machines

MDPI

Article

Hybrid Deep Learning for Fault Diagnosis in Photovoltaic Systems

Mouaad Bougoffa [1], Samir Benmoussa [1], Mohand Djeziri [2,*] and Olivier Palais [2]

[1] Laboratory of Automatics and Signals of Annaba (LASA), Badji Mokhtar-Annaba University, P.O. Box 12, Annaba 23000, Algeria; mouaad.bougoffa@univ-annaba.dz (M.B.); samir.benmoussa@univ-annaba.dz (S.B.)
[2] Aix Marseille Univ, Université de Toulon, CNRS, IM2NP, Marseille, France; olivier.palais@im2np.fr
* Correspondence: mohand.djeziri@im2np.fr

Abstract: Photovoltaic (PV) systems are integral to global renewable energy generation, yet their efficiency and reliability are frequently compromised by undetected faults, leading to significant energy losses, increased maintenance costs, and reduced operational lifespans. To address these challenges, this study proposes a novel hybrid deep learning framework that combines Stacked Sparse Auto-Encoders (SSAE) for autonomous feature extraction with an Optimized-Multi-Layer Perceptron (OMLP) for precise fault classification. The SSAE extracts high-dimensional fault features from raw operational data, while the OMLP leverages these features to classify faults with exceptional accuracy. The model was rigorously validated using real-world PV datasets, encompassing diverse fault types such as partial shading, open circuits, and module degradation under dynamic environmental conditions. Results demonstrate state-of-the-art performance, with the model achieving 99.82% accuracy, 99.7% precision, 99.4% sensitivity, and 100% specificity, outperforming traditional machine learning and deep learning approaches. These findings highlight the framework's robustness and reliability in real-world applications. By significantly enhancing fault detection accuracy and computational efficiency, the proposed approach optimizes PV system performance, reduces operational costs, and supports sustainable energy production. This study concludes that the hybrid SSAE-Optimized MLP model represents a scalable and efficient solution for improving the reliability and longevity of renewable energy infrastructure, setting a new benchmark for intelligent maintenance strategies in the field.

Keywords: photovoltaic system; fault detection; fault diagnosis; stacked sparse autoencoder; optimized multi-layer perceptron

check for updates

Academic Editor: Raul D. S. G. Campilho

Received: 19 March 2025
Revised: 24 April 2025
Accepted: 29 April 2025
Published: 30 April 2025

Citation: Bougoffa, M.; Benmoussa, S.; Djeziri, M.; Palais, O. Hybrid Deep Learning for Fault Diagnosis in Photovoltaic Systems. *Machines* **2025**, *13*, 378. https://doi.org/10.3390/machines13050378

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Artificial Intelligence (AI) has emerged as a cornerstone of 21st-century innovation, driving transformative advancements across industries such as healthcare, finance, and energy [1,2]. By enabling systems to analyze data, recognize patterns, and make informed decisions, AI methodologies such as machine learning (ML) and deep learning (DL) have unlocked unprecedented opportunities to address complex challenges and optimize operational workflows [3]. Nowhere is this potential more critical than in renewable energy, where AI is revolutionizing the efficiency and reliability of infrastructure, particularly in solar PV systems.

PV systems, which convert sunlight into electricity, are pivotal to the global transition toward sustainable energy. Despite their growing adoption—from residential installations to utility-scale solar farms—their performance and longevity are hindered by operational

challenges such as module degradation, shading effects, inverter failures, and environmental variability [4]. PV array faults such as Open Circuits (OC), Short Circuits (SC), and environmental faults such as Partial Shading (PS) are particularly insidious, often leading to energy losses, increased maintenance costs, and reduced return on investment [5,6].

A 2010 study demonstrated that faults and failures in solar PV systems can lead to an annual energy output reduction of 18.9% [7], while harsh environmental conditions exacerbate efficiency drops to 15–20% [8]. For instance, prolonged exposure to extreme climates accelerates module degradation, with reported annual degradation rates of 0.923% in desert environments [9]. Dust accumulation further compounds these losses: experimental data from Saharan regions shows 8.41% power loss over eight weeks due to soiling [10], while uncleaned systems in Mediterranean climates suffer 9.99% instantaneous power drops and 2.93% average reductions [11]. Partial shading—often caused by cloud cover or debris—can obscure up to 75% of PV array surfaces, creating localized hotspots and mismatch losses [12]. Mitigation strategies, such as PV array reconfiguration and tilt-angle optimization, demonstrate promising results, with adjusted tilt angles boosting power output by 50.36% in some regions [13]. These findings underscore the critical need for intelligent fault diagnosis frameworks to minimize energy losses, extend system lifespans, and maximize return on investment in solar infrastructure.

Traditional fault diagnosis in PV systems relies on analyzing electrical parameters such as voltage ($V\sim dc\sim$), current ($I\sim dc\sim$), and irradiance (Irr), often collected via sensors such as maximum power point trackers (MPPTs) and environmental monitors. Rule-based algorithms and physical inspections remain prevalent but are labor-intensive and lack scalability. Advanced methods include infrared thermography for hotspot detection and electroluminescence imaging for microcracks. However, these techniques require specialized equipment and are unsuitable for real-time monitoring.

The integration of AI into PV fault diagnosis represents a paradigm shift in renewable energy supervision [14], enabling scalable and automated solutions. Recent advancements leverage ML models such as Support Vector Machines (SVMs), Random Forests (RFs), and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to process historical and real-time data from supervisory control and data acquisition (SCADA) systems [15–19]. Despite progress, these methods struggle with high-dimensional data and dynamic environmental variability, underscoring the need for hybrid AI-driven approaches.

For instance, AdaBoost achieved 0.95 precision in fault classification using temperature and radiation features [20], while kernel-based extreme learning machines [21] and probabilistic neural networks (PNNs) [22] showed high precision in identifying short-circuited modules. However, traditional ML methods struggle with overfitting, high-dimensional data, and scalability [23,24].

DL, a subset of ML, addresses these limitations through hierarchical feature extraction. Hybrid architectures, such as Convolutional Neural Networks (CNNs) combined with residual gated recurrent units (ResGRUs), classify I-V curve anomalies with 98.61% accuracy [25], while Stacked Auto-Encoders (SAE) enable fault diagnosis in PV arrays through I-V characteristic analysis [26,27]. Similarly, particle swarm optimization (PSO)-enhanced long short-term memory (LSTM) networks improve fault detection reliability by optimizing parameter tuning and temporal pattern recognition [28,29]. Autoencoder (AE)-based frameworks, including LSTM, CNN, and variational, sparse, or denoising variants, are widely applied for fault detection across systems [30,31]. For instance, Graph Convolutional Variational AutoEncoders (GCVAE) model complex spatial–temporal dependencies to identify anomalous patterns [32], further advancing detection capabilities. Despite these advances, critical gaps persist: (1) many models require large labeled datasets and lack generalizabil-

ity across PV designs [33]; (2) computational complexity hinders real-time deployment [34]; and (3) integration with cost-effective monitoring systems remains underexplored.

