as the research by Sohani et al. (2020b), relative humidity was considered in the PV system.

2.1.2. Electrical system modeling

Electrical system modelling aims to determine key output parameters, primarily focusing on power. Additionally, the current and voltage values, as significant parameters, can also be ascertained through this modelling process.

Correlation-based approach

In the present research endeavours, when the objective is to retrieve power, the initial step involves employing Equation 4 to compute efficiency. Subsequently, the power output (P) is calculated by multiplying the efficiency (η) with the incident solar radiation (G) and the module's surface area (A_{mod}), both of which are established parameters (Rashidi et al., 2021):

$$P = \eta G A_{mod} \tag{4}$$

However, the efficiency (η) is further defined in Equation (5).

$$\eta = \eta_{ref}\left\{1 - \beta_{ref}(T_{\mathrm{mod}} - T_{ref}) + \gamma \log_{10}(G)\right\} \tag{5}$$

Here T_{mod} is a known parameter deducible from thermal modelling. Furthermore, η_{ref} signifies the module's efficiency under reference conditions while γ representing the coefficient applicable to conditions yielding maximum power. Manufacturers typically provide values for both η_{ref} and γ.

Equivalent circuit method

The equivalent circuit method is valuable when focusing on electrical parameters beyond power. This methodology uses an equivalent circuit to represent the module's electrical behaviour, which is comprised of multiple resistors and diodes. Three pivotal components within each equivalent circuit are as follows:

- Photocurrent (Iph): This component measures the current produced due to incident solar radiation.
- Series Resistance (Rs): A fraction of the energy produced by the solar module dissipates by means of the metallic connections and semiconductor due to the current flow.

Parallel resistance (RP), also called shunt resistance, attracts various phenomena, including the conduction of current from non-idealities, the crystal geometry holes, and the module's edges.

1) Single diode

The single diode model is referred to as a one-diode method where an equivalent circuit resembling **Figure 2** is employed. This circuit consists of a current source that measures the photocurrent produced, a series resistance as well as shunt resistance.

Additionally, a diode is present within the circuit. A portion of the sunlight-generated current flows through the diode, reducing the voltage received at the terminals.

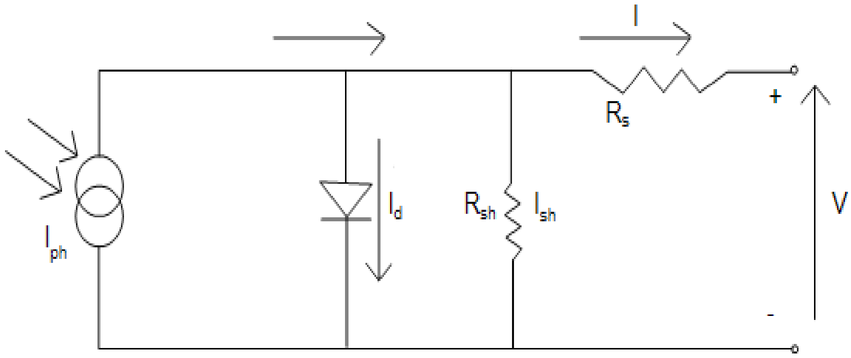

Figure 2. A practical single-diode equivalent circuit with parasitic series and shunt resistances (Gholami et al., 2021).

Abiding by the principles of electrical governance, the connections between voltage and current for the single-diode model are expressed as follows in Equation (6) (Nižetić et al., 2021b):

$$I = I_{ph} - I_0[exp(\frac{1}{V_t}(\frac{V}{N_s} + IR_s)) - 1] - \frac{1}{R_p}(\frac{V}{N_s} + R_s I) \tag{6}$$

In addition to the variables R_p, I_{ph}, and R_s, there are three additional crucial parameters within these equations (Nižetić et al., 2021a):

Thermal voltage of diode (V_t): This parameter characterizes the thermal driving force experienced by electrons in a semiconductor (Gholami et al., 2022). Notably, this motion is contrary to that of the photocurrent.

Diode saturation current (I_0): This describes the motion executed by the minority charge carriers within a semiconductor from a neutral region to a depletion layer.

Number of cells connected in series in a module (N_S): Modules are constructed by linking cells in series electrically. Besides the current and voltage, the other variables in Equation (6) are obtained from Equations (7)–(11):

$$R_p = \frac{G_{STC}}{G} R_{p,STC} \tag{7}$$

$$R_s = R_{s,STC} \tag{8}$$

$$I_0 = I_{0,STC}(\frac{T_{\mathrm{mod}}}{T_{\mathrm{mod}.STC}})^3 \exp[\frac{qE_g}{K}(\frac{1}{T_{\mathrm{mod},STC}} - \frac{1}{T_{\mathrm{mod}}})] \tag{9}$$

$$\mathrm{V}_t = \frac{T_{\mathrm{mod}}}{T_{\mathrm{mod},STC}} V_{t,STC} \tag{10}$$

$$I_{ph} = \frac{G}{G_{STC}}\{I_{ph,STC} + \alpha(T_{\mathrm{mod}} - T_{\mathrm{mod},STC})\} \tag{11}$$

Values at STC are needed to find out the values in the condition examined. At STC, the values of I_0, V_t, and I_{ph} are computed from Equations (12)–(14):

$$I_{0,STC} = I_{sc,STC}\, exp(\frac{-V_{oc,STC}}{N_s V_{t,STC}}) \quad (12)$$

$$V_{t,STC} = \frac{\beta T_{mod,STC} - V_{oc,STC}}{\frac{N_s T_{mod,STC}\alpha}{I_{ph,STC}} - 3N_s - \frac{E_{g,N_s}}{KT_{mod,STC}}} \quad (13)$$

$$I_{ph,STC} = I_{sc,STC} \quad (14)$$

When considering the right-hand side of Equations (12)–(14), the values for all the parameters found are established and well-known. These values encompass the module characteristics readily accessible through the module directory (for example, $I_{sc,STC}$, $V_{oc,STC}$, β, and α) or unchanging constants (like E_g and K).

2) Double diode model

The ideality factor is an essential parameter for a PV module, which shares similarities with a diode. In the lower voltage range, the ideality factor tends to be approximately 2, primarily due to junction recombination playing a dominant role. Conversely, the ideality factor converges towards unity in the higher voltage range as the primary recombination processes shift to the bulk zone and surface recombination within the PV module. While the single-diode model assumes a constant ideality factor, introducing a second diode in parallel enhances accuracy. The mathematical expressions for the two-diode model closely resemble those of the single-diode model, with adjustments stemming from including the second diode in **Figure 3**. Furthermore, at STC, determining the series resistance value involves computing Equation (15), where this parameter stands as the only unknown entity.

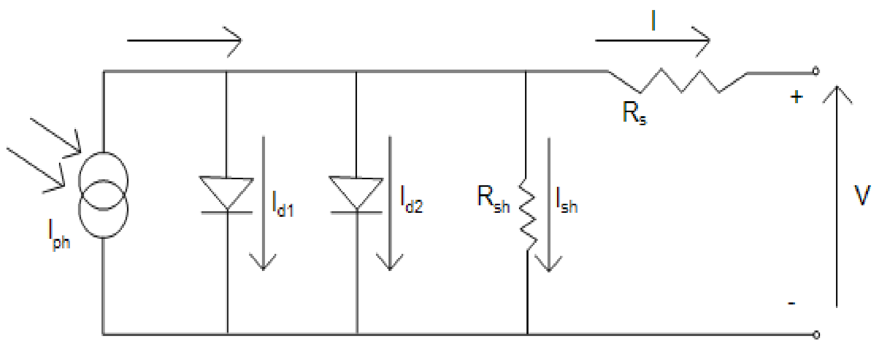

Figure 3. The p-n junction of two diodes equivalent circuit model.

$$I_{mp,STC} = I_{ph,STC} - I_{0,STC}\left[exp\left(\frac{V_{mp,STC} + I_{mp,STC}R_{s,STC}}{N_s V_t}\right) - 1\right]$$
$$\frac{(V_{mp,STC} + I_{mp,STC}R_{s,STC})[-N_s V_{t,STC} I_{mp.STC} + (V_{mp,STC} - I_{mp,STC}R_{s,STC})(I_{sc,STC} - I_{mp,STC})]}{(-I_{mp,STC}R_{s,STC} + V_{mp,STC})(V_{mp,STC} - N_s V_{t,STC})} \quad (15)$$

Similarly, at STC, the shunt resistance can equally be computed from Equation (16).

$$R_{p,STC} = \frac{\left(-I_{mp,STC}R_{s,STC} + V_{mp,STC}\right)\left(V_{mp,STC} - N_sV_{t,STC}\right)}{-N_sV_{t,STC}I_{mp,STC} + \left(V_{mp,STC} - I_{mp,STC}R_{s,STC}\right)\left(-I_{mp,STC} + I_{sc,STC}\right)} \quad (16)$$

These modifications can be found in various sources, such as Babu and Gurjar, (2014), Shannan et al. (2013), Chennoufi et al. (2021), Sangeetha et al. (2021). The two unknown diode quality factors do not only increase the number of equations but also the unknown parameters, thereby producing much more complex calculations.

2.2. Machine learning algorithms

The extensive applications of a machine learning algorithm in solar energy have been achieved because of the huge number of variants and models that could reach the demands of data clustering, classification, and regression. Yet, the algorithm matches the needs of solar predicting and builds forecast models according to past data. The machine-learning algorithm uses an input vector x as a function $f(x)$ and produces an output vector y. Machine learning classification involves two approaches, supervised and unsupervised learning, as presented in **Figure 4**.

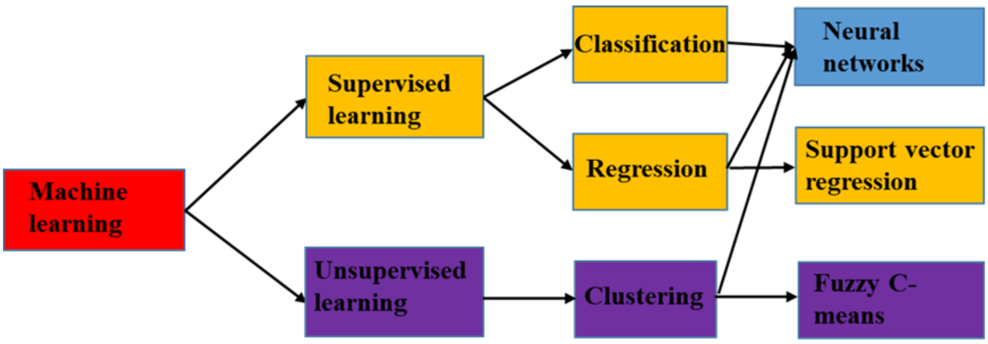

Figure 4. Categories of machine learning.

Unlike supervised learning models, these types of models operate without requiring expert intervention. These models can distinctly uncover concealed patterns within inputs, regardless of output knowledge. Unsupervised learning mirrors the statistical challenge of density estimation, but it extends beyond this by encompassing various techniques aiming to condense and elucidate the essential characteristics of the data. Numerous techniques utilized in unsupervised learning are rooted in data mining approaches commonly applied for data preprocessing purposes. Nevertheless, performance forecasting is one of the most significant uses of ML methods for PV connections. This can be undertaken with the intention of defining several variables. The different ML approaches that have been applied in research works are shown in **Figure 5**.

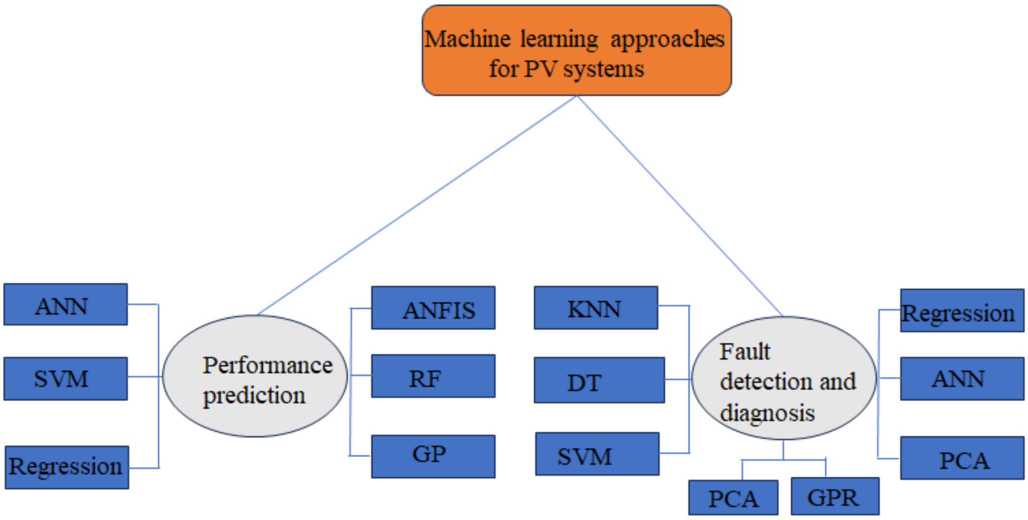

Figure 5. Machine learning approaches for performance prediction and fault detection in PV systems.

