Article

Assessment of Machine and Deep Learning Approaches for Fault Diagnosis in Photovoltaic Systems Using Infrared Thermography

Sahbi Boubaker [1,*], Souad Kamel [1], Nejib Ghazouani [2] and Adel Mellit [3]

1 Department of Computer & Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 21959, Saudi Arabia
2 Department of Civil Engineering, College of Engineering, Northern Border University, Arar 73222, Saudi Arabia
3 The Abdus Salam International Centre for Theoretical Physics (ICTP), Str. Costiera, 34151 Trieste, Italy
* Correspondence: sboubaker@uj.edu.sa

Abstract: Nowadays, millions of photovoltaic (PV) plants are installed around the world. Given the widespread use of PV supply systems and in order to keep these PV plants safe and to avoid power losses, they should be carefully protected, and eventual faults should be detected, classified and isolated. In this paper, different machine learning (ML) and deep learning (DL) techniques were assessed for fault detection and diagnosis of PV modules. First, a dataset of infrared thermography images of normal and failure PV modules was collected. Second, two sub-datasets were built from the original one: The first sub-dataset contained normal and faulty IRT images, while the second one comprised only faulty IRT images. The first sub-dataset was used to develop fault detection models referred to as binary classification, for which an image was classified as representing a faulty PV panel or a normal one. The second one was used to design fault diagnosis models, referred to as multi-classification, where four classes (Fault1, Fault2, Fault3 and Fault4) were examined. The investigated faults were, respectively, failure bypass diode, shading effect, short-circuited PV module and soil accumulated on the PV module. To evaluate the efficiency of the investigated models, convolution matrix including precision, recall, F1-score and accuracy were used. The results showed that the methods based on deep learning exhibited better accuracy for both binary and multiclass classification while solving the fault detection and diagnosis problem in PV modules/arrays. In fact, deep learning techniques were found to be efficient for the detection and classification of different kinds of defects with good accuracy (98.71%). Through a comparative study, it was confirmed that the DL-based approaches have outperformed those based on ML-based algorithms.

Keywords: photovoltaic modules; fault detection and diagnosis; machine learning and deep learning; hybrid model; transfer learning

1. Introduction

According to the International Energy Agency [1], the total cumulative installed capacity of photovoltaic (PV) energy at the end of 2021 reached at least 942 GW worldwide. To keep these plants working effectively and reliably over time and for safety reasons, they should be carefully protected, monitored, and supervised [2]. Designing effective fault detection and diagnosis approaches is a crucial task that plays an important role in protecting PV plants from many risks such as fire. Effective fault detection approaches can help in detecting faults at an early stage and then avoiding power losses. While effective fault diagnosis helps operation and maintenance (O&M) to make quick interventions with appropriate decisions (clear fault, change PV modules, etc.). Such techniques could be developed and integrated with the monitoring system for real-time application. Based on

the logged PV power values, monitoring systems can detect whether plants are working correctly or not [3].

To date, various fault detection and diagnosis methods have been developed [4]. In this work, we will focus on an approach based on infrared thermography images (IRTs) for the classification of four major defects that may occur on PV modules. The investigated defects are, in turn: dust and sand accumulated on PV module surfaces, damaged bypass diodes, shunted PV modules, and partial shading effect.

The main idea developed in this study consists of solving a classification problem based on the collected IRT images. Thus, binary classification for fault detection and multiclass classification for fault diagnosis purpose are used. According to the literature, various methods have been developed. Recently, the application of machine learning (ML)- and deep learning (DL)-based algorithms for the diagnosis of PV modules based on IRT images has attracted the attention of many researchers. For example, the classification of normal RGB images of PV modules (PVM) using a combination of deep learning and machine learning algorithms was addressed in [5]. A set of lazy classifiers, including locally weighted learning (LWL), the K-star algorithm (KS), nearest neighbor (NN), and k-nearest neighbor (kNN), were adopted to perform the classification process on the chosen image features. With the kNN classifier, whose computation time was 0.04 s, a higher accuracy— reaching a maximum of 98.95%—was achieved when compared to the other classifiers.