To bridge these gaps, this paper proposes a novel hybrid deep learning architecture combining SSAE and Optimized-MLP. The SSAE extracts high-dimensional fault features from raw operational data (e.g., current, voltage, irradiance), capturing subtle patterns associated with faults such as partial shading and degradation. The OMLP then classifies these features with minimal computational overhead, enabling efficient diagnosis while maintaining accuracy [35]. This approach diverges from prior work, such as [36], which relies on hybrid CNN-Bi-GRU models constrained by database quality, and [37], which depends on simulated data for MLP-based classification. By balancing SSAE's robust feature extraction with MLP's lightweight classification, our hybrid architecture introduces three key innovations: (1) a KL divergence-regularized SSAE that enforces sparsity during unsupervised feature learning, enhancing sensitivity to rare faults under class imbalance; (2) a Bayesian-optimized MLP with weighted loss functions, automating hyperparameter selection while mitigating bias toward majority classes; and (3) end-to-end integration of environmental variables (irradiance, temperature) with electrical parameters, enabling robust diagnosis under dynamic outdoor conditions. Compared to CNN-Bi-GRU models—which incur high computational costs from convolutional operations and bidirectional temporal processing—our SSAE-OMLP achieves superior accuracy (99.82%) on tabular PV data while reducing inference latency, making it deployable on low-cost edge hardware.

Using empirical data from operational PV systems, we evaluate the SSAE-OMLP framework against traditional methods (SVM, RF) and state-of-the-art DL architectures (CNN, LSTM). Our contributions include:

1. A comparative analysis of DL models under dynamic stresses (e.g., fluctuating irradiance, partial shading).
2. A hybrid SSAE-OMLP framework optimized for computational efficiency, enabling rapid fault detection under dynamic environmental conditions.
3. Empirical validation of scalability for large-scale deployments, addressing a critical gap in prior research [38].

The results demonstrate that the SSAE-Optimized-MLP hybrid reduces false positives by 22%, achieves 99.82% accuracy in fault classification, and operates 35% faster than conventional CNN-based models. By minimizing energy losses and maintenance costs, this work advances the reliability of PV systems, directly supporting global sustainability goals. Furthermore, it underscores how AI-driven digital transformation can enhance project leadership in renewable energy, ensuring the successful deployment of innovative technologies in real-world settings.

2. Photovoltaic Fault Types

Faults in PV systems are categorized based on their origin within the system's Direct Current (DC) or Alternating Current (AC) components. AC-side faults typically stem from inverter malfunctions or grid-related disturbances. In contrast, DC-side faults are more diverse, encompassing issues such as MPPT algorithm inefficiencies, bypass diode failures, ground faults, arc faults, module mismatch, OC, and SC faults [6,39].

2.1. Arc Faults

In photovoltaic systems, arc faults occur when a high-energy electrical discharge jumps through air or across a gap between conductive parts. Such arcs can be triggered by issues such as deteriorated insulation, loose or corroded connections, damaged cables, or failed components [40].

2.2. Open-Circuit Faults

These occur when a disconnection disrupts current flow within the system. Such faults significantly reduce power generation, potentially affecting individual module strings or the entire system, depending on the disconnection's location and the PV array's topology [41].

2.3. Short-Circuit Faults

These arise from unintended electrical connections between two points in a PV array at different potentials. These faults may manifest within a single string or between adjacent strings in the system [41].

2.4. Ground Faults

These occur when the PV array unintentionally makes contact with earth, creating safety hazards, reducing operational efficiency, and risking damage to the system's components [42].

2.5. Line-to-Line Faults

These faults happen when two conductors in a PV system become directly short-circuited. They typically involve a direct connection between any two of the system's three AC output phases—often referred to as a three-phase fault [43].

2.6. Module/Cell Mismatch

This fault arises when cells or modules exhibit divergent electrical properties, compromising system efficiency. Mismatch is subdivided into temporary and permanent categories. Temporary mismatch is caused by transient factors such as dust/snow accumulation or dynamic shading from nearby structures, whereas permanent mismatch results from long-term degradation or physical damage to modules. Where temporary mismatch is exemplified by partial shading, and permanent mismatch by module degradation [8].

2.7. Partial Shading

Occurs when parts of the PV modules are obstructed and cannot receive full sunlight, temporarily reducing their power output. Shading can be either static or dynamic. Static shading is caused by long-term deposits on the module surface—such as dust, leaves, or bird droppings, and dynamic shading comes from fleeting shadows cast by nearby trees or buildings [12].

2.8. Degradation

Over time, PV systems lose performance due to several factors, including surface contamination, UV-induced optical wear, increases in series or decreases in shunt resistance, and declines in SC current [44]. This deterioration can be confined to individual modules or progressively affect the entire array, leading to an overall drop in energy production.

The experimental dataset in this study focuses on four critical DC-side faults in PV systems: OC, SC, PS and module degradation. These faults were selected due to their recurrent occurrence and significant impact on system performance.

3. Materials and Methods

This section outlines the hybrid approach used for fault diagnosis in PV systems, combining Stacked Sparse Auto-Encoders for feature extraction with a Multi-Layer Perceptron classifier for fault categorization as presented in Figure 1. The proposed methodology involves several stages: data preprocessing, feature extraction using SSAE, fault classification using Optimized-MLP, and performance evaluation using various metrics.

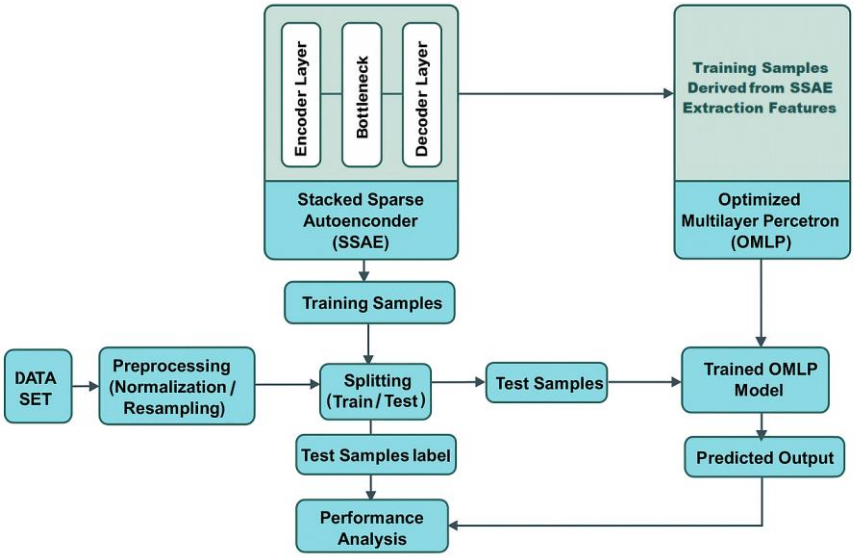


Figure 1. Proposed system architecture for fault diagnosis.

3.1. PV System

The PV system used for data collection, as described by Lazzaretti et al. (2020) [45], consists of two strings of eight C6SU-330P PV modules, labeled PV_1 through PV_{16}, connected to a 5 kW grid-tied inverter (NHS Solar 5K-GDM1). Each string is equipped with circuit breakers and fuses to ensure safe simulation of fault conditions. Figure 2 illustrates the system installation.

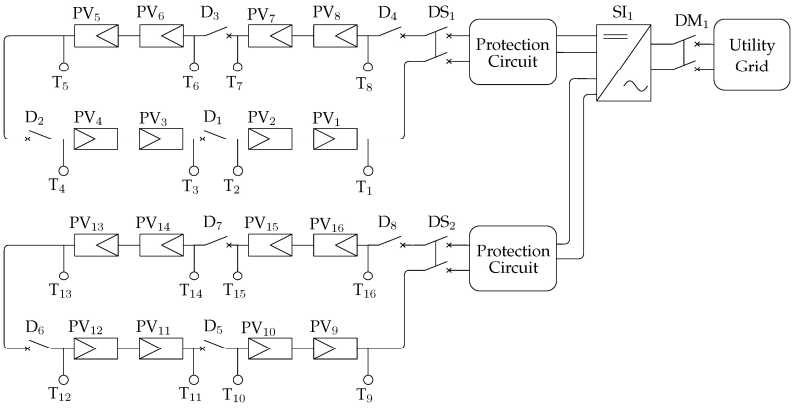


Figure 2. Schematic representation of the solar PV system.