2.2.1. Time series algorithms

Time-series forecasting (N_2) refers to an algorithm that forecasts the future results of any system with historical information. Meanwhile, this prediction method requires the description of the data either by non-linear or linear autoregressive method. Therefore, the time series algorithm Equation is presented in Equation (17).

$$x(k+h) = f[x(k), x(k-1), x(k-2), \ldots, x(k-n-1)] \quad (17)$$

where function f represents the current value and the historical value of x. However, different methods of the possible forecast involve forecasting the h next values of the time series as defined in two techniques. The first method is the independent value forecast (preparing the direct model to predict $x(k + h)$), while the other step involves an iterative technique and reiterating one-method-upfront forecasting until the preferred possibility.

2.2.2. Supportive vector machine (SVM)

This type of ML is a classification technique rooted in statistical information theory and fundamental risk minimization constructs multiple hyperplanes to categorize data samples. Among these, the most optimal hyperplane is selected, granting SVM an advantage in mitigating overfitting.

SVM stands out as a preferred machine learning approach within the energy domain, particularly for predicting renewable power sources like PV systems and wind turbines. Its ability to handle fluctuations and time-varying parameters is noteworthy. For instance, SVM is used by Wolff et al. (2016) to harness historical data encompassing numerical weather prediction (NWP), PV power, and solar irradiation inputs to establish short-term PV power forecasts using SVM. The efficacy of their strategy is assessed using BIAS and root mean square error (RMSE) statistical error metrics. Similarly, SVM is utilized by Leone et al. (2015) for forecasting energy yield from an Italian solar PV plant, utilizing ambient temperature, solar irradiance, and past PV power data in 15-minute intervals. The study achieves >90% performance using the R score, with weather forecasts playing a crucial role. Kim et al. (2017) studied the relationship between SVM and ANN algorithms for predicting daily PV output from a PV system in Korea. The results show that SVM, leveraging cloud cover and temperature features, outperformed ANN in terms of average Percentage Prediction Error (PPE).

In a tropical region, an investigation is applied in both SVM and multilayer perceptron (MLP) methods for hourly solar irradiation prediction. Using data retrieved from a solar farm in Malaysia, SVM exhibited better prediction results than the MLP model, as evidenced by metrics such as the Kolmogorov-Smirnov integral test, MAE, MBE, and RMSE. However, Dong et al. (2013) devised an innovative hybrid method involving particle swarm optimization, SVR, and Self-Organizing Maps (SOM) for solar irradiation prediction. SVR showcased superior accuracy compared to traditional methods like linear, simple exponential smoothing (SES), Autoregressive integrated moving average (ARIMA), exponential smoothing (LES), and random walk (RW). Many other research works, such as Mellit et al. (2013), introduced a least-squares SVM algorithm for 24-hour solar irradiation forecasts, incorporating inputs like wind speed, solar irradiation, wind direction, air temperature, atmospheric pressure, and relative humidity. The model outperformed recurrent NN (RNN), MLP, and radial basis function NN with respect to performance metrics, including R_2 and mean absolute percentage error (MAPE). The least squares SVM centering on the PV power prediction system was developed by Wang et al. (2015). The study considers sky cover, relative humidity, and wind speed as input parameters, highlighting SVM's superiority over other experimental methods (for example, ANN) in handling time-varying and non-linear parameters.

The Naive Bayes (NB) model method is a widely utilized supervised learning approach centered on Bayes' theorem. This algorithm calculates conditional probabilities on training data for classification. The key principle of NB is that every feature exerts a self-regulating and equivalent effect on the outcome, contributing to its ease of implementation and strong classification performance. Consequently, Kwon et al. (2019) conducted a two-day ahead solar irradiation forecast utilizing the NB classifier, achieving a 2.73% root Mean Bias Error (RMBE) with a 333.04 Wh/m^2 of Global Horizontal Irradiance (GHI). In a similar study, Persson et al. (2017) applied the NB classifier for day-ahead solar power predictions in one-hour intervals, demonstrating its superior accuracy and reliability compared to other experimented methods through normalized Mean Absolute Error (MAE) evaluation.

2.2.3. Multivariable regression method (MLR)

The MLR is an ML technique established with a focus on the general formula depicted in Equation (18).

$$R_s = m_1 + m_2X_1 + m_3X_2 + m_4X_3 .. + m_{n+1}X_n \quad (18)$$

while X_i stands for meteorological parameters, m_1, m_2, m_3, m_4…, m_{n+1} represent regression coefficients. MLR is mostly an algorithm used to study the relationship between multiple independent and dependent parameters. Additionally, it is broadly used in assessing solar radiation research (He et al., 2020).

2.2.4. Artificial neural network

The artificial neural network is a supervised learning approach stimulated by the intricate structure of the human brain's neurons. It comprises interrelated layers of nodes, each layer equipped with weights and activation functions. The input layer obtains input samples, which are then processed through hidden layers to produce an output that represents the final prediction. The capability of ANN to take care of noisy and incomplete data makes it a preferred choice in knowledge discovery endeavours. Presently, there exist six distinct types of ANNs employed in ML research, including self-organizing map (SOM) NN, multilayer perceptron (MLP), back-propagation NN, feed-forward NN, extreme learning (EL) and radial basis function NN. The utilization of ANNs within the realm of solar energy, particularly in power prediction, has seen extensive exploration and presentation. As an illustration, Mandal et al. (2012) introduced a hybrid approach that combines radial basis function NN and wavelet transform (WT) methods to predict one-hour-ahead PV power using weather temperature and solar irradiation inputs. Empirical findings demonstrate that this hybrid technique captures non-linear PV fluctuations more effectively than other methods tested (WT + BPNN, radial basis function NN, BPNN, and WT), albeit with some limitations during rainy conditions. A hybrid system was devised by Cervone et al. (2017) to merge analogue ensemble (AnEn) ANN models for PV power generation forecast in an Italian plant. The hybrid model, evaluated against other methods ANN and AnEn), yields notably more accurate results. In another instance, Muhammad et al. (2017) utilized an ANN regressor for day-ahead power prediction of a 20 kWp grid-connected PV plant in Tiruchirappalli, India. The assessment of this model using statistical error analysis, specifically MAPE, demonstrates a remarkable accuracy of 0.855%. Ramsami and Oree (2015) amalgamated ANN and regression techniques for stochastic energy output prediction of PV systems, leveraging stepwise regression to improve forecasting accuracy. The hybrid ANN + LR model outperforms single-stage models across various evaluation metrics.

Asrari et al. (2016) proposed a hybrid prediction model for hour-ahead solar power prediction, combining meta-heuristic and gradient-descent optimization techniques. This algorithm demonstrates improved prediction accuracy while managing computational complexity. In a different context in a grid-connected PV system, an enhanced multilayer feed-forward NN with a fuzzy rule-based classifier was developed for maximum power point tracking (MPPT) (Chaouachi et al., 2010). Izgi et al. (2012) employed ANN for short and medium-term power forecasts, determining optimal time horizons for different seasons. In the same vein, Khandakar et al., (2019) compared M5P regression tree learners, linear regression, ANN, and Gaussian process regression models for PV power forecast, with the ANN model outperforming other regression approaches.

Several research works in the systematic literature review (SLR) focus on solar irradiation forecasting, given its critical role in PV power generation. ANNs are a prominent choice for estimating solar irradiation and PV power production. For instance, Jang et al. (2016) introduced an innovative solar irradiation prediction approach using ANN and fuzzy logic algorithm. In another study, radial basis NN is employed for day-ahead solar irradiation forecasting, achieving a statistical error of 12% (Jang et al., 2016). Jang et al. (2016) presented a back-propagation NN-based short-term prediction algorithm that enhances generating capacity prediction precision. Moreover, Amrouche and Le Pivert (2014) demonstrated that ANNs could adapt to noisy and missing data and successfully predict solar irradiation for locations lacking direct measurements by utilizing nearby meteorological data.

2.2.5. Random forest

This algorithm constitutes an ensemble-driven classification technique wherein it generates multiple Decision Trees (DTs) and subsequently merges their outcomes to enhance the precision of class labelling (Basaran et al., 2019). This methodology hinges on a consensus-based aggregation of predictions from all the trees within the ensemble, collectively known as a forest, to foresee the classification of new sample instances.

In the research paper by Ahmad et al. (2015), a pioneering ensemble-focused system was introduced for forecasting stochastic PV production output hourly. This system harnesses a combination of tree-based ensemble and SVM methods, specifically Extra Trees (ET) and Random Forest (RF), to predict the hourly power yield from a PV system situated in Cardiff, UK. The findings underscore that ET and RF furnish more precise prediction outcomes while demanding reduced training time in comparison to SVM. Another study proposes an innovative framework that synergizes wavelet decomposition and bias-corrected RF models for PV output prediction. This approach employs present PV output values and meteorological sensor data (Antonanzas et al., 2017) to predict PV power production. The expected technique is benchmarked against Back-propagation Neural Network (BPNN), RF, and Wavelet-SVM models, gauged by Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) metrics. As per the results, the proposed approach exhibits diminished error rates in contrast to the other employed algorithms.

Almeida et al. (2015) introduce Quantile Regression Forests, an extension of RFs tailored for quantile regression, to devise a solar power prediction system. This model assimilates multiple forecasts of meteorological variables and genuine AC power measurements from PV plants as input. The model’s efficacy is evaluated using data from five PV plants in northern Spain, sampled at a 5-second resolution. The findings from the assessment demonstrate that the model accurately predicts hourly and daily power generation, achieving an absolute Mean Bias Error (MBE) of less than 1.3% and MAE of less than 1.3% for both timeframes, respectively. In a distinct study by Zamo et al. (2014), a predictive model is constructed to estimate the hourly PV electricity production for the following day at various power plants in mainland France. This model employs binary regression trees, bagging, RF, gradient boosting (GB), and SVM learning models. The evaluation results unveil that RF outperforms the other experimented models in terms of accuracy for hourly PV production forecasts.

3. Research methodology

A bibliographic analysis is a study area that describes the lines of action among keywords, authors, and topics to graph a correlation between them and detect trends as well as areas of concentration in publication activities. These forms of analyses direct scholars on how earlier studies proceeded over time and where they can focus in the future. A bibliographic review is significant since it irradiates their networks and demonstrates which fields are highly intertwined while others lack correlations (David et al., 2022). This exposes parts that have attracted less interest and have the prospect of being discovered.

Several software are mapped out for bibliometric review (Daneshgar et al., 2022). However, the application of VOSviewer software stands out as one of the greatest and most appropriate tools utilized to generate link data and maps. This package was designed by van Eck and Waltman (2009). The maps generated by VOSviewer comprise different words, including publications, researchers, and keywords, which attract great attention from users (Apeh and Nwulu, 2024). Each map concentrates on a single form of the word, and the relationship between these words is virtually represented by bows. These connections form links, with words and their connotation mutually shaping the networks. These words are assembled into groups that typically have correlated topics.

In search of the study objective, VOSviewer software (Savaresi, 2013) is utilised to execute the bibliographic analyses and examine the correlations between diverse study fields. It is an open software designed by Ech and Waltman with the assistance of the University of Leiden. This software concepts web and bibliometric maps in accordance with data built on authors, journals and keywords.

3.1. Data sources and search procedures

This study limits the search terms to English language. To examine the performance analysis of the appropriate articles for this review, an extensive search was undertaken across famous global databases, including Scopus, ScienceDirect and Google Scholar. However, Scopus has been identified as the highest academic abstract and citation database globally; it comprises a constant broad variety of subjects and contents encircling virtually 50 million literature parts published from 1823 (Agbodjan et al., 2022). This stated evidence correlated to its significance in the published research content validates why Scopus has been utilised as the database to execute the bibliometric analysis of machine learning for solar energy. During the search process, various keywords were employed in diverse combinations, encompassing terms such as “failure,” “modeling,” “machine learning,” “fault”, “artificial intelligence,” “performance prediction,” “neural network” as well as several methodologies within ML example ANN or SVM, and various fault categories examples, “line to line”, “short circuit,” “degradation,” and their standard abbreviations. The flowchart in **Figure 6** illustrates the overview of the research methodology used in this study.

However, the scope of articles considered was up to the most recent submission dated 11 January 2025, with corresponding revisions accepted until 17 February 2025. The culmination of this search yielded a selection of over 2258 articles that have been incorporated into the present review. These keywords are examined distinctly, and the outcomes of all the keywords are imported into the VOSviewer concurrently.