In ref. [6], true-color images captured by unmanned aerial vehicles (UAVs) were dually processed by CNN-ML classifiers to detect PVM faults. The AlexNet, VGG16, GoogLeNet, and ResNet50 models were assessed and demonstrated standing feature extraction capabilities. According to the study, support vector machine (SVM) classifiers demonstrated best-in-class performance and, as a result, can be regarded as ready for real-time multiclass classification applications.

The authors in [7] conducted a comparative study to assess the anomaly detection performance of three ML models. Data collected from two working PV power plants in India allowed the determination of some correlations between internal and external plant feature parameters. The study concluded that the auto-encoder–long short-term memory (AE-LSTM) method was able to efficiently detect anomalies and healthy signals.

As an alternative to signal processing for PVM fault diagnosis, IRT-based fault classification was proposed in [8,9]. The authors' more recent work provides an automated fault detection methodology utilizing texture feature analysis and supervised machine learning. The suggested approach uses IRT to identify PVM anomalies and fuzzy edge detection to determine the orientation of anomalous modules. A gray level co-occurrence matrix was used to extract picture texture. The retrieved characteristics were labeled and trained using SVM to identify the PVM failure type.

Utilizing three segmentation models (FPN, U-Net, and DeepLabV3+), ref. [10] presented a deep learning-based photovoltaic defect detection method using UAV thermal imaging to identify defective solar panels on large-scale farms. Two datasets were merged to boost accuracy. Among the different segmentation models, the highest dice scores were obtained with the FPN, U-Net, and DeepLabv3+ networks. Achieving dice scores of 0.94 (U-Net), 0.87 (FPN), and 0.87 (DeepLabV3+), these tests were able to produce competitive semantic segmentation results. This research revealed that picture segmentation algorithms based on deep learning outperform conventional approaches.

An automatic UAV-based PVM inspection system [11] was tailored for large-scale farms using an offline scheme. For defect detection purposes, a seven-learned-layer CNN-based deep learning structure was adopted to detect visible PVM defects. An 8400-image dataset was used in the study to assess six typical defects. To make sure all defects were sufficiently represented in the dataset, images from three photovoltaic (PV) plants of PV modules with artificially produced defects were either taken manually or with UAVs. The proposed approach was validated and compared with classical pattern recognition approach and the pre-trained VGG-16 in ImageNet.

In [12], the authors suggested an automatic PVM fault detection study using IRT images collected experimentally in indoor and outdoor conditions. Isolated and develop-model transfer deep learning techniques were used. An isolated trained model based on a light CNN architecture achieved 98.67% of accuracy. Transfer learning begins with a foundational model pre-trained on the PVM electroluminescence (EL) image dataset, which was subsequently refined using the IR image datasets. The accuracy of the transfer-learning model was 99.23%. The times required to make a prediction for a single image using isolated and transfer learning networks were 13.12 and 13.07 ms, respectively.

In [13], a semi-supervised learning model (ppFDetector) was proposed as a novel method for detecting PV panel faults quickly and accurately, with only a small number of negative samples being required. A generative adversarial network (GAN) was used to reconstruct positive images using a centralized gradient. By comparing the original and reconstructed images (32,000 images reconstructed from a smaller dataset of images taken at a large-scale PV power plant located at Zhejiang, China), ppFDetector was used, along with the convolutional block attention module (CBAM) and SmoothL1 loss algorithms, to identify defective and healthy PVMs. The ppFDetector model outperformed all selected semi-supervised models, but demonstrated lower performance compared with a fully supervised pre-trained VGG-16 model with 200 negative images. The study did not report the computational cost required to achieve the model's performance.

The authors in [14] investigated an IRT image-based approach using an isolated convolutional neural model (ICNM) with a transfer learning approach to detect PVM faults of a small-sized solar PV system in Pakistan. The performance metrics of the proposed multi-class classification strategy were measured, including training loss, training accuracy, validation loss, validation accuracy, testing accuracy, and execution time. The results for PV health-based classification were provided for the ICNM, as well as publicly available pre-trained transfer learning networks (SqueezeNet, GoogleNet, and ShuffleNet). The seven-layered ICNM trained model achieved better performance than the aforementioned pre-trained networks with respect to execution time, training loss and training/validation accuracy.