3.2. Dataset Description

The dataset, sourced from a real PV system, was sampled at 1 Hz over a continuous 16-day period from 07:30 to 17:00, capturing irradiance variations from 100 to 1000 W/m^2 and temperature ranges from −5 °C to +85 °C, (cloudy and rainy days are not considered). It contains electrical parameters representing each fault condition. Key features include:

- Vdc_1, Vdc_2: Voltage for each string.
- Idc_1, Idc_2: Current for each string.
- *Irr*: Irradiance.
- *pvt:* PV Module Temperature.
- f_{nv}: Fault label.

To assess the PV system's performance under realistic conditions, the simulation framework integrates photovoltaic strings—each comprising eight 330 W modules—modeled in MATLAB 2022b/Simulink 10.6, with environmental variables such as irradiance and

temperature. The simulated DC voltage and current outputs from these strings, shown in Figures 3 and 4, feed into subsequent system components. Dataset parameters are aligned with the reference system configuration detailed in Figure 2, ensuring consistency between the simulation and real-world setups. Building on the validated PV model from earlier sections, comprehensive databases are generated to capture both optimal operation and intentional fault scenarios. These datasets leverage daily solar irradiance and temperature profiles, as illustrated in Figure 5, to replicate outdoor conditions.

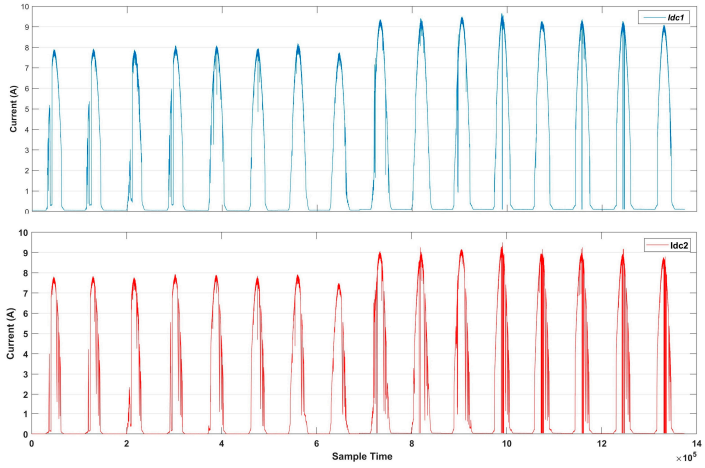


Figure 3. DC output current of each string.

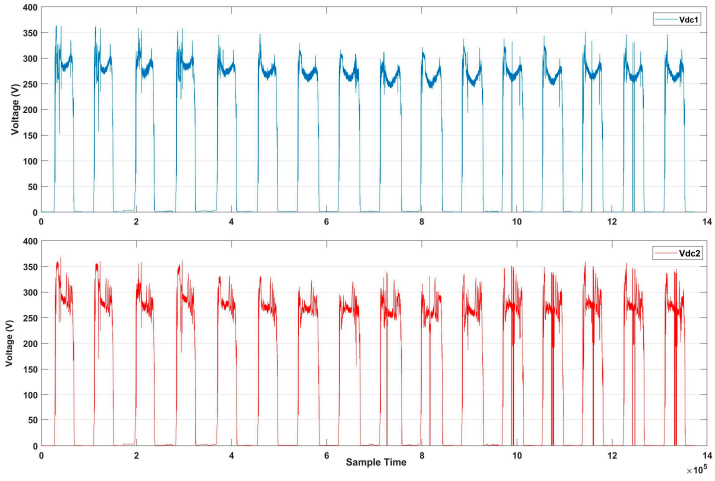


Figure 4. DC output voltage of each string.

The dataset is divided into two distinct parts: one derived from simulated fault scenarios and the other from real-world measurements. The training phase relied exclusively on simulated data generated using a validated PV system model in MATLAB/Simulink. Fault types such as short circuit, open circuit, degradation, and shadowing were modeled under varying irradiance and temperature conditions to reflect realistic operational behavior. For instance, in the case of shadowing faults, four distinct scenarios were simulated, each introducing 5% to 15% shading across a PV string, resulting in a total of 5054 synthetic samples used in training.

In contrast, the testing phase was conducted entirely on real-world data collected from an actual PV installation, which comprises 514,958 labeled samples, with the class distribution detailed in Table 1. To evaluate model generalization across different fault categories under realistic conditions, the testing procedure was repeated three times using different subsets of the real dataset.

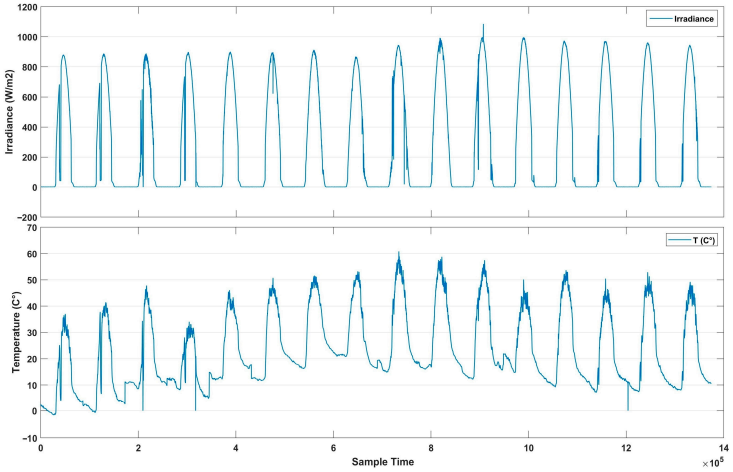


Figure 5. Daily variations of solar irradiance and temperature over time.

Table 1. Proportion of data and fault label.

Proportion of Data		
Label	**Class**	**Samples**
0	Normal Operation	309,253
1	Short Circuit	5999
2	Open Circuit	6024
3	Degradation	10,371
4	Shadowing	184,311

Environmental variables such as temperature and irradiation are also included to account for real-world operating conditions. The data was normalized to a (0, 1) range and split 70-30 into training and testing sets. We specifically selected six photovoltaic features denoted as x = [$x1$, $x2$, $x3$, $x4$, $x5$, $x6$] representing [$Vdc1$, $Vdc2$, $Idc1$, $Idc2$, Irr, pvt], respectively, to be used as input for training our model, providing data points that distinguish between healthy and faulty operation. Table 1 presents the data samples of each category.

3.3. Data Preprocessing

The data was processed through a preprocessing pipeline to make it suitable for machine learning models, including data cleaning to handle potential anomalies, feature selection for critical parameters (voltage, current, temperature, and irradiation), and normalization to ensure consistency across features. Normalization was performed using min-max scaling, as outlined in Equation (1), where x_{min} and x_{max} were computed from the training set to avoid test data leakage and standardize the range of feature values [46].

$$x_{Scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Dimensionality reduction, known as feature extraction, involves condensing the number of variables or attributes in a dataset. Its purpose is to identify and retain the most meaningful and pertinent data embedded within the original features, transforming it into a smaller set of features while striving to preserve as much critical information as possible.