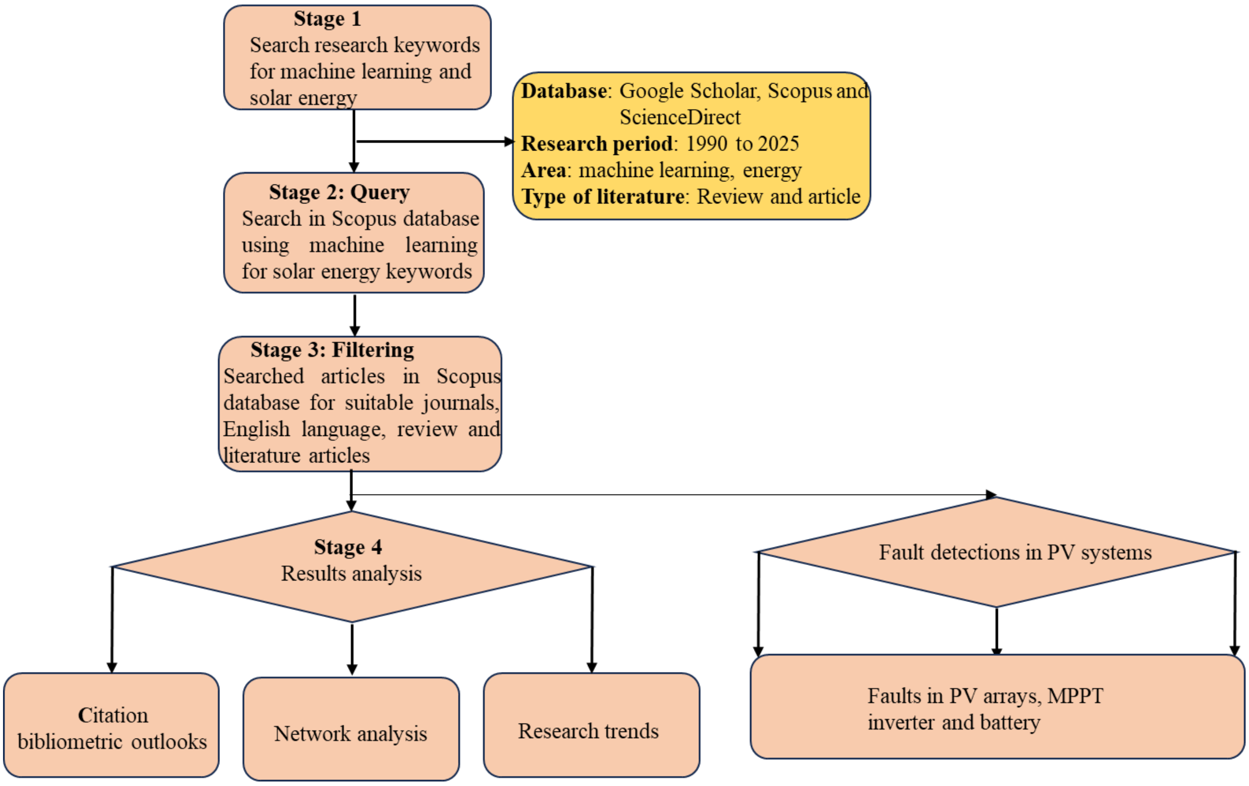

Figure 6. Research methodology and the summary of the process for each step.

3.2. Selection process

3.2.1. Inclusion criteria

- Complete-article accessibility.
- Publication in the final stage.
- Academic journals and conference papers.
- Published between 1990 and January 2025.
- Focus areas connected to operations research, engineering, energy, and decision sciences.

3.2.2. Exclusion criteria

- Book chapters, abstracts, technical reports, and dissertations were excluded.
- Papers covering alternative renewable or non-renewable energy sources, other than PV technologies were exempted.
- Articles that cannot be accessed.
- Restricted to English language only.

3.3. Fault detections in PV system

Effective and efficient detection and diagnosis of faults in a PV system requires a systematic knowledge of the characteristics of every fault and its respective challenges. A typical Stand-Alone Photovoltaic System (SAPVS) includes different components such as PV arrays, inverters, batteries, charge controllers, Maximum Power Point Tracking (MPPT) systems, connection wires, and additional protective devices.

3.3.1. Faults in PV array

An open circuit (OC) occurs due to accidental disconnection within a closed loop, resulting from objects hitting the panels, cable scratches, loose connections, or unintentional disconnections at current-carrying conductors. Also, cracked cells, cable joint failures, loose connections, and old power cables close to terminals can cause OC faults (Ali et al., 2020). Notwithstanding the existence of bypass diodes sustaining some current flow during an OC fault, substantial power loss occurs because of a voltage drop in a string (Ali et al., 2020).

3.3.2. Faults in solar battery

Batteries constitute approximately 43% of SAPVS lifecycle prices, requiring attention to guarantee optimal operation. The possible faults comprise external short circuits, degradation, internal faults such as Short Circuits (SC) and Ground Faults (GF), open circuits, undercharging and overcharging (Zenebe et al., 2021). These faults can decrease battery performance, shorten its lifespan, increase maintenance charges, and expose it to the dangers of fire or explosion (Zenebe et al., 2021). Challenges in sensing internal faults consist of the lack of guidelines for choosing fuses and circuit breakers, worsened by the slow fluctuations in battery current and voltage over time.

3.3.3. Faults in inverter

Inverter faults involve different issues, such as switch OC, switch SC, filter failure, and gating failure (Mellit et al., 2018). Gating failure, mostly incipient faults in the Insulated Gate Bipolar Transistor (IGBT), is dangerous as it frequently causes inverter failures. Recognizing such faults can improve system reliability, although generating these faults for training and validating ML needs exact procedures. The illustration is presented in **Figure 7**.

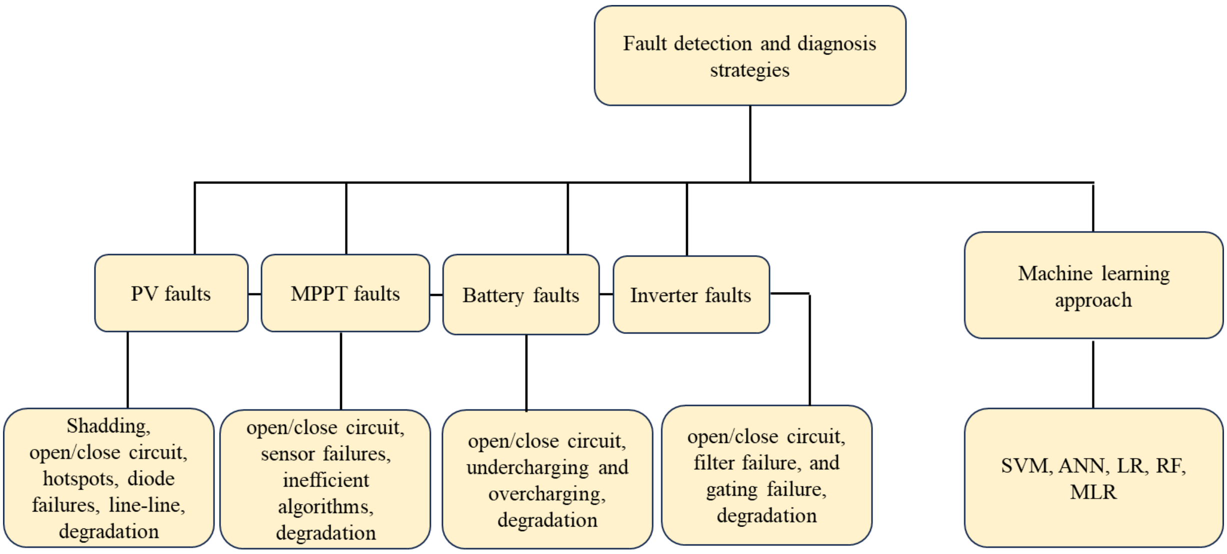

Figure 7. Fault detection and diagnosis strategies.

3.3.4. Faults in MPPT

The MPPT system, essential for optimizing PV array performance, depends on sensors for irradiance, temperature, current, and voltage measurements, along with optimization algorithms. Faults within the MPPT system, such as sensor failures or inefficient algorithms, cause incorrect operating points and significant cuts in system output power (Jäger-Waldau, 2022).

4. Results and discussion

4.1. Faults evaluations in PV systems

Table 1 shows how several research has been analysing faults in PV arrays, with more than 80% of papers focusing on this field. Nevertheless, research into faults within batteries and inverters has also been undertaken. Particularly, faults in Grid-Connected Photovoltaic Systems (GCPVS) have attracted more emphasis than SAPVS. The most analysed types of faults within PV arrays include SC, OC, and Partial Shading (PS), despite Ground Faults (GF), Arc Faults (AF), and Line Losses (LL) being the most critical. SVM and MLP have been extensively used among several ML methods for fault detection and classification in PV systems.

Table 1. Summary of reviewed literature for fault detection using machine learning method.

Fault identified			**Methods**							
Fault	**component**	**Input data**	**Algorithm**	**Type**	**Simulated**	**Experimental**	**Type of PV system**	**Accuracy**	**Year**	**References**
OC, SC, Degradation	PV array	-	KNN, DT, SVM, ANN	ML	yes	No	GCPVS	-	2020	Lazzaretti et al., 2020
Hotspot	PV array	Image	SVM	ML	No	Yes	SAPVS	> 92%	2020	Ali et al., 2020
LL	PV array	I-V char	SVM, NB, KN	EL	Yes	Yes	-	> 99%	2020	Eskandari et al., 2020
OC, Norma, SC	PV array	I, V, P, T, G, string	DT, RF, DA	EL	No	Yes	GCPVS	> 97%	2021	Kapucu and Cubukcu, 2021
Module SC, String, OC	PV array	I, V, T, G	ANN, PNN	ML	Yes	Yes	GCPVS	-	2017	Garoudja et al., 2017
OC, degradation, SC	PV array	G, T, P	GPR, SVM	ML	Yes	Yes	GCPVS	-	2021	Harrou et al., 2021
String-string, string-ground, and OC	PV array	-	RF, CNN, LSTM	DL	Yes	No	GCPVS	> 99%	2023	Mustafa et al., 2023
LL, degradation, PS, bypass dode	PV array	I, V	ML, DL	ML	-	-	-	-	2022	Mellit and Kalogirou, 2022
Norma, SC	Inverter	I, V	ResNet-50 CNN	TL	Yes	No	GCPVS	> 97%	2021	Malik et al., 2021
SC, OC	PV aray, battery	I, V	ANN	ML	Yes	Yes	GCPVS	> 96%	2018	Sabri et al., 2018

The assessment of these models mostly depends on accuracy and confusion metrics, even though some researchers present their metrics and consider implementation time. Given the stochastic nature of ML, it is important to define performance after a reasonable number of model implementations despite the time investment it entails. In terms of geographical location and data source, research on experimental PV systems is largely conducted in Algeria, followed by China and Korea. Assessing ML methods under various climatic and geographical conditions is crucial because of the different challenges possessed by the PV array and the battery in various locations. For example, while snow poses a great challenge in polar regions, equatorial regions face issues such as dust accumulation, soiling, and high operating temperatures.

Several studies rely on simulated PV system data for input features, with only a minority integrating experimental data due to the challenges of setting up operational PV systems for data collection. Input features usually comprise irradiance, temperature, and main points from I-V characteristics for PV array faults, while current and voltage data are employed for faults in batteries, inverters, MPPT, and others. Even though electrical and meteorological data are mostly utilized in ML, image data is common for deep learning algorithms, including Convolutional Neural Networks (CNN). Recently, there has been a drift towards using electrical and meteorological data in deep learning algorithms as well, following transformations to change 1-D data into 2-D. For faults such as arc faults that do not show in I-V characteristics, devices analysing signal waveforms (for example, wavelet transformation) capable of capturing signal distortion effects may be suitable. Preprocessing methods like normalization frequently improve accuracy, but when not possible, deep learning models excel because of their ability to automatically extract features. Despite substantial developments in using ML methods for fault detection in PV systems, the commercial application remains limited, with only one paper employing ML in a prototype based on results from the literature review. Therefore, authors have acknowledged different challenges that hinder the progress of PV systems as analysed in the following:

- PV size and type are very scarce to discover, training, validate, and test data set that fit at least the main fault in a PV system.
- Even if most researchers have established their own data set, many are simulation data. In addition, in DL-based approaches, collecting the image data using a drone and camera is costly.
- Most measuring instruments and sensors are required because there is no suitable technique for efficient input feature selection.
- There is inadequate skill to detect severe but scarce faults.
- Choices of model configurations are done with trial and error.
- The model developed so far does not have the modularity and generalization capacity; consequently, the type of fault, size, and input data type rely on different ML model selections.
- Research that control, including the approaches with the present protective plans, are scarce. Besides, most of the articles do not offer detailed solutions on how to address the faults. Once the fault is categorized, a technique and plan are required to organize it with protective devices to clear the fault automatically and send the signal to the operators for solutions.
- The accuracy of the model is flexible as it relies on data quality, data size, and input and output features.