Overheated PVM regions were investigated in [15] using IRT image analysis with a deep convolutional neural network (DeepCNN). To address the deep CNNs' limited effectiveness in the detection of overheated regions, the authors proposed a skip-connected DeepCNN architecture (SC-DeepCNN). A dataset was employed that consisted of 1428 IRT images captured at a 120 kWp PV solar farm in Jilin, China, of which 695 images included overheated region defects, including hotspots and overheated substrings and strings. IRT images with overheated regions were selected using SC-DeepCNN. A series of regions of interest (RoIs) for each raw IRT image was predicted using selective search (SS). Therefore, an RoI-DeepCNN model was developed and used along with joint learning for overheated region object detection. The results were presented and compared with other ML models, demonstrating that the method outperformed SC-DeepCNN and RoI-DeepCNN in terms of relevant performance metrics. Features and feature extraction techniques are reviewed and explained in [16,17].

In [18] a fine-tuned model based on Visual Geometry Group (VGG-16) and a DeepCNN were developed to classify five faults often observed in PV modules (short-circuit, open circuit, dust deposit, bypass diode failure, and partially covered PV module). The VGG-16 achieved an accuracy of 99.91% for fault detection and 99.80% for fault diagnosis for the five types of defect.

Other methods based on deep learning for fault detection and diagnosis can be found in [19–22].

The main objective of this study is to develop and assess various fault detection and diagnosis methods for PV modules by means of IR thermography images and deepCNN.

In the context of the literature review presented above, the most important contributions of the current study are:

- The performance of different state-of-the-art ML and DL methods is compared for the fault detection and diagnosis of PV modules.
- To the best of the authors' knowledge, no comprehensive comparative study has previously been performed on such applications that has included DL, ML, and hybrid approaches.
- The investigations are conducted using real-world IR thermography images of PV Si mono-crystalline technology.
- Four common faults in PV modules were examined: bypass diode failure, shading effect, short-circuited PV module, and soil accumulation on the PV module.

The rest of this paper is organized as follows: The materials and methods are described in Section 2, including the dataset used in this study, as well as the classification processes based on machine learning, deep learning, transfer learning, and hybrid approaches. Section 3 illustrates the results obtained, along with a discussion. A comparative study is reported in Section 4. Finally, conclusions and perspectives are presented in Section 5.