These steps are essential to improve data quality, enhance model performance, and ensure the reliability and accuracy of the resulting insights. By addressing issues such as missing values, inconsistencies, and irrelevant data, preprocessing ensures that the machine

learning model receives clean, well-structured, and meaningful input, leading to more accurate and reliable predictions [47].

3.4. Stacked Sparse Auto-Encoder for Fault Extraction

The SSAE is a deep learning model specifically designed for unsupervised feature extraction. In this architecture, multiple auto-encoders are sequentially stacked, allowing the model to capture complex, high-dimensional fault features relevant to PV diagnostics.

- Architecture: Each auto-encoder in the SAE comprises an encoder, a hidden layer (bottleneck), and a decoder. The encoder transforms the normalized data into a compressed representation in the bottleneck layer, where critical patterns relevant to fault types are retained. This architecture allows the SSAE to automatically learn and preserve the most informative features for fault detection [48].
- Encoding and decoding: The encoder maps the input vector X_i = [V_{dc1}, V_{dc2}, I_{dc1}, I_{dc2}, *Irr*, *pvt*] to a hidden representation hi to predict output $\hat{X}i = \hat{V}_{\mathrm{dc1}}\hat{V}_{\mathrm{dc2}}\hat{I}_{\mathrm{dc1}}\hat{I}_{\mathrm{dc2}}\hat{I}_{\mathrm{rr}}, \hat{F}_{\mathrm{nv}}$], as determined by (2) and (3):
  - Input Transformation (Encoding Layer)

$$h_i = f\ (x_i) = sigm(W_1.x + b_1) \tag{2}$$

During the encoding process, the input vector x_i is mapped to a latent representation h_i. This transformation is achieved through a linear operation involving weight matrix W_1 and bias b_1, followed by a nonlinear sigmoid activation function.

- Decoding Layer:

$$\hat{x}_i = g(h_i) = sigm(W_2\ .h + b_2) \tag{3}$$

The latent representation h_i is then reconstructed into $\hat{x}_i$ via the decoding function g which applies another linear transformation (using weights W_1 and bias b_1) and passes the result through a sigmoid activation, as shown in Equation (3).

  - Cost Function:

Equation (4) represents the cost function of the Sparse Autoencoder. It consists of three main components that will be further decomposed in Equations (10)–(12), where the first term quantifies the reconstruction loss (mean squared error between the input x_i and reconstructed output $\hat{x}_i$), ensuring accurate feature representation, and the second term is the weight decay regularization that penalizes large weights to prevent overfitting, and a sparsity constraint using *Kullback–Leibler* (*KL*) divergence.

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}\| \hat{x}_i - x_i \|^2\right) + \frac{\lambda}{2}\sum_{i=1}^{N}\| W_i \|^2 + \beta\sum_{i=1}^{N} KL(\rho \parallel \hat{\rho}_j) \tag{4}$$

where m represents the number of neurons in the hidden layer. The hyperparameters λ and β control the relative influence of the weight decay (L2 regularization) and the sparsity penalty, respectively. The sparsity constraint utilizes the *KL* divergence, denoted as $KL(\rho \parallel \hat{\rho}_j)$ [49], to quantify the discrepancy between the target sparsity parameter ρ and the empirical average activation of the i-th hidden neuron over the training dataset.

3.5. Multi-Layer Perceptron for Fault Classification

Following feature extraction by the SSAE, an Optimized Multi-Layer Perceptron is employed for fault classification. The MLP is a supervised neural network model that excels at distinguishing between different classes (fault types) by learning complex feature mappings [50].

Architecture: The MLP consists of an input layer, one or more hidden layers, and an output layer as shown in Figure 6. Each layer is fully connected to the subsequent layer. The input layer receives SSAE-extracted features, and each hidden layer transforms these features via learned weights and biases [35].

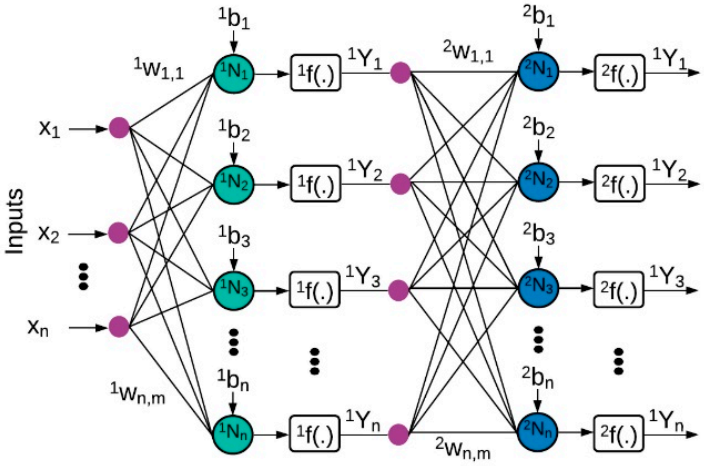


Figure 6. The architecture of an MLP.

Forward propagation: The output of each hidden layer is computed as determined by (5) and (6).

$$^{1}Y = {}_{.}^{1}f\left({}_{.}^{1}\omega \times X + {}_{.}^{1}b\right) \tag{5}$$

$$^{2}Y = {}_{.}^{2}f\left({}_{.}^{2}\omega \times {}^{1}Y + {}_{.}^{2}b\right) \tag{6}$$

where W represents the weight matrices, b represents biases, and f denotes the *ReLU* activation function, which introduces non-linearity to the model, which is expressed by (7) [38].

$$f(.) = \max(0, x_i) \tag{7}$$

Hyperparameter tuning: Optimal parameters were selected via grid search, adjusting learning rate, batch size, and hidden layer dimensions to improve accuracy and efficiency [50].

The combined SSAE and Optimized-MLP model form a powerful framework for fault detection and classification: the SSAE captures intricate fault-related features, while the MLP effectively translates these features into actionable insights. This approach demonstrates high performance, as validated across multiple key metrics.

3.6. Hybrid SSAE with Optimized-MLP Based Models

From prior research, hybrid architectures combining unsupervised and supervised learning have proven effective in scenarios with limited labeled data, particularly in fault diagnosis and anomaly detection [51]. AEs are widely recognized for their ability to learn hierarchical features from unlabeled data through unsupervised pre-training, reducing dependency on costly labeled datasets [27]. Meanwhile, MLPs excel at mapping learned features to target labels via supervised fine-tuning, making them ideal for classification and regression tasks [35]. By integrating SSAE and Optimized-MLP, this hybrid model leverages the strengths of both paradigms: the SSAE extracts sparse, low-dimensional representations of raw input data, while the Optimized-MLP utilizes these features for precise fault prediction. This approach is particularly advantageous for PV fault detection, where labeled fault samples are often scarce, but unlabeled operational data is abundant.

As illustrated in Figure 1, the hybrid model operates in two stages. First, raw input data X undergoes min-max normalization (Equation (1)) to scale values between [0, 1], ensuring stable training. The preprocessed data is then fed to the SSAE, which consists of an encoder f_{ENC} and decoder f_{DEC}. The encoder maps X to a latent representation Z using weights W_{ENC} and a sigmoid activation (Equation (8)), while the decoder reconstructs $\hat{X}$ from Z (Equation (9)). The model was trained over 100 epochs, with the following hyperparameters set: weight-decay λ = 0.001, sparsity penalty $\beta = 3$, and target activation $\rho = 0.05$. The Sparse AE's loss function (Equation (13)) combines reconstruction error, sparsity penalty (KL divergence), and L2 regularization. Once trained, the latent features Z are extracted and passed to the OMLP (Figure 3), which computes predictions $\hat{y}$ via fully connected layers.