- There are no common platforms or testing standards based on execution time, accuracy, cost, or memory usage for comparing ML devices.

4.2. Advantages and disadvantages of using machine learning in diagnosing faults in PV systems

4.2.1. Time series algorithms

Advantages of time series algorithms

Temporal patterns: Time series algorithms are designed to capture and analyze temporal patterns and trends, making them highly suitable for data that exhibit seasonality, trends, and other time-dependent patterns. These algorithms excel at forecasting future values based on historical data. They can predict stock prices, weather conditions, sales, demand, and other time-dependent variables. Time series algorithms can provide real-time insights and predictions, enabling organizations to make timely and informed decisions. Time series algorithms can compress and summarize large volumes of time-based data, making it easier to store, visualize, and analyze the information. Time series algorithms can extract relevant features from time-dependent data, aiding in identifying important patterns and relationships. They are broadly employed in several domains, including finance, economics, healthcare, manufacturing, and more, for anomaly detection, quality control, and resource allocation.

Disadvantages of time series algorithms

Data preprocessing: Time series data often require extensive preprocessing to handle missing values, outliers, and noise, which can impact the quality of predictions. Advanced time series algorithms can be complex and require a deep understanding of mathematical and statistical concepts, making them challenging to implement and interpret. Dealing with multivariate time series (multiple time-dependent variables) can increase the complexity of analysis and modelling. Depending on the complexity of the algorithm and the amount of available data, there's a risk of overfitting, where the model fits noise in the data rather than capturing meaningful patterns. Time series data can exhibit non-stationarity, where statistical properties change over time. Handling non-stationary data can be tricky and may require additional techniques. Accurate predictions often rely on a sufficient amount of historical data. If data is limited, the performance of time series algorithms can be compromised. Choosing the right algorithm and model parameters for a specific time series problem can be challenging and might require experimentation. Some advanced time series algorithms can be computationally intensive, especially when dealing with large datasets or complex models.

4.2.2. Supportive vector machine

Advantages of supportive vector machine

SVM offers a range of merits. Firstly, they exhibit versatility by handling both classification and regression tasks, making them well-suited for diverse data types, including structured and semi-structured datasets. A pivotal component of SVM is the kernel function, a collection of mathematical functions defined by various SVM algorithms. These kernels encompass various types, such as linear, non-linear, and sigmoid, enabling them to handle intricate data patterns when appropriately tailored effectively. SVM shines within high-dimensional spaces when the number of samples is smaller than the dimensions, showcasing enhanced efficacy and reduced susceptibility to overfitting. This inherent universality steers clear of local optima entanglements. When optimality is approximated, SVM can generate a unique solution. This distinguishes SVM from Neural Networks, which often yield several solutions prompted by local minima, making their results less reliable and consistent across various datasets (Karamizadeh et al., 2014).

Disadvantages of supportive vector machine

When dealing with extensive data collection, SVM's performance can decrease over time, constituting a drawback. Moreover, its consistency is compromised in the presence of noisy data. Another disadvantage arises from the intricate process of selecting an apt kernel function, which demands substantial effort and time. Unlike DT, SVM algorithms are notably complex, posing challenges in comprehending outcomes, especially for datasets rich in features. Lastly, SVM's performance deteriorates as the number of features surpasses the count of training instances.

4.2.3. Multivariable linear regression (MLR)

Advantages of multivariable linear regression

The MLR allows one to account for the impact of multiple independent variables on a single dependent variable. This is particularly useful when the outcome is influenced by several factors simultaneously. Moreover, it provides coefficients for each independent variable, indicating the direction and magnitude of their impact on the dependent variable. This makes it easier to interpret the relationships between variables. When the underlying assumptions of MLR are met, it can provide accurate predictions for the dependent variable with a focus on the values of the independent variables. MLR can help identify which independent variables have a statistically essential impact on the dependent variable, allowing for variable selection and focusing on the most important predictors. MLR can help uncover complex relationships between variables. It can reveal whether the relationships are linear, positive, negative, or more nuanced. Even though the MLR is a fast technique, however, the geometric mean of NN is better than the MLR. Hence, one may argue that the MLR is biased towards several data and, its precision for the minority is fairly bad.

Disadvantages of multivariable linear regression

This regression relies on several assumptions, including linearity, independence of errors, homoscedasticity (constant variance of errors), and normal distribution of errors. Violation of these assumptions can lead to biased or inaccurate results. Suppose too many independent variables are included in the model. In that case, there is a risk of overfitting, where the algorithm executes well on the training data but fails to generalize to original, hidden data. In the same way, high correlation or multicollinearity among independent variables can lead to difficulties in interpretation and unstable coefficient estimates. It can also make it challenging to identify the individual contributions of each predictor. MLR assumes a linear connection between the dependent and independent variables. If the true relationship is non-linear, MLR may not capture it accurately. Outliers and influential data points can heavily influence the results of MLR, leading to inaccurate parameter estimates and predictions.

4.2.4. Artificial neural networks

Advantages of artificial neural networks

The ANN excels in swiftly addressing challenges characterized by uncertain behaviours or intricate structures, owing to their utilization of non-linear activation functions. Their adaptability stands out prominently; they can dynamically alter their architecture to suit specific usage scenarios, harnessing the inherent cognitive capabilities embedded in their algorithms. The configuration adjustments are steered by the input data traversing the NN, facilitating pattern modifications. The NN's non-linear activation function empowers it to seamlessly handle data across varying dimensions, as long as the input conforms to a continuously differentiable function. ANN is recognized for its high precision and processing capability.

Disadvantages of artificial neural network

The main limitation of ANN is its substantial demand for computational resources, driven by the extensive prerequisites of input data volume. Achieving optimal predictive performance necessitates a considerable volume of training data. A noteworthy concern lies in its susceptibility to the initial randomization of network parameters. Furthermore, the processing time escalates exponentially with an increase in the number of hidden layers, posing a potential efficiency challenge. Moreover, ANN is designed to process numerical data, involving problems to be transformed into numerical values before being input into the network (Khalilov et al., 2021). The selected representation approach plays a vital function in determining the network's performance, which mostly relies on the user's knowledge in choosing a proper encoding approach.

4.3. Bibliometric analysis of performance indicators

Even though machine learning and solar PV systems are conventional areas that have existed since the mid-fifties, the initial article on the energy-connected area was published in 1969; nevertheless, because of several causes, such as inadequate computational power, such systems were very unattractive. However, in 2000, ML began to gain greater attention with computer technology developments, and articles in this field improved by unbelievable development. ML-connected articles presented a sluggish rise from 2000 to 2011, but after 2011, there was a clear explosion in research; the number of articles suddenly improved to the fact that it developed fifteen times the number in 2011 in 12 years. **Figure 8** illustrates the number of articles associated with solar energy and machine learning.

As presented in **Figure 8**, the growth of publications on ML and solar energy can be described in three different stages. The first stage describes the growth period spanning from 1990 to 2009, during which a total of 6 articles were published. Then, in 2010, the second stage of publications began and continued until 2015, producing a significant surge in publications with a total of 62 articles being published. This period is categorized as a progressive stage when European countries and the United States presented reports and strategies aimed at accomplishing buildings with zero energy. Notably, in 2008, the US Department of Energy issued a statement on plans and initiatives related to building technologies to commence between 2008 and 2012 for the actualization of zero-energy buildings (Pyloudi et al., 2015). Moreover, the International Energy Agency (IEA) hosted a global-level scientific platform in 2008 termed "Towards Net Zero Energy Solar Buildings" (Voss and Riley, 2009). Also, in 2007, the agency of the American Society of Heating Refrigerating and Air-Conditioning Engineers established a vision for 2020 to offer technologies that will allow the building community to generate market-feasibility of energy independence by 2030 (ASHRAE, 2020). The last stage can be characterized as a fast expansion period, and an amazing 2184 articles are available. This stage experienced an outstanding publication growth rate of 96.9% from 2016 to 14 January 2025. The full year in 2025 was not explored, hence the decline in the number of publications. To further verify the study methodologies, **Figure 9** is used to illustrate the number of publications by type.

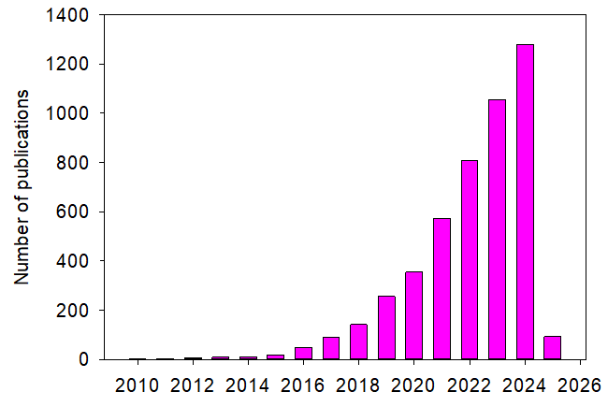

Figure 8. The number of articles associated with machine learning and solar energy from 2010 to 2025.

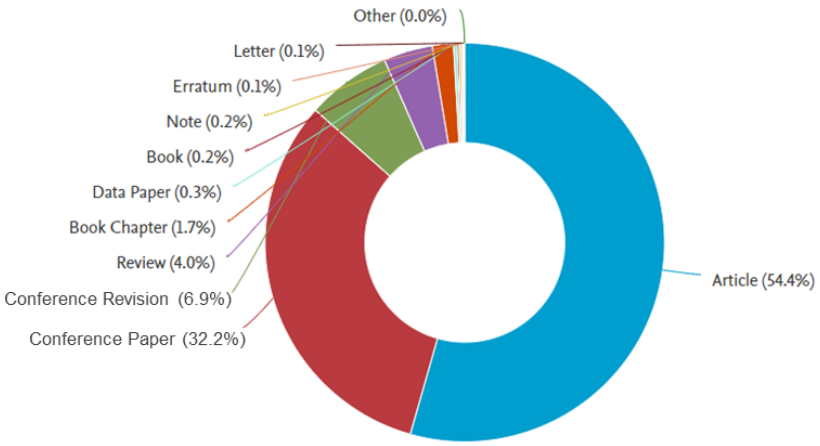

Figure 9. Comparing different publication types of machine learning and solar PV from 2010 to 2025.

Besides, with the rising alarm regarding challenges in air pollution and global warming associated with fossil fuels, several researchers are considering reaching out to the wider world through publications. Hence, articles contribute to 54.4% of literature works, followed by conferences 32.2%. However, Data Papers, Books, Notes, Erratum, and Letters are rarely considered when reaching out to the wider community regarding machine learning for solar energy. Furthermore, **Figure 10** shows documents by subject areas where solar energy links with machine learning.

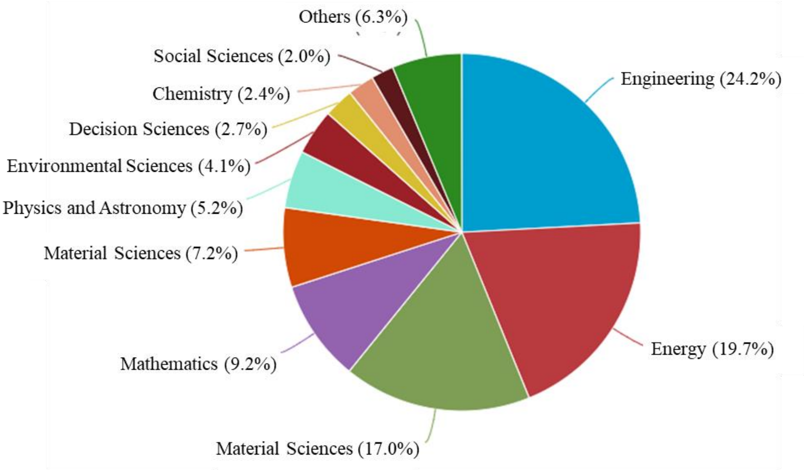

Figure 10. Comparing the document by subject area of machine learning and solar PV from 2010 to 2025.

As can be observed in **Figure 10**, subject areas such as Engineering, accounting for 24.2%, Energy, recording 19.7%, and Computer science, accounting for 17%, have published most of these research papers in the field of ML and solar PV systems.