2. Materials and Methods

2.1. Research Methodology

To achieve these research objectives, a design methodology was designed composed of three steps (Figure 1). The main purpose of this study implementation is to test the feasibility of PV modules defects classification based on ML and DL techniques in order to subsequently extend the developed algorithms to a real case study in the Saudi Arabian context, in which PV images will be taken using drone-based equipment (Figure 2).
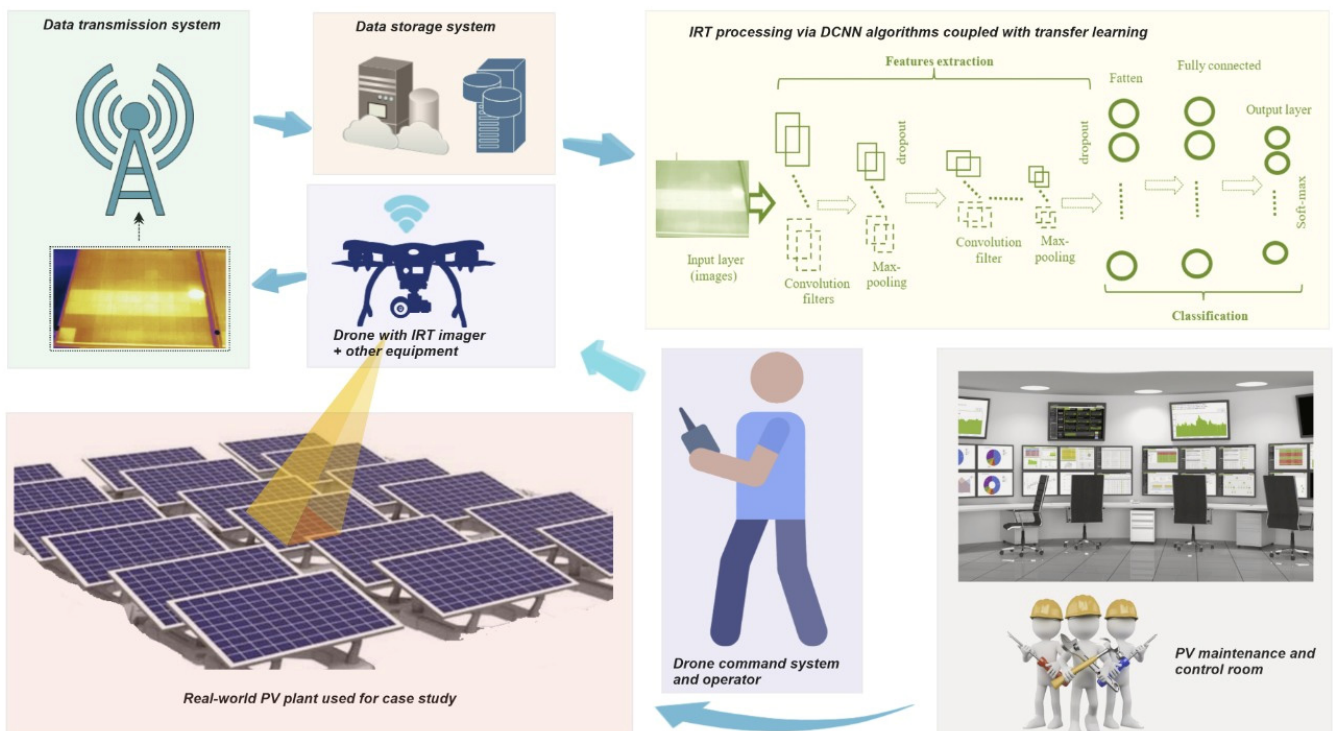

Figure 1. Block diagram of the proposed framework for the research.

In fact, this IRT-based PV fault detection investigation is the first part of a comprehensive case study targeting the first Saudi utility-scale solar energy project, the Sakaka photovoltaic power plant in the Al-Jouf region, northern Saudi Arabia. The power facility comprises 1.2 million PV panels over a surface of 6 square kilometers. It is expected to produce 300 MWp. After entering service, the farm will require maintenance and follow-up to ensure its full operation with the desired efficiency. Since the current maintenance approach depends on reading the electrical outputs at the inverter level, the proposed

research approach based on IRT and ML/DL represents a more effective and safer solution for solar panels, as it provides an opportunity for early detection of defects before they develop. Furthermore, the facility size challenge with respect to field inspection is easily overridden. Overall, this research consists of the following steps:

Figure 2. Aerial, lateral, and front views of the solar farm and solar panels, respectively.

2.1.1. Data Collection and Preparation

During this step, the research team, which specializes in IRT imaging, conducts an aerial survey of the solar panels in the study area using a drone equipped with an IR camera. The work team ensures the suitability of the weather and operational conditions at the photovoltaic plant before starting the onsite survey work in order to obtain high-quality IRT images. The first dataset is built by the structural health monitoring (SHM) expert, including the classification and labeling of IRT images into normal and faulty classes. Further preprocessing results in a second dataset, which in this project will take three or four classes into consideration (e.g., image with dust, image with sand, image with shading, etc.). It is worth noting that the identification of fault classes to build the IRT dataset is complex. This is made more difficult by the fact there is not an equal number of thermal images in each fault category.

2.1.2. Fault Detection Procedure Based on DeepCNNs

A DeepCNN classifier (binary classification) is developed for detection purposes. To this end, the IRT images are used to train and test the DeepCNN, following which new images taken by drone are used to test the DeepCNN classifier.