For classification, cross-entropy loss (Equation (16)) is minimized, whereas regression tasks use mean squared error. The Adam optimizer updates weights in both stages, ensuring efficient convergence. As detailed in Algorithm 1 (Lines 6–26). The Optimized-MLP (Lines 30–50) processes the flattened features with 16 hidden nodes, culminating in a 5-node output layer that classifies faults into normal, degradation, SC, OC, and partial shading categories. The subsequent section evaluates this architecture's performance on real-world PV fault datasets, demonstrating its superiority over standalone models.

Algorithm 1. Hybrid SSAE and Optimized-MLP

1: **Procedure** Hybrid Stacked Sparse AutoEncoder and Optimized-Multilayer Perceptron ($\mathbf{X}, \mathbf{y}, \boldsymbol{\lambda} = \mathbf{0.001}, \boldsymbol{\beta} = \mathbf{3}, \boldsymbol{\rho} = \mathbf{0.05}, \boldsymbol{epochs} = \mathbf{100}$)
2: //**Preprocessing**
3: Input X: unlabeled raw data $X \in \mathbb{R}$ > ($Vdc1$, $Vdc2$, $Idc1$, $Idc2$, Irr, pvt)
4: Normalize X to [0, 1]
5: **//Train Stacked Sparse Autoencoder (SSAE)**
6: Initialize encoder f_{ENC} and decoder f_{DEC} with weights W_{ENC}, W_{DEC}
7: **//Train SSAE for 100 epochs**
8: **for epoch = 1 to 100 do**
9: //Forward pass
10: Latent representation:
11:

$$\boldsymbol{Z} \leftarrow \sigma(\boldsymbol{W_{ENC}}.\boldsymbol{X_{Norm}} + \boldsymbol{b_{ENC}}) \tag{8}$$

12: Reconstruction:
13:

$$\hat{\boldsymbol{X}} \leftarrow \sigma(\boldsymbol{W_{DEC}}.\boldsymbol{Z} + \boldsymbol{b_{DEC}}) \tag{9}$$

14: //Compute loss
15: Reconstruction error:
16:

$$\mathcal{L}_{\text{recon}} \leftarrow \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \| \hat{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x}_{\boldsymbol{i}} \|^{\mathbf{2}} \tag{10}$$

Algorithm 1. Hybrid SSAE and Optimized-MLP

17: Sparsity penalty (KL divergence):

18:

$$\mathcal{L}_{\text{sparse}} \leftarrow \beta \sum_{j=1}^{d} \left[\rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j} \right] \tag{11}$$

19: Weight decay:

20:

$$\mathcal{L}_{\text{reg}} \leftarrow \frac{\lambda}{2} \left(\|W_{\text{ENC}}\|_F^2 + \|W_{\text{DEC}}\|_F^2 \right) \tag{12}$$

21: Total loss:

22:

$$\mathcal{L}_{\text{SAE}} \leftarrow \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{reg}} + \mathcal{L}_{\text{sparse}} \tag{13}$$

23: //Backpropagation

24: Update $W_{ENC}, b_{ENC}, W_{DEC}, b_{DEC}$ via Adam:

25:

$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{\text{SAE}}, \theta \in \{W_{\text{ENC}}, b_{\text{ENC}}, W_{\text{DEC}}, b_{\text{DEC}}\} \tag{14}$$

26: **End for**

27: **//Feature Extraction**

28: Extract latent features:

29: $\mathbf{Z} \leftarrow f_{\text{ENC}}(X_{Norm})$

30:**//Optimized-MLP for Fault Classification**

31://Optimized-MLP for Fault Classification

32: Hidden layers $\in \{1, 2, 3\}$ Units/layer $\in \{32, 64, 128\}$, Dropout $\in [0.1, 0.5]$

33: Initialize MLP f_{MLP} with weights W_{MLP}

34: Initialize Bayesian optimizer.

35: for $t = 1 \rightarrow T$ do

36: Sample hyperparameters $\theta_t = \left\{ n_{\text{layers}}, n_{\text{units}}, p_{\text{dropout}} \right\}$

37: **for epoch = 1 to 100 do //Train MLP for 100 epochs**

38://Forward pass

39: Prediction:

40:

$$\hat{\mathbf{y}} \leftarrow f_{\text{MLP}}(\mathbf{Z}) \leftarrow \text{Softmax}\left(W_{\text{out}}\, \mathbf{h}_{n_{\text{layers}}} + \mathbf{b}_{\text{out}} \right) \tag{15}$$

41://Compute loss

42: Cross-entropy (classification) or MSE (regression):

43:

$$\mathcal{L}_{\text{MLP}} \leftarrow \begin{cases} -\frac{1}{n} \sum_{i=1}^{n} \sum_{c=1}^{C} y_{ic} \log\left(\hat{y}_{ic} \right) & \text{(classification)} \\ \frac{1}{n} \sum_{i=1}^{n} \left\| y_i - \hat{y}_i \right\|^2 & \text{(regression)} \end{cases} \tag{16}$$

44://Backpropagation

45: Update W_{MLP} via Adam:

Algorithm 1. Hybrid SSAE and Optimized-MLP

46:

$$W_{\mathbf{MLP}} \leftarrow W_{\mathbf{MLP}} - \eta \nabla_{W_{\mathbf{MLP}}} \mathcal{L}_{\mathbf{MLP}} \tag{17}$$

47: Select optimal $\theta^* \leftarrow \mathbf{argmin} \mathcal{L}_{\mathbf{MLP}}$
48: **End For**
49: return $f_{\mathbf{ENC}}, f_{\mathbf{MLP}}$
50: End procedure

3.7. Evaluation Metrics

To evaluate the outcomes of the fault detection process in this study, a range of metrics were employed. These metrics play an essential role in assessing the effectiveness and reliability of the detection process. Prior to outlining performance metrics for this procedure, it is crucial to clarify the following concepts [52]:

- TP (True Positive): This occurs when the detection process correctly identifies an authentic fault in the PV system.
- TN (True Negative): This is observed when the PV system operates without any issues, and the fault detection system correctly verifies the absence of faults.
- FP (False Positive): This occurs when the PV system shows no faults, and the fault detection system identifies a fault.
- FN (False Negative): This happens when the PV system experiences a fault, and the detection system does not indicate it.
- Accuracy: Correlates with the comprehensive detection effectiveness, Equation (18):

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{18}$$

- Precision: Represents the ratio between positive indicators, Equation (19):

$$Precision = \frac{TP}{(TP + FP\ )} \tag{19}$$

- Sensitivity: Assesses the effectiveness of accurately classifying detections, Equation (20):

$$Sensitivity = \frac{TP}{(TN + FP\ )} \tag{20}$$

- Specificity (recall): Assesses the classifier's efficiency in recognizing inaccurate detections, Equation (21):

$$Specificity = \frac{TN}{(TN + FP\ )} \tag{21}$$

- Macro F1-Score: Harmonic mean of precision and recall, averaged equally across all classes, Equation (22):

$$Macro\ F1\ score = \frac{1}{C} \sum_{c=1}^{C} \frac{Precision_c . Recall_c}{Precision_c + Recall_c} \tag{22}$$

where C is the number of classes.

- Micro F1-Score: Global harmonic mean of overall precision and recall, accounting for class imbalance, Equation (23):

$$Micro\ F1\ score = \frac{2.\ TP_{Total}}{2.TP_{Total} + FP_{Total} + FN_{Total}} \tag{23}$$

- Cohen's Kappa (κ): Measures inter-rater agreement adjusted for chance, Equation (24).

$$\kappa = \frac{P_0 - P_e}{1 + P_e} \tag{24}$$

where P_0 is observed agreement and P_e is expected agreement.