4.4. Co-occurrence mapping

This research identifies the highly and regularly utilized paper keywords, titles, and author-provided keywords, including machine learning, solar power generation and forecasting. These recognized keywords are then used to produce an inclusive keyword link, enabling correlations among the keywords applied in the research papers. A notable strong point in this connection depends on its competence in assessing papers with these three keywords. Nevertheless, the network map that was created controls the reoccurrence of a keyword in the article and the similarity of other related keywords in the paper. **Figure 11** demonstrates the author-provided keyword links based on machine learning and solar energy within the broader area of energy study.

Five main clusters, each denoting an exact study area, designed the total map in this link. These clusters are Energy efficiency and forecasting (Blue), Network analysis (Green), Unexplored fields of study representing fault detection using machine learning (Purple), Machine learning and application in energy systems (Red) as well as the hot research zones representing (yellow circle) the intersection of other four clusters. However, each node in the network indicates a keyword, and the connections between nodes stand for the relationships between these keywords. Hence, in **Figure 11**, in this visualization, the node represents the most frequent keywords in machine learning and energy systems, as indicated by its larger size and central position. On the other hand, every link shows a co-occurrence of keywords in a similar paper. The thickness of the line characterizes the frequency of co-occurrences, this means that as the lines get thicker, the more often the keywords get closer. As can be observed in **Figure 10**, mapping a network of ML and solar energy provides a prospect to discover new and unexplored fields with the possibility for future research explorations. **Table 2** presents the highly current application of ML in solar energy, finding out hot research areas, using methods and objectives.

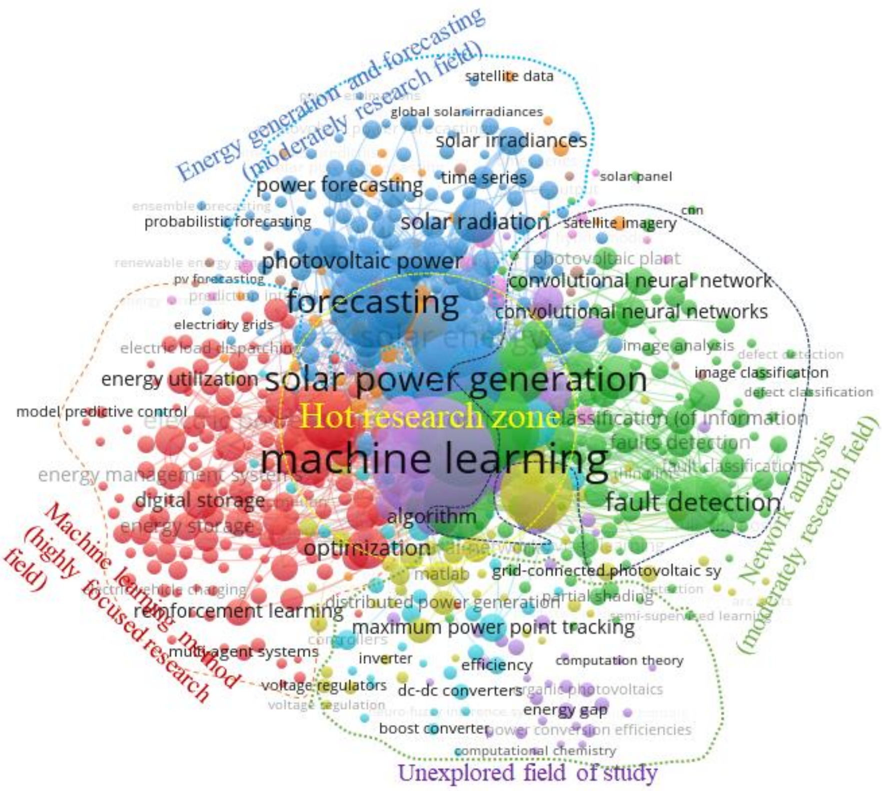

Figure 11. Overlay visualization of author keywords occurrence network.

Table 2. The basic applications of machine learning in solar PV systems.

Targets	Algorithm/methods	References
Forecasting	Time series	Voyant et al., 2017
PV performance assessment	Supportive vector	Bakay and Ağbulut, 2021
Decision making	Numerical model	Ziemba and Szaja, 2023
PV system management	Fuzzy analysis	Mocanu et al., 2018
Sustainable development	Uncertainty analysis	Ismaen et al., 2023
Energy storage	Decision tree	Luo et al., 2021
Building and fault detection	Artificial neural network	Jyothy and Sindhu, 2018

4.5. Study of co-authorship mapping

The ‘bibliographic coupling’ was used to determine countries with the greatest contributions to the growth of machine learning for solar energy. The co-authorship mapping refers to examining the relationships between different items with the same item citations. **Figure 12** illustrates the results of this study, which was conducted, and the document is set to a minimum publication of 6 per country with a total link strength of 1010. The size of every node represents the number of documents published by countries. The results analysis shows that China (281), the USA (241), India (227), Germany (79), Australia (78), Saudi Arabia (78), and the UK (76) represent the most relevant countries with the greatest number of publications published in the area of solar energy for machine learning.

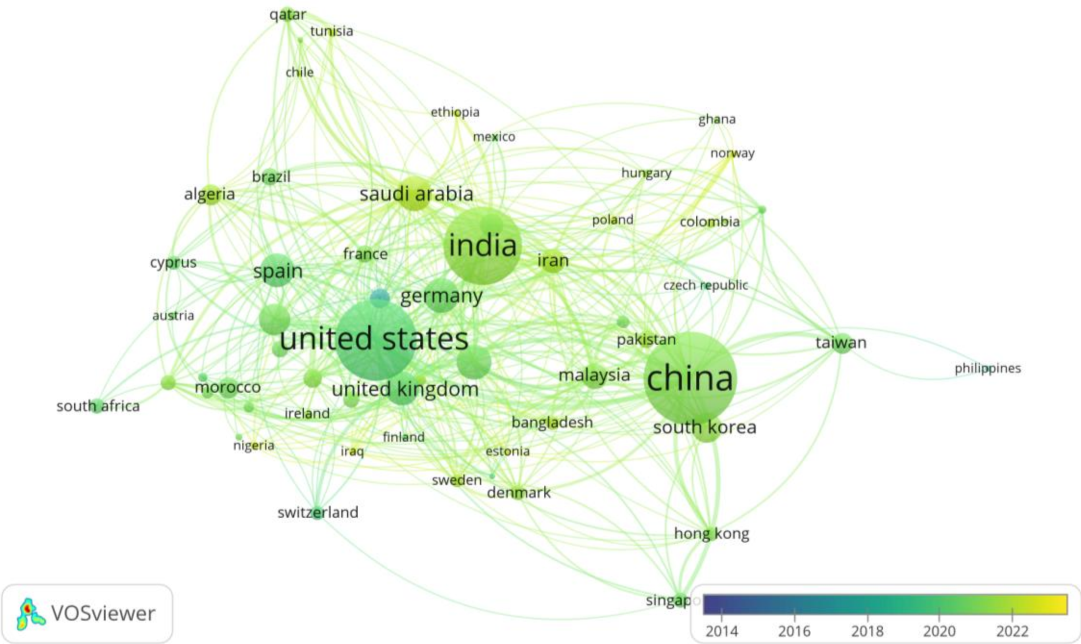

Figure 12. Bibliographic analysis of countries' co-authorship with overlay visualization.

Even though the highest number of publications were produced in developed countries, a huge contribution of research works has likewise been published in developing countries, including India. Nevertheless, the inability of several countries to publish research works can be noticed, for example, in West African countries. In view of the fact expected urbanization and development in developing countries, more studies that centres on energy are essential to contribute to the development of these countries to shift towards energy independence (Apeh and Nwulu, 2024). To further highlight the countries with relevant publications in the field, **Figure 13** compares the number of articles of the top ten countries in the field.

By looking at the number of published works, China has published a vast quantity of ML and solar PV articles. The statistics of countries' quantity of researched works is observed in **Figure 13**, where China recorded 422, accounting for 25.23% of the total, and this is followed by the United States, which published 330 researched papers representing a share of 19.76 % and India recording 277 articles accounting for 16.59% of the total articles. The increasing trend of article publications in the top ten countries can be attributed to the constant government support to alleviate energy poverty and meet the sustainable development goals projected that global sector energy applications will rise to about 30% by 2050 (Zakari et al., 2022).

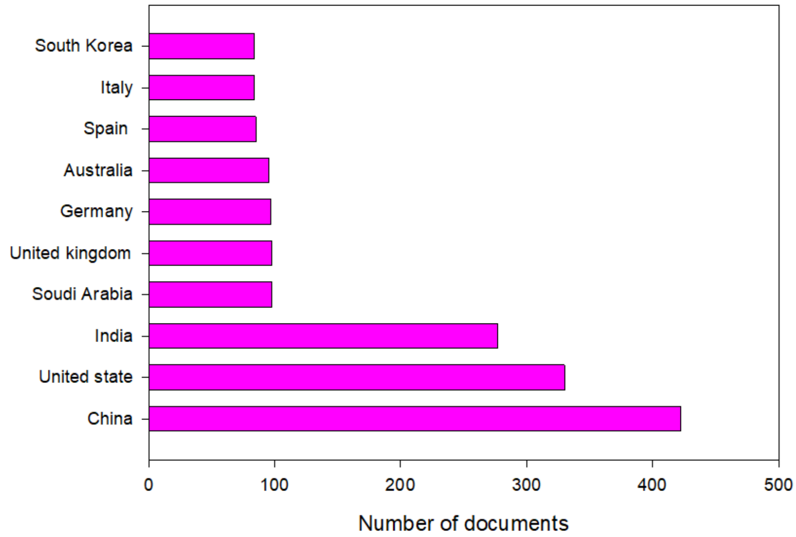

Figure 13. Comparing the top ten countries by the number of articles between 1990 to 2025.

To describe more on the research in energy outlooks, it is also necessary to analyze the top journal publications in this area of study. **Table 3** presents the number of articles published in machine learning for solar energy. Moreover, the research trend focused on studies submitted to several journals. It is observed that work regarding any subject is initially available in journals that character originality. The most current papers in the area of PV systems relating to machine learning between 2010 and 2025 include Energies, with a total article of 150, followed by IEEE Access, recording a total of 64 published papers, Applied Energy has a total of 58 while conference records for IEEE PV specialist conference accounts for a total of 54 articles.

Table 3. The number of articles related to machine learning in the selected energy field from 2010–2025.

Year	Energies	IEEE Access	Solar Energy	Conference records of IEEE PV specialist conference	Applied Energy	Sustainability Switzerland	Energy Report	Energy	Lecture note in Electrical Engineering	Renewable Energy
2010			1							
2011				2						
2012				1						
2013				1						
2014										
2015			2						1	
2016	1		2	3				1		2
2017	3		4		3				2	2
2018	33		4		4			1		4
2019	10	4	7	8	1	1		1		1
2020	18	13	7	12	3	1	2	4	7	3
2021	32	21	4	15	9	8	6	8	5	10
2022	39	11	5	12	20	11	14	7	20	5
2023	14	15	22		6	15	11	10	11	8

With respect to this analysis, it can be concluded that ML systems have advanced in several areas in recent years.

4.6. Using machine learning method for faults diagnosis in PV systems

The research work undertaken by Wu et al. (2009) is an illustration of the initial works in the area of fault diagnosis using machine learning. They utilized the forward error back propagation model (BP) to identify faults in a grid-connected PV system. Their innovative application of the predicted model, which was focused on signal processing, proved a higher possibility. Also, Rezgui et al., (2014) suggested a technique for diagnosing PV system faults. Their anticipated method employed SVM, which was enhanced via the KNN process. Similarly, Wu et al. (2009) identified five fault categories: gadget overheating and irregular grid voltage readings. In contrast, Rezgui et al. (2014) delved into short circuits, along with blocking diodes and bypassing. Moreover, Mostefa Khelil et al. (2020) proposed fault analysis in a grid-connected PV system using a logic block neural network-supported model (Khelil et al., 2020). The model was assigned to two forms of fault, including short circuit failures and string disconnections. Their model had four input parameters. These were the solar radiation, operating temperature, highest current, and voltage characteristics curve. To decrease the error and increase the simplification competence, Kapucu and Cubukcu (2021) presented a model in collaboration with ML applications (Kapucu and Cubukcu, 2021). The research was a collective model to detect both partial shading and short circuit faults. A neuro-fuzzy algorithm to diagnose defects in a PV component was established by Cabeza and Potts (2021), where both current and voltage faults were examined (Cabeza and Potts, 2021). Basnet et al. (2020) discovered the utilization of a smart neural network of probabilistic form in the study to analyze different defects in PV power production with the application of PNN during the winter period. They classified four different faults, including arc, line-to-line, open and short circuit conditions. Hajji et al. (2021), Meyer et al. (2020) examined a fault identification technique for PV units (Hajji et al., 2021; Meyer et al., 2020). Their research utilized supervised ML with a detailed analysis of grid-connected systems. Two faults from the network side and three from the PV sector were considered, comprising inverter and connection defects. On the other hand, connections, partial shading and sensors were considered faults in the PV area. Failure detections in a string of PV modules, including voltage and current variation charges, were accomplished using the MATLAB algorithm interface (Abd el-Ghany et al., 2021; Apeh et al., 2021). This specified technique skillfully categorizes the faults arising from broad and incomplete shading as well as cell and array connections. Lin et al. (2022) presented a model for an enhanced failure diagnosis that stressed a PV array's multiple failures. That was a SE-ResNet network method. The research measured different faults, such as dust, open circuits, line-to-line, shadow on the modules and abnormal degradation. The importance of selecting the ANN category was examined (Khelil et al. 2021). Their research considered PNN, RBF, and generalized and back-propagation sets (GNN and BPNN). The optimum response was observed for RBF, whereas BPNN and GRNN were detected to produce more precise outcomes by assessment with others (Khelil et al. 2021). The study revolved around a grid-connected PV array, with the measured failures encompassing short circuit currents and connection faults.