2.1.3. Fault Classification Procedure Based on DeepCNNs

An advanced type of deepCNN using a transfer learning approach is designed for fault classification (based on the collected images, the DeepCNN classifier will be able to classify the type of faults). This classifier can assist users in making informed decisions related to interventions in the PV installation, in cases where it is necessary to immediately clean, remove, or replace faulty panels (if fault hotspots persist undetected for a long time, they may cause a fire hazard).

The following flowchart (Figure 3) comprises two procedures for the use of DCNN-based processing of IRT datasets for fault detection and classification:

2.2. Databases

The databases used are summarized in Table 1. An augmentation technique was used to increase the size of these databases. A range of techniques are supported, as well as pixel scaling methods. We focused on five main types of data augmentation technique for image data, specifically:

- Image shifts via the width_shift_range and height_shift_range arguments.
- Image flips via the horizontal_flip and vertical_flip arguments.

- Image rotations via the rotation_range argument.
- Image brightness via the brightness_range argument.
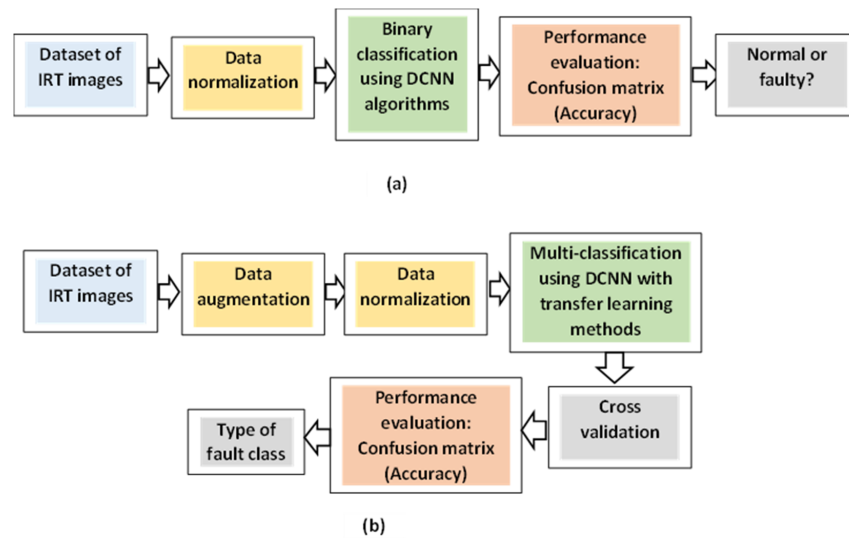- Image zoom via the zoom_range argument.

Figure 3. The proposed methodology for fault detection (**a**) and classification (**b**) using the IRT dataset.

Table 1. Databases.

Databases	Size	Classification			
Database #1	**4740**	**Binary**			
		Faulty		**Normal**	
Before augmentation		158		158	
After augmentation		2370		2370	
Database #2	**10,845**	**Multiclass**			
		F1	**F2**	**F3**	**F4**
Before augmentation		190	188	187	188
After augmentation		2850	2820	2805	2820

Each image from the datasets was generated 15 times employing the above augmentation techniques. The first database (database #1) consists of 4740 IR thermography images (each class comprises 2370 images). The second one (database #2) includes a total of 10,845 images (for example, F1 = 2850 images for the class F1 faulty PV panels). As an example, Figure 4 shows a sample of the collected images—(a) normal and (b) faulty. All investigated defects on PV modules (shunted, dust accumulated, shading effect and damaged bypass diode) are natural.
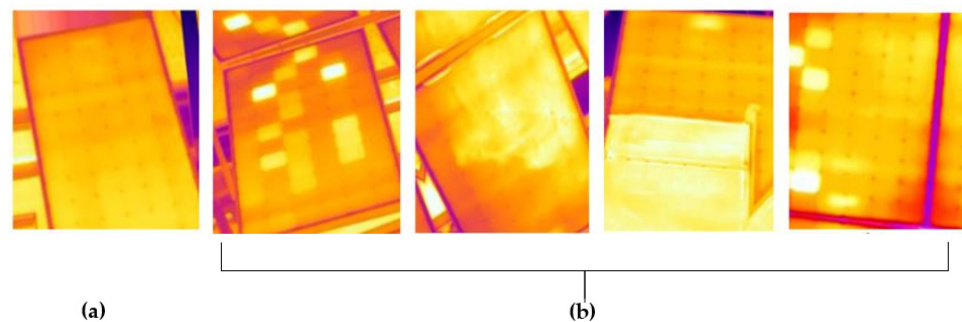