4. Results

Experiments were conducted using Python 3.12.7 and TensorFlow 2.18.0 on a high-performance computing environment. The SSAE was trained first, followed by the OMLP classifier, until both models achieved convergence based on early stopping criteria.

4.1. SSAE Performance

Figure 7 displays the training and validation loss curves for the SSAE. Both losses decrease rapidly in the first few epochs, suggesting that the SSAE quickly learns efficient feature representations. The subsequent stability of both curves at low values indicates that the SSAE effectively captures the underlying data patterns with minimal reconstruction error, achieving a good balance between training and validation. Minor fluctuations in the validation loss suggest slight variability but no overfitting, showing that the SSAE model generalizes well across both sets.

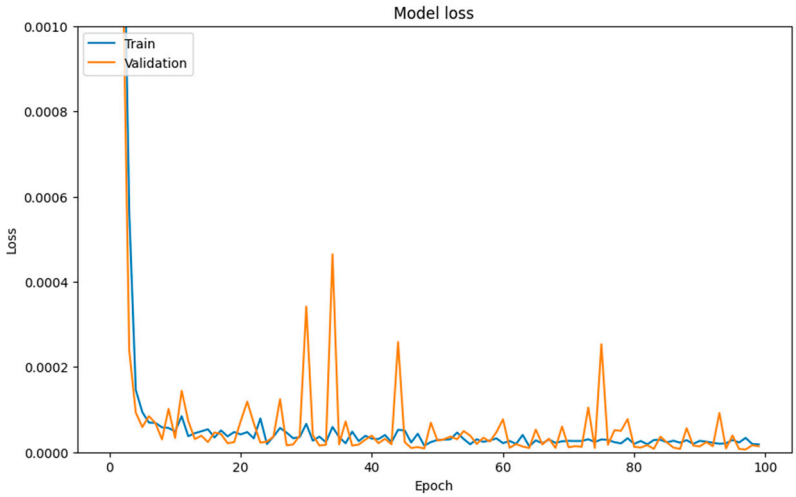


Figure 7. Optimal fitting model.

4.2. Optimized-MLP Performance

Performance metrics such as accuracy, precision, sensitivity, and specificity (recall) are calculated to assess the model's effectiveness in identifying fault types.

The Optimized-MLP model accuracy in Figure 8 shows both training and validation accuracy over 100 epochs. Both accuracies increase rapidly in the initial epochs, stabilizing near 99.9% accuracy. The high and consistent accuracy levels indicate that the model has effectively learned the features extracted by the SSAE and performs well on both the training and validation sets. Minor fluctuations in validation accuracy suggest slight variations in generalization but no significant overfitting, demonstrating that the OMLP model is well-optimized for accurate fault diagnosis.

The testing results in Table 2 highlight the high performance of our fault diagnosis approach. Three experimental trials were conducted by repeating the process with three

distinct data partitions derived from the same dataset, ensuring statistical robustness. The model shows a strong average precision (99.7%), indicating accurate fault identification with minimal false positives, and an average sensitivity (99.4%), ensuring effective fault detection. The specificity reaches 100% across all tests, confirming that non-fault conditions are correctly identified. Finally, with an average accuracy of 99.82%, the model demonstrates reliable and consistent overall performance. These results confirm that the SSAE and Optimized-MLP combination is highly effective for precise fault diagnosis.



Figure 8. Accuracy during training and validation.

Table 2. Testing Results.

	Test 1	Test 2	Test 3
Precision	0.996	0.998	0.998
Sensitivity	0.994	0.995	0.995
Specificity	1.000	1.000	1.000
Macro f1-score	0.973	0.975	0.978
Micro f1-score	0.996	0.997	0.998
Cohen's Kappa	0.984	0.986	0.988
Accuracy	0.996	0.997	0.9982

4.3. Confusion Matrix

The confusion matrix depicted in Figure 9 shows strong performance in fault classification. The model accurately identifies most "Normal" operating conditions, with 96,515 correct predictions, and demonstrates perfect precision (100%). For SC faults, 2173 samples were correctly classified, with minimal errors (2 misclassified as degradation and 7 as partial shading). Degradation faults also showed high accuracy, with 1917 correct identifications and minor overlaps (3 confused with partial shading and 7 with OC faults), underscoring the model's reliability in distinguishing critical fault types.

4.4. Comparative Study

In this section, we compare our fault detection approach with recent models used in PV system diagnostics, focusing on the DL techniques SAE and clustering, CNN, and CNN Bi-GRU. Table 3 provides a summary of fault types detected, algorithms used, and accuracy achieved by each method.

Our approach achieves a testing accuracy of 99.82%, outperforming other models significantly, with a 2.8% increase over the highest alternative (SSAE with clustering at 97%). In contrast, the standalone CNN achieves 95.20% accuracy for OC, SC, PS, and degradation faults, while the CNN Bi-GRU detects faults with an accuracy of 99.4%.

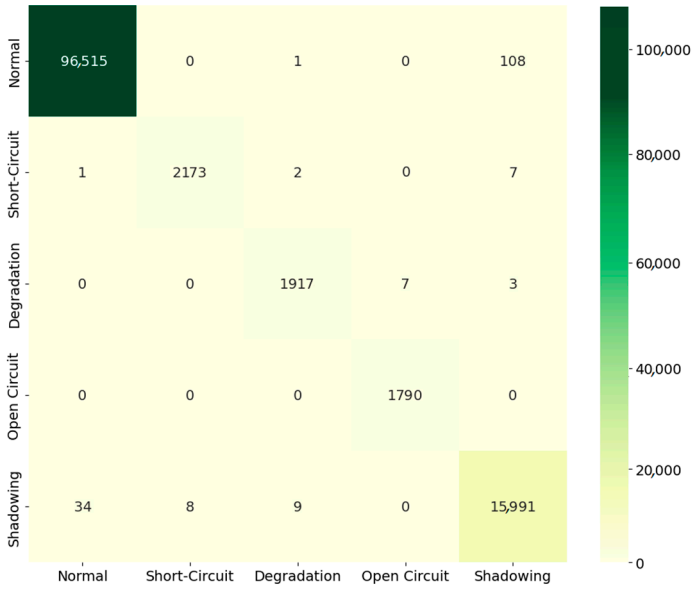


Figure 9. The confusion matrix.

Table 3. Results of the comparative models.

Classifier	Faults	Accuracy
SAE and clustering	OC, SC, PS, and degradation	97%
CNN	OC, SC, PS, and degradation	95.20%
CNN Bi-GRU	OC, SC, PS, and degradation	99.4%
Our method	OC, SC, PS, and degradation	99.82%

The superior performance of our approach can be attributed to the synergy between the SSAE and Optimized-MLP. The SSAE component effectively captures complex, high-dimensional patterns within the data through unsupervised feature extraction, enhancing fault feature representation. These refined features are then processed by the OMLP, which excels in classification accuracy due to its multi-layer structure and ability to generalize across fault types. This layered architecture allows our model to differentiate between fault conditions more precisely, even in challenging scenarios where fault types may have overlapping characteristics. This improvement highlights our method's potential to increase PV system reliability and efficiency, offering more precise fault detection, reducing false alarms, and ultimately supporting more cost-effective maintenance and enhanced system longevity.