5. Conclusion, recommendations, and suggestions for future studies

This study presents a thorough framework of photovoltaic fault diagnosis, covering detection and identification, which is vital for protecting photovoltaic systems against different losses in power, efficiency, and reliability. An in-depth analysis of modern literature reveals that support vector machines and artificial neural network stand out as the most commonly utilized machine learning techniques. The majority of machine learning methods present an accuracy exceeding 90%, emphasizing their efficiency in fault diagnosis. Among photovoltaic array faults, SC, OC, and PS are the most widely researched within photovoltaic systems. Nevertheless, challenges persist, such as issues with dataset quality, model configuration, and continuous integration of machine learning methodologies with existing photovoltaic infrastructures. Remarkably, a complete method to addressing serious faults in photovoltaic components, mainly in stand-alone photovoltaic system, appears lacking. Hence, there is an urgent demand for more research to discover the application of these approaches and search into lesser-known algorithms. The findings from the study equally reveal a significant trend of machine learning around 2009 to 2015, when 62 articles were published, especially taking off after 2015 when the number of machine learning and solar photovoltaic publications encountered exponential growth. Over an interval of just nine years, from 2015 to 14 January 2025, the total number of published articles grew to 2184, accounting for 96.9%. Furthermore, the research reveals different promising yet neglected areas where machine learning can be efficiently utilized. These areas comprise maximum power point tracking, energy storage, DC-DC converter, voltage regulator, boost converter, energy supply chain optimization, and demand response. Regarding unexplored fields in machine learning, there are additional prospects in heat storage, environmental management, thermal energy management, uncertainty analysis, and carbon capture techniques. The analysis emphasizes the huge potential for machine learning and solar photovoltaic to contribute to different facets of the energy field, cutting across in improving efficiency to addressing critical environmental challenges and optimizing energy systems.

It is recommended that the governments should encourage and fund research on machine learning to capture the components of photovoltaic involved in failure and for higher assessments of regular patterns. Moreover, effective fault detection and diagnosis methods should be considered by their efficiency, ease of application, fast detection and diagnostic algorithms, generalization capability for large-scale photovoltaic plants (LS-PVP), adaptability to several PV technologies, robust fault classification, the ability to identify numerous faults simultaneously, and the capacity to detect emerging or new faults. Hence, this study suggests that in future work, anomaly detection and fault diagnosis in solar photovoltaic systems will be properly addressed by improving energy prediction models.

Author contributions: Conceptualization, OOA; methodology, OOA; first draft, OOA; final corrections, OOA; supervision, NIN. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

Nomenclature

AAKR	Auto-associative kernel regression
AC	Alternating current
AI	Artificial intelligent
ANFIS	Adaptive neuro-fuzzy inference
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
BIAS	Bias absolute error
BP	Back propagation
BPNN	Back-propagation Neural Network
CO2	Carbon (iv) oxide
DC (A)	Direct current
DNN	Deep neural networks
DT	Decision Trees
EIBCM	Expanded-Improved Bristow-Campbell Model
EL	Extreme learning
ET	Extra Trees
GB	Gradient boosting
GHI	Global horizontal irradiance
GP	Genetic programming
HCSAFRIMA	Harmonically coupled seasonal autoregressive fractionally integrated moving average
IEA	International Energy Agency
Io	Diode saturation current
Iph	Photocurrent
I-V	Current-voltage
IYHM	Improved Yang Hybrid Model
LES	Linear exponential smoothing
LL	Line-to-line
LR	Linear regression
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MBE	Mean bias error
MF	Membership functions
ML	Machine learning
MLP	Multilayer perceptron
MPPT	Maximum power point tracking
NB	Naive Bayes
NMOT	Nominal module operating temperature

NN	Neural network
NOCT	Nominal operating cell temperature
NOx	Nitrogen oxide
NS	Number of cells in series
NWP	Numerical weather prediction
PPE	Percentage Prediction error
PR	Performance ratio
PV	Photovoltaic
RF	Random Forest
RMBE	Root Mean Bias Error
RMSE	Root mean square error
RP	Parallel resistance
Rs	Series resistance
RW	Random walk
SARFIMA	Seasonal autoregressive fractionally integrated moving average
SES	Simple exponential smoothing
SLR	Systematic literature review
SO2	Sulphur (iv) oxide
SOM	Self-Organizing maps
STC	Standard test condition
SVM	Support vector machines
UK	United Kingdom
USA	United States of America
Voc	Open circuit voltage
WT	Wavelet transform

References

Abd el-Ghany, H. A., ELGebaly, A. E., & Taha, I. B. M. (2021). A new monitoring technique for fault detection and classification in PV systems based on rate of change of voltage-current trajectory. International Journal of Electrical Power & Energy Systems, 133, 107248. https://doi.org/10.1016/j.ijepes.2021.107248

Ağbulut, Ü., Gürel, A. E., & Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. Renewable and Sustainable Energy Reviews, 135, 110114. https://doi.org/10.1016/j.rser.2020.110114

Aghaei, M., Kolahi, M., Nedaei, A., et al. (2023). A Holistic Study on Failures and Diagnosis Techniques in Photovoltaic Modules, Components and Systems. In: Proceedings of 2023 International Conference on Future Energy Solutions (FES). pp. 1–6.

Ali, M. U., Khan, H. F., Masud, M., et al. (2020). A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography. Solar Energy, 208, 643–651. https://doi.org/10.1016/j.solener.2020.08.027

Almeida, M. P., Perpiñán, O., & Narvarte, L. (2015). PV power forecast using a nonparametric PV model. Solar Energy, 115, 354–368. https://doi.org/10.1016/j.solener.2015.03.006

Alrahim Shannan, N. M. A., Yahaya, N. Z., & Singh, B. (2013). Single-diode model and two-diode model of PV modules: A comparison. 2013 IEEE International Conference on Control System, Computing and Engineering; 29 November 2013–1 December 2013; Penang, Malaysia. pp. 210–214.

Alzahrani, A., Kimball, J. W., & Dagli, C. (2014). Predicting Solar Irradiance Using Time Series Neural Networks. Procedia Computer Science, 36, 623–628. https://doi.org/10.1016/j.procs.2014.09.065

Amrouche, B., & Le Pivert, X. (2014). Artificial neural network based daily local forecasting for global solar radiation. Applied Energy, 130, 333–341. https://doi.org/10.1016/j.apenergy.2014.05.055

Antonanzas, J., Urraca, R., Pernía-Espinoza, A., et al. (2017). Single and Blended Models for Day-Ahead Photovoltaic Power Forecasting. In: Martínez de Pisón, F., Urraca, R., Quintián, H., Corchado, E. (editors). Hybrid Artificial Intelligent Systems. Springer, Cham.

Apeh, D. Oliver. O., & Nwulu, P. N. (2024). The Food-Energy-Water Nexus Optimization: A Systematic Literature Review. Research on World Agricultural Economy, 247–269. https://doi.org/10.36956/rwae.v5i4.1170

Apeh, O. O., & Nwulu, N. I. (2024). The water-energy-food-ecosystem nexus scenario in Africa: Perspective and policy implementations. Energy Reports, 11, 5947–5962. https://doi.org/10.1016/j.egyr.2024.05.060

Apeh, O. O., & Nwulu, N. I. (2024). Unlocking economic growth: Harnessing renewable energy to mitigate load shedding in Southern Africa. E-Prime—Advances in Electrical Engineering, Electronics and Energy, 10, 100869. https://doi.org/10.1016/j.prime.2024.100869

Apeh, O. O., & Nwulu, N. I. (2025). Improving traceability and sustainability in the agri-food industry through blockchain technology: A bibliometric approach, benefits and challenges. Energy Nexus, 17, 100388. https://doi.org/10.1016/j.nexus.2025.100388

Apeh, O. O., Meyer, E. L., & Overen, O. K. (2021). Modeling and experimental analysis of battery charge controllers for comparing three off-grid photovoltaic power plants. Heliyon, 7(11), e08331. https://doi.org/10.1016/j.heliyon.2021.e08331

Apeh, O. O., Meyer, E. L., & Overen, O. K. (2022). Contributions of Solar Photovoltaic Systems to Environmental and Socioeconomic Aspects of National Development—A Review. Energies, 15(16), 5963. https://doi.org/10.3390/en15165963

Apeh, O. O., Overen, O. K., & Meyer, E. L. (2021). Monthly, Seasonal and Yearly Assessments of Global Solar Radiation, Clearness Index and Diffuse Fractions in Alice, South Africa. Sustainability, 13(4), 2135. https://doi.org/10.3390/su13042135

Ashrae, A. V. (2020). Producing Net Zero Energy Buildings, 2008. Whole Building Design Guide.

Asrari, A., Wu, T. X., & Ramos, B. (2017). A Hybrid Algorithm for Short-Term Solar Power Prediction—Sunshine State Case Study. IEEE Transactions on Sustainable Energy, 8(2), 582–591. https://doi.org/10.1109/tste.2016.2613962

Babu, B. C., & Gurjar, S. (2014). A Novel Simplified Two-Diode Model of Photovoltaic (PV) Module. IEEE Journal of Photovoltaics, 4(4), 1156–1161. https://doi.org/10.1109/jphotov.2014.2316371

Badrudeen, T. U., Nwulu, N. I., & Gbadamosi, S. L. (2023). Neural Network Based Approach for Steady-State Stability Assessment of Power Systems. Sustainability, 15(2), 1667. https://doi.org/10.3390/su15021667

Bakay, M. S., & Ağbulut, Ü. (2021). Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. Journal of Cleaner Production, 285, 125324. https://doi.org/10.1016/j.jclepro.2020.125324

Basaran, K., Özçift, A., & Kılınç, D. (2019). A New Approach for Prediction of Solar Radiation with Using Ensemble Learning Algorithm. Arabian Journal for Science and Engineering, 44(8), 7159–7171. https://doi.org/10.1007/s13369-019-03841-7

Basnet, B., Chun, H., & Bang, J. (2020). An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems. Journal of Sensors, 2020, 1–11. https://doi.org/10.1155/2020/6960328

Cabeza, R. T., & Potts, A. S. (2021). Fault diagnosis and isolation based on Neuro-Fuzzy models applied to a photovoltaic system. IFAC-PapersOnLine, 54(14), 358–363. https://doi.org/10.1016/j.ifacol.2021.10.380

Cervone, G., Clemente-Harding, L., Alessandrini, S., et al. (2017). Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. Renewable Energy, 108, 274–286. https://doi.org/10.1016/j.renene.2017.02.052

Chaouachi, A., Kamel, R. M., & Nagasaka, K. (2010). A novel multi-model neuro-fuzzy-based MPPT for three-phase grid-connected photovoltaic system. Solar Energy, 84(12), 2219–2229. https://doi.org/10.1016/j.solener.2010.08.004

Chennoufi, K., Ferfra, M., & Mokhlis, M. (2021). An accurate modelling of Photovoltaic modules based on two-diode model. Renewable Energy, 167, 294–305. https://doi.org/10.1016/j.renene.2020.11.085

Chérifa, K. M. K., Badia, A., Abou Soufiane, B. (2020). New Intelligent Fault Diagnosis (IFD) approach for grid-connected photovoltaic systems. Energy, 211, 118591. https://doi.org/10.1016/j.energy.2020.118591

Cristaldi, L., Leone, G., & Ottoboni, R. (2017). A hybrid approach for solar radiation and photovoltaic power short-term forecast. In: Proceedingd of 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC); 22–25 May 2017; Turin, Italy. pp. 1–6.