Figure 4. Sample of collected IR thermography images for PV modules: (**a**) healthy and (**b**) faulty (shunted, dust accumulation, shading effect, and damaged bypass diode).

As can be seen, hotspots can be clearly observed in each faulty PV image with variable profiles. The main problem faced by maintenance decision makers is the distinguish between those profiles and the efficient classification of defects such that precise decisions are possible (i.e., to determine whether to replace the PV module, clean the PV module, or remove the fault). Furthermore, if such phenomena persist for a long time without detection, they may lead to a fire hazard [2], resulting in damage to the PV module or even to part of the PV plant. To remove noise from the collected IRT images (image denoising), all images are then preprocessed using median filters (Figure 5).
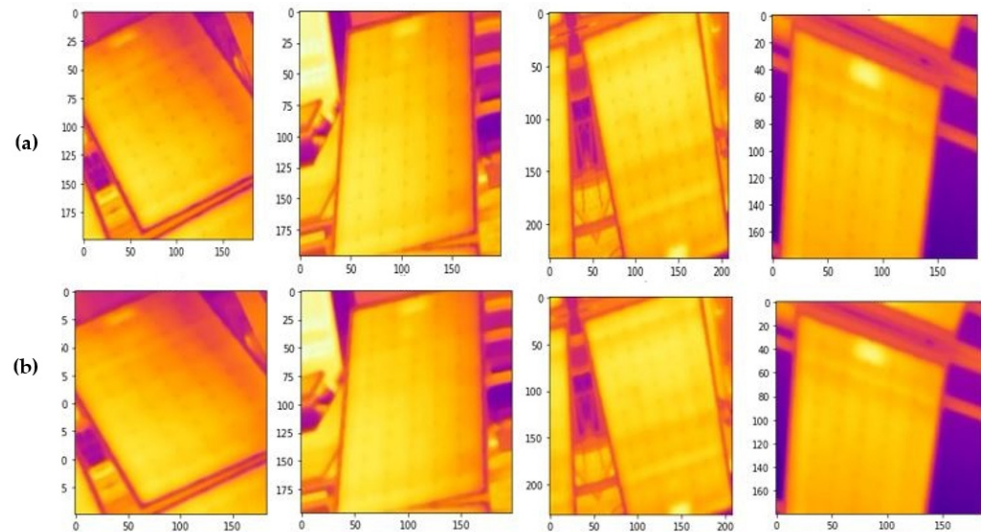

Figure 5. Case of healthy PV modules: (**a**) original IRT images; and (**b**) IRT images denoised by means of median filter.