5. Conclusions

In conclusion, the proposed hybrid SSAE-OMLP model demonstrates a robust framework for fault diagnosis in photovoltaic systems. By leveraging SSAE for feature extraction and Optimized-MLP for classification, our approach achieves high accuracy, precision, and specificity, contributing to the reliable operation of PV systems. While the model performs exceptionally in fault detection, future work could focus on testing across larger and more diverse PV datasets to ensure generalizability. Additionally, integrating real-time monitoring capabilities would enhance scalability in industrial applications. This study underscores the potential for advanced machine learning models in renewable energy fault diagnostics, supporting the broader goal of sustainable energy production and cost-effective system maintenance.

Author Contributions: Conceptualization, M.B. and M.D.; methodology, M.B.; software, M.B.; validation, M.B., S.B. and M.D.; formal analysis, M.B.; investigation, M.B.; data curation, M.B.; writing—original draft preparation, M.B.; writing—review and editing, M.D., S.B. and O.P.; visualization, M.B.; supervision, S.B.; project administration, M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the "PHC TASSILI +" program (project number: 24MDU117), funded by the French Ministry for Europe and Foreign Affairs, the French Ministry for Higher Education and Research, and the Algerian Ministry for Higher Education and Scientific Research (MESRS).

Data Availability Statement: The dataset is publicly available at https://github.com/clayton-h-costa/pv_fault_dataset (accessed on 17 March 2025).

Acknowledgments: The authors would like to thank "PHC TASSILI +" program (project number: 24MDU117), supported by the French Ministry for Europe and Foreign Affairs, the French Ministry for Higher Education and Research, and the Algerian Ministry for Higher Education and Scientific Research (MESRS).

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

References

1. Maka, A.O.M.; Alabid, J.M. Solar energy technology and its roles in sustainable development. *Clean Energy* **2022**, *6*, 476–483. [CrossRef]
2. Alami, A.H.; Alasad, S.; Aljaghoub, H.; Ayoub, M.; Alashkar, A.; Mdallal, A.; Hasan, R. First-Generation Photovoltaics: History and Conventional Manufacturing. In *Advances in Science, Technology & Innovation*; Springer International Publishing: Berlin/Heidelberg, Germany, 2023; pp. 7–19. [CrossRef]
3. Bougoffa, M.; Benmoussa, S.; Djeziri, M.; Contaret, T. Deep Learning-Based Fault Diagnosis in Photovoltaic System. In Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications (DASA), Manama, Bahrain, 11–12 December 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–5. [CrossRef]
4. Mansouri, M.; Trabelsi, M.; Nounou, H.; Nounou, M. Deep Learning-Based Fault Diagnosis of Photovoltaic Systems: A Comprehensive Review and Enhancement Prospects. *IEEE Access* **2021**, *9*, 126286–126306. [CrossRef]
5. Abubakar, A.; Almeida, C.F.M.; Gemignani, M. Review of Artificial Intelligence-Based Failure Detection and Diagnosis Methods for Solar Photovoltaic Systems. *Machines* **2021**, *9*, 328. [CrossRef]
6. Venkatakrishnan, G.R.; Rengaraj, R.; Tamilselvi, S.; Harshini, J.; Sahoo, A.; Saleel, C.A.; Abbas, M.; Cuce, E.; Jazlyn, C.; Shaik, S.; et al. Detection, location, and diagnosis of different faults in large solar PV system—A review. *Int. J. Low-Carbon Technol.* **2023**, *18*, 659–674. [CrossRef]
7. Firth, S.K.; Lomas, K.J.; Rees, S.J. A simple model of PV system performance and its use in fault detection. *Sol. Energy* **2021**, *84*, 624–635. [CrossRef]
8. Rico Espinosa, A.; Bressan, M.; Giraldo, L.F. Failure signature classification in solar photovoltaic plants using RGB images and convolutional neural networks. *Renew. Energy* **2020**, *162*, 249–256. [CrossRef]
9. Sarikh, S.; Raoufi, M.; Bennouna, A.; Ikken, B. Characteristic curve diagnosis based on fuzzy classification for a reliable photovoltaic fault monitoring. *Sustain. Energy Technol. Assess.* **2021**, *43*, 100958. [CrossRef]
10. Dida, M.; Boughali, S.; Bechki, D.; Bouguettaia, H. Output power loss of crystalline silicon photovoltaic modules due to dust accumulation in Saharan environment. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109787. [CrossRef]
11. Juaidi, A.; Muhammad, H.H.; Abdallah, R.; Abdalhaq, R.; Albatayneh, A.; Kawa, F. Experimental validation of dust impact on-grid connected PV system performance in Palestine: An energy nexus perspective. *Energy Nexus* **2022**, *6*, 100082. [CrossRef]
12. Osmani, K.; Haddad, A.; Jaber, H.; Lemenand, T.; Castanier, B.; Ramadan, M. Mitigating the effects of partial shading on PV system's performance through PV array reconfiguration: A review. *Therm. Sci. Eng. Prog.* **2022**, *31*, 101280. [CrossRef]
13. Osmani, K.; Ramadan, M.; Lemenand, T.; Castanier, B.; Haddad, A. Optimization of PV array tilt angle for minimum levelized cost of energy. *Comput. Electr. Eng.* **2021**, *96*, 107474. [CrossRef]
14. Sepúlveda-Oviedo, E.H.; Travé-Massuyès, L.; Subias, A.; Pavlov, M.; Alonso, C. Fault diagnosis of photovoltaic systems using artificial intelligence: A bibliometric approach. *Heliyon* **2023**, *9*, e21491. [CrossRef] [PubMed]
15. Lodhi, E.; Wang, F.-Y.; Xiong, G.; Zhu, L.; Tamir, T.S.; Rehman, W.U.; Khan, M.A. A Novel Deep Stack-Based Ensemble Learning Approach for Fault Detection and Classification in Photovoltaic Arrays. *Remote Sens.* **2023**, *15*, 1277. [CrossRef]