Cuce, E., Cuce, P. M., & Bali, T. (2013). An experimental analysis of illumination intensity and temperature dependency of photovoltaic cell parameters. Applied Energy, 111, 374–382. https://doi.org/10.1016/j.apenergy.2013.05.025

Cuce, E., Cuce, P. M., Karakas, I. H., et al. (2017). An accurate model for photovoltaic (PV) modules to determine electrical characteristics and thermodynamic performance parameters. Energy Conversion and Management, 146, 205–216. https://doi.org/10.1016/j.enconman.2017.05.022

David, L. O., Nwulu, N. I., Aigbavboa, C. O., et al. (2022). Integrating fourth industrial revolution (4IR) technologies into the water, energy & food nexus for sustainable security: A bibliometric analysis. Journal of Cleaner Production, 363, 132522. https://doi.org/10.1016/j.jclepro.2022.132522

De Leone, R., Pietrini, M., & Giovannelli, A. (2015). Photovoltaic energy production forecast using support vector regression. Neural Computing and Applications, 26(8), 1955–1962. https://doi.org/10.1007/s00521-015-1842-y

Dong, Z., Yang, D., Reindl, T., et al. (2013). Short-term solar irradiance forecasting using exponential smoothing state space model. Energy, 55, 1104–1113. https://doi.org/10.1016/j.energy.2013.04.027

Eskandari, A., Milimonfared, J., & Aghaei, M. (2020). Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on I-V characteristics. Solar Energy, 211, 354–365. https://doi.org/10.1016/j.solener.2020.09.071

Fan, J., Wang, X., Wu, L., et al. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Conversion and Management, 164, 102–111. https://doi.org/10.1016/j.enconman.2018.02.087

Fan, J., Wu, L., Zhang, F., et al. (2019). Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. Renewable and Sustainable Energy Reviews, 105, 168–186. https://doi.org/10.1016/j.rser.2019.01.040

Garoudja, E., Chouder, A., Kara, K., et al. (2017). An enhanced machine learning based approach for failures detection and diagnosis of PV systems. Energy Conversion and Management, 151, 496–513. https://doi.org/10.1016/j.enconman.2017.09.019

Gholami, A., Ameri, M., Zandi, M., et al. (2021). A single-diode model for photovoltaic panels in variable environmental conditions: Investigating dust impacts with experimental evaluation. Sustainable Energy Technologies and Assessments, 47, 101392. https://doi.org/10.1016/j.seta.2021.101392

Gholami, A., Ameri, M., Zandi, M., et al. (2022). Electrical, thermal and optical modeling of photovoltaic systems: Step-by-step guide and comparative review study. Sustainable Energy Technologies and Assessments, 49, 101711. https://doi.org/10.1016/j.seta.2021.101711

Gholami, A., Ameri, M., Zandi, M., et al. (2022). Predicting solar photovoltaic electrical output under variable environmental conditions: Modified semi-empirical correlations for dust. Energy for Sustainable Development, 71, 389–405. https://doi.org/10.1016/j.esd.2022.10.012

Gholami, A., Ameri, M., Zandi, M., et al. (2022). Step-By-Step Guide to Model Photovoltaic Panels: An Up-To-Date Comparative Review Study. IEEE Journal of Photovoltaics, 12(4), 915–928. https://doi.org/10.1109/jphotov.2022.3169525

Gholami, A., Ameri, M., Zandi, M., et al. (2023). Impact of harsh weather conditions on solar photovoltaic cell temperature: Experimental analysis and thermal-optical modeling. Solar Energy, 252, 176–194. https://doi.org/10.1016/j.solener.2023.01.039

Gopi, A., Sharma, P., Sudhakar, K., et al. (2022). Weather Impact on Solar Farm Performance: A Comparative Analysis of Machine Learning Techniques. Sustainability, 15(1), 439. https://doi.org/10.3390/su15010439

Hajizadeh, Y. (2019). Machine learning in oil and gas; a SWOT analysis approach. Journal of Petroleum Science and Engineering, 176, 661–663. https://doi.org/10.1016/j.petrol.2019.01.113

Hajji, M., Harkat, M.-F., Kouadri, A., et al. (2021). Multivariate feature extraction based supervised machine learning for fault detection and diagnosis in photovoltaic systems. European Journal of Control, 59, 313–321. https://doi.org/10.1016/j.ejcon.2020.03.004

Harrou, F., Saidi, A., Sun, Y., et al. (2021). Monitoring of Photovoltaic Systems Using Improved Kernel-Based Learning Schemes. IEEE Journal of Photovoltaics, 11(3), 806–818. https://doi.org/10.1109/jphotov.2021.3057169

He, C., Liu, J., Xu, F., et al. (2020). Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods. Energy Conversion and Management, 220, 113111. https://doi.org/10.1016/j.enconman.2020.113111

Hussain, S., & AlAlili, A. (2017). A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks. Applied Energy, 208, 540–550. https://doi.org/10.1016/j.apenergy.2017.09.100

Ibrahim, I. A., & Khatib, T. (2017). A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Conversion and Management, 138, 413–425. https://doi.org/10.1016/j.enconman.2017.02.006

Ismaen, R., Kucukvar, M., El Mekkawy, T. Y., et al. (2023). Optimization and enviro-economic assessment of solar-cooling systems towards sustainable development: A case study of Qatar. Journal of Cleaner Production, 419, 138253. https://doi.org/10.1016/j.jclepro.2023.138253

İzgi, E., Öztopal, A., Yerli, B., et al. (2012). Short–mid-term solar power prediction by using artificial neural networks. Solar Energy, 86(2), 725–733. https://doi.org/10.1016/j.solener.2011.11.013

Jäger-Waldau, A. (2022). Snapshot of photovoltaics—February 2022. EPJ Photovoltaics, 13, 9. https://doi.org/10.1051/epjpv/2022010

Jain, S., & Jain, P. K. (2017). The rise of Renewable Energy implementation in South Africa. Energy Procedia, 143, 721–726. https://doi.org/10.1016/j.egypro.2017.12.752

Jang, H. S., Bae, K. Y., Park, H.-S., et al. (2016). Solar Power Prediction Based on Satellite Images and Support Vector Machine. IEEE Transactions on Sustainable Energy, 7(3), 1255–1263. https://doi.org/10.1109/tste.2016.2535466

Jia, D., Yang, L., Lv, T., et al. (2022). Evaluation of machine learning models for predicting daily global and diffuse solar radiation under different weather/pollution conditions. Renewable Energy, 187, 896–906. https://doi.org/10.1016/j.renene.2022.02.002

Joshua, S. R., Yeon, A. N., Park, S., et al. (2024). A Hybrid Machine Learning Approach: Analyzing Energy Potential and Designing Solar Fault Detection for an AIoT-Based Solar–Hydrogen System in a University Setting. Applied Sciences, 14(18), 8573. https://doi.org/10.3390/app14188573

Jović, S., Aničić, O., Marsenić, M., et al. (2016). Solar radiation analyzing by neuro-fuzzy approach. Energy and Buildings, 129, 261–263. https://doi.org/10.1016/j.enbuild.2016.08.020

Jyothy, L. P. N., Sindhu, M. R. (2018). An artificial neural network based MPPT algorithm for solar PV system. In: Proceedings of 2018 4th International Conference on Electrical Energy Systems (ICEES); 7–9 February 2018; Chennai, India. pp. 375–380.

Kaba, K., Sarıgül, M., Avcı, M., et al. (2018). Estimation of daily global solar radiation using deep learning model. Energy, 162, 126–135. https://doi.org/10.1016/j.energy.2018.07.202

Kapucu, C., & Cubukcu, M. (2021). A supervised ensemble learning method for fault diagnosis in photovoltaic strings. Energy, 227, 120463. https://doi.org/10.1016/j.energy.2021.120463

Kara Mostefa Khelil, C., Amrouche, B., Kara, K., et al. (2021). The impact of the ANN's choice on PV systems diagnosis quality. Energy Conversion and Management, 240, 114278. https://doi.org/10.1016/j.enconman.2021.114278

Karamizadeh, S., Abdullah, S. M., Halimi, M., et al. (2014). Advantage and drawback of support vector machine functionality. In: Proceedings of 2014 International Conference on Computer, Communications, and Control Technology (I4CT); 2–4 September 2014; Langkawi, Malaysia. pp. 63–65.

Khalilov, D. A., Jumaboyeva, N. A. K., Kurbonova, T. M. K. (2021). Advantages and Applications of Neural Networks. Acadamic Research of Education Science, 2(2), 1153–1159.

Khandakar, A., E. H. Chowdhury, M., Khoda Kazi, M.-, Benhmed, K., et al. (2019). Machine Learning Based Photovoltaics (PV) Power Prediction Using Different Environmental Parameters of Qatar. Energies, 12(14), 2782. https://doi.org/10.3390/en12142782

Kim, J., Kim, D., Yoo, W., et al. (2017). Daily prediction of solar power generation based on weather forecast information in Korea. IET Renewable Power Generation, 11(10), 1268–1273. Portico. https://doi.org/10.1049/iet-rpg.2016.0698

Kwon, Y., Kwasinski, A., & Kwasinski, A. (2019). Solar Irradiance Forecast Using Naïve Bayes Classifier Based on Publicly Available Weather Forecasting Variables. Energies, 12(8), 1529. https://doi.org/10.3390/en12081529

Lazzaretti, A. E., Costa, C. H. da, Rodrigues, M. P., Yamada, G. D., et al. (2020). A Monitoring System for Online Fault Detection and Classification in Photovoltaic Plants. Sensors, 20(17), 4688. https://doi.org/10.3390/s20174688

Li, H., Liu, Z., Liu, K., et al. (2017). Predictive Power of Machine Learning for Optimizing Solar Water Heater Performance: The Potential Application of High-Throughput Screening. International Journal of Photoenergy, 2017, 1–10. https://doi.org/10.1155/2017/4194251

Li, H., Zhang, Z., & Liu, Z. (2017). Application of Artificial Neural Networks for Catalysis: A Review. Catalysts, 7(10), 306. https://doi.org/10.3390/catal7100306

Li, H., Zhang, Z., & Zhao, Z.-Z. (2019). Data-Mining for Processes in Chemistry, Materials, and Engineering. Processes, 7(3), 151. https://doi.org/10.3390/pr7030151

Lin, P., Qian, Z., Lu, X., et al. (2022). Compound fault diagnosis model for Photovoltaic array using multi-scale SE-ResNet. Sustainable Energy Technologies and Assessments, 50, 101785. https://doi.org/10.1016/j.seta.2021.101785

Liu, Z., Li, H., Tang, X., et al. (2016). Extreme learning machine: a new alternative for measuring heat collection rate and heat loss coefficient of water-in-glass evacuated tube solar water heaters. SpringerPlus, 5(1). https://doi.org/10.1186/s40064-016-2242-1

Liu, Z., Liu, K., Li, H., et al. (2015). Artificial Neural Networks-Based Software for Measuring Heat Collection Rate and Heat Loss Coefficient of Water-in-Glass Evacuated Tube Solar Water Heaters. PLOS ONE, 10(12), e0143624. https://doi.org/10.1371/journal.pone.0143624

Luo, X., Xia, J., & Liu, Y. (2021). Extraction of dynamic operation strategy for standalone solar-based multi-energy systems: A method based on decision tree algorithm. Sustainable Cities and Society, 70, 102917. https://doi.org/10.1016/j.scs.2021.102917

Malik, A., Haque, A., Satya Bharath, K. V., Jaffery, Z. A. (2021). Transfer Learning-Based Novel Fault Classification Technique for Grid-Connected PV Inverter. In: Mekhilef, S., Favorskaya, M., Pandey, R. K., Shaw, R. N. (editors). Innovations in Electrical and Electronic Engineering. Springer. pp. 217–224.

Maluta, E. N., Mulaudzi, S. T. (2018). Evaluation of the temperature based models for the estimation of global solar radiation in Pretoria, Gauteng province of South Africa. International Energy Journal, 18(2).

Mandal, P., Madhira, S. T. S., haque, A. U., et al. (2012). Forecasting Power Output of Solar Photovoltaic System Using Wavelet Transform and Artificial Intelligence Techniques. Procedia Computer Science, 12, 332–337. https://doi.org/10.1016/j.procs.2012.09.080

Martín, L., Zarzalejo, L. F., Polo, J., et al. (2010). Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. Solar Energy, 84(10), 1772–1781. https://doi.org/10.1016/j.solener.2010.07.002

Mellit, A., & Kalogirou, S. (2022). Assessment of machine learning and ensemble methods for fault diagnosis of photovoltaic systems. Renewable Energy, 184, 1074–1090. https://doi.org/10.1016/j.renene.2021.11.125

Mellit, A., Pavan, A. M., & Benghanem, M. (2012). Least squares support vector machine for short-term prediction of meteorological time series. Theoretical and Applied Climatology, 111(1–2), 297–307. https://doi.org/10.1007/s00704-012-0661-7

Mellit, A., Tina, G. M., & Kalogirou, S. A. (2018). Fault detection and diagnosis methods for photovoltaic systems: A review. Renewable and Sustainable Energy Reviews, 91, 1–17. https://doi.org/10.1016/j.rser.2018.03.062

Merza, B. N., Wadday, A. G., & Abdullah, A. K. (2024). Study and analysis of fault detection in solar array system based on internet of things. AIP Conference Proceedings, 3092(1).