2.3. ML-Based Classification

The preprocessing of images is an important step in developing an accurate classification model (Figure 6). ML algorithms require that all images in a dataset have a fixed feature vector size. This implies that the images should be preprocessed and scaled to have identical widths and heights. There are numerous ways of achieving this resizing and scaling, ranging from more advanced methods that respect the aspect ratio of the original image to the scaled image to simple methods that ignore the aspect ratio and simply squash the width and height to the required dimensions [16,17]. Three ML-based classifiers are discussed below: k-nearest neighbor (k-NN) [23]; support vector machine (SVM) [24]; and CatBoost [25].
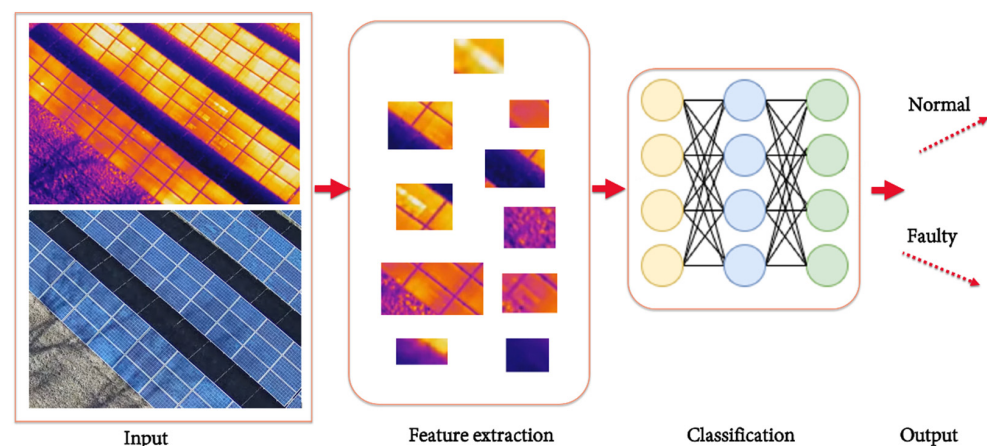

Figure 6. Workflow of ML-based classification.

2.3.1. k-NN

The k-nearest neighbor algorithm, commonly referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to classify or predict the grouping of a single data point. Ref. [23] looked into the combination of k-NN and the local manifold learning categorization of hyperspectral images. Local manifold learning was investigated from both supervised and unsupervised perspectives. Three variants were tested with AVIRIS and Hyperion hyperspectral data. The results showed that the hybrid k-NN performed better than the conventional k-NN. Additionally, the hybrid k-NN including supervised manifold learning outperformed that incorporating unsupervised manifold learning in terms of accuracy. The hybridization of k-NN and supervised manifold learning is interesting, since it depends only on the weights provided by the kernel function of the particular machine learning method, and does not require dimensionality reduction. Furthermore, this classifier should work for other classification problems like the one examined in this paper (normal and defective solar panels).

2.3.2. SVM

Support vector machine (SVM) is a group of supervised learning techniques for classifying data, performing regression analyses, and identifying outliers. SVM is built on the idea of a hyper-plane classifier. It seeks out the hyper-plane that optimizes the margin between two classes. SVM has been widely applied to many classification problems, among them the categorization of MR (Magnetic Resonance) images. Recently, there has been a significant increase in the use of magnetic resonance images in clinical studies due to their safety for the human body and the possibility of obtaining accurate information by scanning the same slice using different frequencies and settings. Ref. [24] used a support vector machine (SVM) to categorize the breast tissues in MR images. In order to confirm the viability and efficacy of this method, the authors in [24] carried out several experiments and compared the results with the widely utilized C-means (CM). In the scenario of SNR = 5 dB, the accuracy of SVM classification was noticeably higher than that of CM. As a result, it may be concluded that the SVM is noise-resistant and capable of segmenting multi-spectral MR images. Doctors can use the classification results of real MR images to achieve more accurate diagnosis and evaluation of patients' conditions.

2.3.3. CatBoost

CatBoost is a technique for decision trees that incorporates gradient boosting. It was created by Yandex researchers and engineers and is used in a variety of tasks, especially for remote sensing image classification. In remote sensing, there are three primary techniques for image classification: (1) classification under supervision; (2) classification without supervision; and (3) image analysis using objects. The third technique works well with high-resolution data. However, the two most common techniques for classifying images are supervised and unsupervised. Ref. [25] introduced and studied CatBoost for the classification of remote sensing (RS) images using a variety of attributes. CatBoost is a kind of ordered gradient boosting (OGB) that has been modified in a novel way to enable categorical features. A new pseudo-label feature extraction approach using multi-size neighboring patch-based multi-clustering has been proposed in order to enhance classification performance by supporting effective and efficient spatial feature extraction. When compared to many other gradient boosting and standard boosting implementations, CatBoost is better able to avoid the overfitting problem by achieving unbiased estimations of the gradient step. Experimental results on one PolSAR and two hyperspectral benchmarks showed that CatBoost demonstrated improved classification accuracy, but worse computing efficiency at higher numbers of boosting iterations. Parallelization performance and GPU-based acceleration of CatBoost could be attractive areas for future research.