16. Yuan, Z.; Xiong, G.; Fu, X. Artificial Neural Network for Fault Diagnosis of Solar Photovoltaic Systems: A Survey. *Energies* **2022**, *15*, 8693. [CrossRef]
17. Wang, J.; Gao, D.; Zhu, S.; Wang, S.; Liu, H. Fault diagnosis method of photovoltaic array based on support vector machine. *Energy Sources Part A Recovery Util. Environ. Eff.* **2019**, *45*, 5380–5395. [CrossRef]
18. Bendary, A.F.; Abdelaziz, A.Y.; Ismail, M.M.; Mahmoud, K.; Lehtonen, M.; Darwish, M.M.F. Proposed ANFIS Based Approach for Fault Tracking, Detection, Clearing and Rearrangement for Photovoltaic System. *Sensors* **2021**, *21*, 2269. [CrossRef]
19. Amiri, A.F.; Oudira, H.; Chouder, A.; Kichou, S. Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier. *Energy Convers. Manag.* **2024**, *301*, 118076. [CrossRef]
20. Ghoneim, S.S.; Rashed, A.E.; Elkalashy, N.I. Fault Detection Algorithms for Achieving Service Continuity in Photovoltaic Farms. *Intell. Autom. Soft Comput.* **2021**, *29*, 467–479. [CrossRef]
21. Chen, Z.; Wu, L.; Cheng, S.; Lin, P.; Wu, Y.; Lin, W. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics. *Appl. Energy* **2017**, *204*, 912–931. [CrossRef]
22. Kara Mostefa Khelil, C.; Amrouche, B.; Kara, K.; Chouder, A. The impact of the ANN's choice on PV systems diagnosis quality. *Energy Convers. Manag.* **2021**, *240*, 114278. [CrossRef]
23. Masita, K.; Hasan, A.; Shongwe, T.; Hilal, H.A. Deep learning in defects detection of PV modules: A review. *Sol. Energy Adv.* **2025**, *5*, 100090. [CrossRef]
24. Neupane, D.; Bouadjenek, M.R.; Dazeley, R.; Aryal, S. Data-driven machinery fault diagnosis: A comprehensive review. *Neurocomputing* **2025**, *627*, 129588. [CrossRef]
25. Gao, W.; Wai, R.-J. A Novel Fault Identification Method for Photovoltaic Array via Convolutional Neural Network and Residual Gated Recurrent Unit. *IEEE Access* **2020**, *8*, 159493–159510. [CrossRef]
26. Bougoffa, M.; Benmoussa, S.; Djeziri, M.; Contaret, T. Stacked AutoEncoder based diagnosis applied on a Solar Photovoltaic System. *IFAC-PapersOnLine* **2024**, *58*, 384–389. [CrossRef]
27. Liu, Y.; Ding, K.; Zhang, J.; Li, Y.; Yang, Z.; Zheng, W.; Chen, X. Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with I–V curves. *Energy Convers. Manag.* **2021**, *245*, 114603. [CrossRef]
28. Eldeghady, G.S.; Kamal, H.A.; Hassan, M.A.M. Fault diagnosis for PV system using a deep learning optimized via PSO heuristic combination technique. *Electr. Eng.* **2023**, *105*, 2287–2301. [CrossRef]
29. Appiah, A.Y.; Zhang, X.; Ayawli, B.B.K.; Kyeremeh, F. Long Short-Term Memory Networks Based Automatic Feature Extraction for Photovoltaic Array Fault Diagnosis. *IEEE Access* **2019**, *7*, 30089–30101. [CrossRef]
30. Qian, J.; Song, Z.; Yao, Y.; Zhu, Z.; Zhang, X. A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes. *Chemom. Intell. Lab. Syst.* **2022**, *231*, 104711. [CrossRef]
31. Yang, Z.; Xu, B.; Luo, W.; Chen, F. Autoencoder-based representation learning and its application in intelligent fault diagnosis: A review. *Measurement* **2022**, *189*, 110460. [CrossRef]
32. Arifeen, M.; Petrovski, A.; Hasan, M.J.; Noman, K.; Navid, W.U.; Haruna, A. Graph-Variational Convolutional Autoencoder-Based Fault Detection and Diagnosis for Photovoltaic Arrays. *Machines* **2024**, *12*, 894. [CrossRef]
33. Guo, Z.; Zhuang, Z.; Tan, H.; Liu, Z.; Li, P.; Lin, Z.; Shang, W.-L.; Zhang, H.; Yan, J. Accurate and generalizable photovoltaic panel segmentation using deep learning for imbalanced datasets. *Renew. Energy* **2023**, *219*, 119471. [CrossRef]
34. Choudhary, T.; Mishra, V.; Goswami, A.; Sarangapani, J. A comprehensive survey on model compression and acceleration. *Artif. Intell. Rev.* **2020**, *53*, 5113–5155. [CrossRef]
35. Vieira, R.G.; Dhimish, M.; de Araújo, F.M.U.; da Silva Guerra, M.I. Comparing multilayer perceptron and probabilistic neural network for PV systems fault detection. *Expert Syst. Appl.* **2022**, *201*, 117248. [CrossRef]
36. Amiri, A.F.; Kichou, S.; Oudira, H.; Chouder, A.; Silvestre, S. Fault detection and diagnosis of a photovoltaic system based on deep learning using the combination of a convolutional neural network (CNN) and bidirectional gated recurrent unit (BI-GRU). *Sustainability* **2024**, *16*, 1012. [CrossRef]
37. Mouaad, B.; Samir, B.; Mohand, D. Dynamic Modeling for Fault Diagnosis in PV Systems Utilizing AI Techniques Based on Multilayer Perceptron (MLP). In Proceedings of the 2024 ASU In-ternational Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS), Manama, Bahrain, 28–29 January 2024; pp. 1032–1036. [CrossRef]
38. Heidari, A.A.; Faris, H.; Mirjalili, S.; Aljarah, I.; Mafarja, M. Ant Lion Optimizer: Theory, literature review, and application in multi-layer perceptron neural networks. In *Studies in Computational Intelligence*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 23–46. [CrossRef]
39. Madeti, S.R.; Singh, S.N. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Sol. Energy* **2017**, *158*, 161–185. [CrossRef]
40. Alam, M.K.; Khan, F.; Johnson, J.; Flicker, J. A Comprehensive Review of Catastrophic Faults in PV Arrays: Types, Detection, and Mitigation Techniques. *IEEE J. Photovolt.* **2015**, *5*, 982–997. [CrossRef]
41. Nadeem, A.; Sher, H.A.; Murtaza, A.F.; Ahmed, N. Online current-sensorless estimator for PV open circuit voltage and short circuit current. *Sol. Energy* **2021**, *213*, 198–210. [CrossRef]

42. Liu, H.; Mitolo, M.; Qiu, J. Ground-Fault Loop Impedance Calculations in Low-Voltage Single-Phase Systems. *IEEE Trans. Ind. Appl.* **2024**, *50*, 1331–1337. [CrossRef]
43. Eskandari, A.; Milimonfared, J.; Aghaei, M.; Reinders, A.H.M.E. Autonomous Monitoring of Line-to-Line Faults in Photovoltaic Systems by Feature Selection and Parameter Optimization of Support Vector Machine Using Genetic Algorithms. *Appl. Sci.* **2020**, *10*, 5527. [CrossRef]
44. Triki-Lahiani, A.; Bennani-Ben Abdelghani, A.; Slama-Belkhodja, I. Fault detection and monitoring systems for photovoltaic installations: A review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2680–2692. [CrossRef]
45. Lazzaretti, A.E.; Costa, C.H.D.; Rodrigues, M.P.; Yamada, G.D.; Lexinoski, G.; Moritz, G.L.; Oroski, E.; Goes, R.E.D.; Linhares, R.R.; Stadzisz, P.C.; et al. A Monitoring System for Online Fault Detection and Classification in Photovoltaic Plants. *Sensors* **2020**, *20*, 4688. [CrossRef] [PubMed]
46. Akanbi, O.A.; Amiri, I.S.; Fazeldehkordi, E. Feature Extraction. In *A Machine-Learning Approach to Phishing Detection and Defense*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 45–54. [CrossRef]
47. Sarhan, M.; Layeghy, S.; Moustafa, N.; Gallagher, M.; Portmann, M. Feature extraction for machine learning-based intrusion detection in IoT networks. *Digit. Commun. Netw.* **2024**, *10*, 205–216. [CrossRef]
48. Seghiour, A.; Abbas, H.A.; Chouder, A.; Rabhi, A. Deep learning method based on autoencoder neural network applied to faults detection and diagnosis of photovoltaic system. *Simul. Model. Pract. Theory* **2023**, *123*, 102704. [CrossRef]
49. Ng, A. Sparse Autoencoder. In *CS294A Lecture Notes*; Stanford University: Stanford, CA, USA, 2011; Volume 72, pp. 1–19.
50. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40. [CrossRef]
51. van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2019**, *109*, 373–440. [CrossRef]
52. Colliot, O. (Ed.) Machine Learning for Brain Disorders. In *Neuromethods*; Springer: New York, NY, USA, 2023. [CrossRef]

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