Meyer, E. L., Apeh, O. O., & Overen, O. K. (2020). Electrical and Meteorological Data Acquisition System of a Commercial and Domestic Microgrid for Monitoring PV Parameters. Applied Sciences, 10(24), 9092. https://doi.org/10.3390/app10249092

Mocanu, E., Nguyen, P. H., & Gibescu, M. (2018). Deep Learning for Power System Data Analysis. Big Data Application in Power Systems, 125–158. https://doi.org/10.1016/b978-0-12-811968-6.00007-3

Mohamed, A., Hamdi, M. S., & Tahar, S. (2015). A Machine Learning Approach for Big Data in Oil and Gas Pipelines. In: Proceedings of 2015 3rd International Conference on Future Internet of Things and Cloud; 24–26 August 2015; Rome, Italy. pp. 585–590.

Muhammad Ehsan, R., Simon, S. P., & Venkateswaran, P. R. (2016). Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. Neural Computing and Applications, 28(12), 3981–3992. https://doi.org/10.1007/s00521-016-2310-z

Mustafa, Z., Awad, A. S. A., Azzouz, M., et al. (2023). Fault identification for photovoltaic systems using a multi-output deep learning approach. Expert Systems with Applications, 211, 118551. https://doi.org/10.1016/j.eswa.2022.118551

Naderi, S., Banifateme, M., Pourali, O., et al. (2020). Accurate capacity factor calculation of waste-to-energy power plants based on availability analysis and design/off-design performance. Journal of Cleaner Production, 275, 123167. https://doi.org/10.1016/j.jclepro.2020.123167

Nižetić, S., Jurčević, M., Čoko, D., Arıcı, M. (2021a). A novel and effective passive cooling strategy for photovoltaic panel. Renewable and Sustainable Energy Reviews, 145, 111164. https://doi.org/10.1016/j.rser.2021.111164

Nižetić, S., Jurčević, M., Čoko, D., et al. (2021b). Implementation of phase change materials for thermal regulation of photovoltaic thermal systems: Comprehensive analysis of design approaches. Energy, 228, 120546. https://doi.org/10.1016/j.energy.2021.120546

Odabaşı, Ç., & Yıldırım, R. (2020). Machine learning analysis on stability of perovskite solar cells. Solar Energy Materials and Solar Cells, 205, 110284. https://doi.org/10.1016/j.solmat.2019.110284

Olatomiwa, L., Mekhilef, S., Shamshirband, S., et al. (2015). A support vector machine–firefly algorithm-based model for global solar radiation prediction. Solar Energy, 115, 632–644. https://doi.org/10.1016/j.solener.2015.03.015

Olatomiwa, L., Mekhilef, S., Shamshirband, S., et al. (2015). Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. Renewable and Sustainable Energy Reviews, 51, 1784–1791. https://doi.org/10.1016/j.rser.2015.05.068

Olorunfemi, B. O., Nwulu, N. I., & Ogbolumani, O. A. (2023). Solar panel surface dirt detection and removal based on arduino color recognition. MethodsX, 10, 101967. https://doi.org/10.1016/j.mex.2022.101967

Pandiyan, P., Sitharthan, R., Saravanan, S., et al. (2022). A comprehensive review of the prospects for rural electrification using stand-alone and hybrid energy technologies. Sustainable Energy Technologies and Assessments, 52, 102155. https://doi.org/10.1016/j.seta.2022.102155

Persson, C., Bacher, P., Shiga, T., et al. (2017). Multi-site solar power forecasting using gradient boosted regression trees. Solar Energy, 150, 423–436. https://doi.org/10.1016/j.solener.2017.04.066

Prasad, R., Ali, M., Kwan, P., et al. (2019). Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. Applied Energy, 236, 778–792. https://doi.org/10.1016/j.apenergy.2018.12.034

Pyloudi, E., Papantoniou, S., & Kolokotsa, D. (2014). Retrofitting an office building towards a net zero energy building. Advances in Building Energy Research, 9(1), 20–33. https://doi.org/10.1080/17512549.2014.917985

Quej, V. H., Almorox, J., Arnaldo, J. A., et al. (2017). ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. Journal of Atmospheric and Solar-Terrestrial Physics, 155, 62–70. https://doi.org/10.1016/j.jastp.2017.02.002

Ramli, M. A. M., Twaha, S., & Al-Turki, Y. A. (2015). Investigating the performance of support vector machine and artificial neural networks in predicting solar radiation on a tilted surface: Saudi Arabia case study. Energy Conversion and Management, 105, 442–452. https://doi.org/10.1016/j.enconman.2015.07.083

Ramsami, P., & Oree, V. (2015). A hybrid method for forecasting the energy output of photovoltaic systems. Energy Conversion and Management, 95, 406–413. https://doi.org/10.1016/j.enconman.2015.02.052

Ranganai, E., & Sigauke, C. (2020). Capturing Long-Range Dependence and Harmonic Phenomena in 24-Hour Solar Irradiance Forecasting: A Quantile Regression Robustification via Forecasts Combination Approach. IEEE Access, 8, 172204–172218. https://doi.org/10.1109/access.2020.3024661

Rashidi, S., Esfahani, J. A., & Hosseinirad, E. (2021). Assessment of solar chimney combined with phase change materials. Journal of the Taiwan Institute of Chemical Engineers, 124, 341–350. https://doi.org/10.1016/j.jtice.2021.03.001

Rezgui, W., Mouss, L.-H., Mouss, N. K., et al. (2014). A smart algorithm for the diagnosis of short-circuit faults in a photovoltaic generator. In: Proceedings of 2014 First International Conference on Green Energy (ICGE 2014); 25–27 March 2014; Sfax, Tunisia. pp. 139–143.

Sabri, N., Tlemcani, A., & Chouder, A. (2018). Intelligent fault supervisory system applied on stand-alone photovoltaic system. In: Proceedings of 2018 International Conference on Applied Smart Systems (ICASS); 24–25 November 2018; Medea, Algeria. pp. 1–5.

Salcedo-Sanz, S., Deo, R. C., Cornejo-Bueno, L., et al. (2018). An efficient neuro-evolutionary hybrid modelling mechanism for the estimation of daily global solar radiation in the Sunshine State of Australia. Applied Energy, 209, 79–94. https://doi.org/10.1016/j.apenergy.2017.10.076

Sangeetha, M., Manigandan, S., Ashok, B., et al. (2021). Experimental investigation of nanofluid based photovoltaic thermal (PV/T) system for superior electrical efficiency and hydrogen production. Fuel, 286, 119422. https://doi.org/10.1016/j.fuel.2020.119422

Sareh, D., Rahim, Z., & Omid Noudeh, F. (2022). Evaluation of the concentration of suspended particles in underground subway stations in Tehran and its comparison with ambient concentrations. Annals of Environmental Science and Toxicology, 6(1), 019–025. https://doi.org/10.17352/aest.000048

Savaresi, A. (2013). Just Another Climate Conference. Journal of Environmental Law and Policy, 43, 284.

Shahverdian, M. H., Sohani, A., Sayyaadi, H., et al. (2021). A dynamic multi-objective optimization procedure for water cooling of a photovoltaic module. Sustainable Energy Technologies and Assessments, 45, 101111. https://doi.org/10.1016/j.seta.2021.101111

Sharma, A., & Kakkar, A. (2018). Forecasting daily global solar irradiance generation using machine learning. Renewable and Sustainable Energy Reviews, 82, 2254–2269. https://doi.org/10.1016/j.rser.2017.08.066

Sharma, N., Sharma, P., Irwin, D., et al. (2011). Predicting solar generation from weather forecasts using machine learning. In: Proceedings of 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm); 17–20 October 2011; Brussels, Belgium. pp. 528–533.

Sireci, O. (2006). Turkish State Meteorological Service Radar Network Feasibility Studies. Technical Report Turkish State Meteorological Service.

Sohani, A., Naderi, S., Torabi, F., et al. (2020a). Application based multi-objective performance optimization of a proton exchange membrane fuel cell. Journal of Cleaner Production, 252, 119567. https://doi.org/10.1016/j.jclepro.2019.119567

Sohani, A., Sayyaadi, H., Cornaro, C., et al. (2022). Using machine learning in photovoltaics to create smarter and cleaner energy generation systems: A comprehensive review. Journal of Cleaner Production, 364, 132701. https://doi.org/10.1016/j.jclepro.2022.132701

Sohani, A., Shahverdian, M. H., Sayyaadi, H., et al. (2020b). Impact of absolute and relative humidity on the performance of mono and poly crystalline silicon photovoltaics; applying artificial neural network. Journal of Cleaner Production, 276, 123016. https://doi.org/10.1016/j.jclepro.2020.123016

Souley Agbodjan, Y., Wang, J., Cui, Y., et al. (2022). Bibliometric analysis of zero energy building research, challenges and solutions. Solar Energy, 244, 414–433. https://doi.org/10.1016/j.solener.2022.08.061

Venkatakrishnan, G. R., Rengaraj, R., Tamilselvi, S., et al. (2023). Detection, location, and diagnosis of different faults in large solar PV system—a review. International Journal of Low-Carbon Technologies, 18, 659–674. https://doi.org/10.1093/ijlct/ctad018

Venkatasubramanian, V. (2018). The promise of artificial intelligence in chemical engineering: Is it here, finally? AIChE Journal, 65(2), 466–478. Portico. https://doi.org/10.1002/aic.16489

Voss, K., Riley, M. (2009). IEA joint project: towards net zero energy solar buildings (NZEBs). International Energy Agency.

Voyant, C., Notton, G., Kalogirou, S., et al. (2017). Machine learning methods for solar radiation forecasting: A review. Renewable Energy, 105, 569–582. https://doi.org/10.1016/j.renene.2016.12.095

Wang, F., Zhen, Z., Mi, Z., et al. (2015). Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. Energy and Buildings, 86, 427–438. https://doi.org/10.1016/j.enbuild.2014.10.002

Wolff, B., Kühnert, J., Lorenz, E., et al. (2016). Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. Solar Energy, 135, 197–208. https://doi.org/10.1016/j.solener.2016.05.051

Wu, Y., Lan, Q., & Sun, Y. (2009). Application of BP neural network fault diagnosis in solar photovoltaic system. In: Proceedings of 2009 International Conference on Mechatronics and Automation; 9–12 August 2009; Changchun, China. pp. 2581–2585.

Yaïci, W., & Entchev, E. (2016). Adaptive Neuro-Fuzzy Inference System modelling for performance prediction of solar thermal energy system. Renewable Energy, 86, 302–315. https://doi.org/10.1016/j.renene.2015.08.028

Zakari, A., Khan, I., Tan, D., et al. (2022). Energy efficiency and sustainable development goals (SDGs). Energy, 239, 122365. https://doi.org/10.1016/j.energy.2021.122365

Zamo, M., Mestre, O., Arbogast, P., et al. (2014). A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. Solar Energy, 105, 792–803. https://doi.org/10.1016/j.solener.2013.12.006

Zang, H., Xu, Q., & Bian, H. (2012). Generation of typical solar radiation data for different climates of China. Energy, 38(1), 236–248. https://doi.org/10.1016/j.energy.2011.12.008

Zenebe, T. M., Midtgård, O. M., Völler, S., Cali, Ü. (2022). Machine Learning for PV System Operational Fault Analysis: Literature Review. In: Sanfilippo, F., Granmo, O. C., Yayilgan, S. Y., Bajwa, I. S. (editors). Intelligent Technologies and Applications. Springer. pp. 337–351.

Zhang, Q., Cui, N., Feng, Y., et al. (2018). Comparative Analysis of Global Solar Radiation Models in Different Regions of China. Advances in Meteorology, 2018, 1–21. https://doi.org/10.1155/2018/3894831

Ziemba, P., & Szaja, M. (2023). Fuzzy Decision-Making Model for Solar Photovoltaic Panel Evaluation. Energies, 16(13), 5161. https://doi.org/10.3390/en16135161

Zou, L., Wang, L., Xia, L., et al. (2017). Prediction and comparison of solar radiation using improved empirical models and Adaptive Neuro-Fuzzy Inference Systems. Renewable Energy, 106, 343–353. https://doi.org/10.1016/j.renene.2017.01.042