2.4. DL-Based Classification

Deep learning (DL) enables the acquisition of data representations at various levels of abstraction by computational models made up of several processing layers (Figure 7).
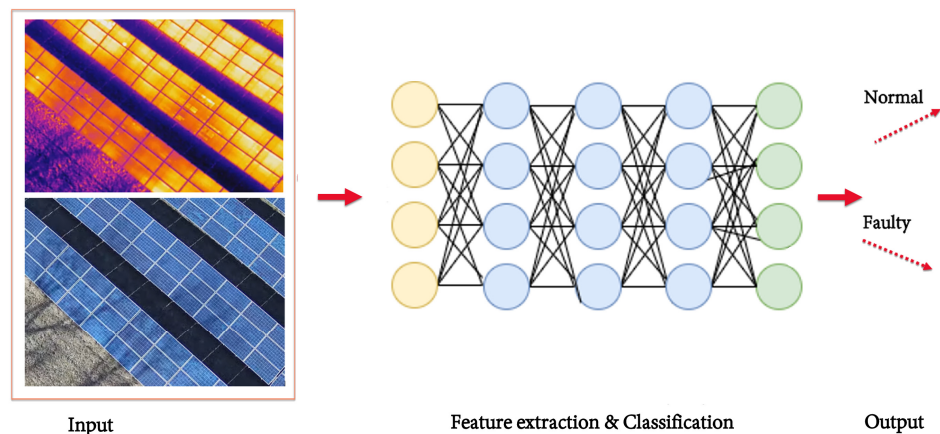

Figure 7. Workflow of DL-based classification.

The addition of deep learning techniques has enhanced the current state of the art in many fields, including drug discovery and genomics, object identification, visual object recognition, and speech recognition. By applying the backpropagation algorithm to determine how a machine should modify the internal parameters required to compute the representations in each layer on the basis of those in the previous layer, deep learning can uncover detailed structures in large data sets. Recurrent nets have shed light on sequential data types like text and speech, whereas deep convolutional nets have led to improvements in the processing of images, video, speech, and audio. Recently, deep learning approaches have shown tremendous success in image classification. As an example, two DL-based classifiers were selected: Visual Geometry Group from Oxford (VGG-16) [26] and Convolutional Neural Network (CNN) [27].

2.4.1. Classification Based on Transfer Learning Methods (VGG-16)

Recently, the idea of transfer learning has gained popularity and has been proven to be useful for classification problems. In transfer learning, convolutional layer weights from previously learned models are employed, whereby only the final layers require training on the basis of data from more recent classes. In ref. [26], the capacity of the well-known pre-trained CNN model VGG-16 with transfer learning was studied with respect to the automatic classification of MR brain images into normal and abnormal categories. The name VGG-16 refers to the group that developed it, while the number 16 refers to the fact that it is a deep neural network with 16 layers. To validate the pre-trained model with transfer learning, the dataset from the Harvard Medical School repository, containing both normal and pathological MR images of various neurological illnesses, was used. In comparison to other existing state-of-the-art studies, the validation using the test set with respect to accuracy, specificity, and sensitivity showed that the pre-trained VGG-16 model with transfer learning yielded the best performance. Additionally, the method offers end-to-end classification on unprocessed imagery without the need for manual attribute extraction. However, the drawback of this model is its extremely lengthy training period.

2.4.2. Classification Based on DeepCNN

Recently, convolutional neural networks (CNN), which represent a strong deep learning model, have been widely employed in the field of image processing, because they are effective at handling difficulties related to image recognition and classification and have achieved a significant increase in the accuracy of many ML tasks (Figure 8). CNN, which is a multilayer neural network, has demonstrated excellent performance in hyperspec-

tral image classification. CNN employs a feature detection layer to implicitly learn from training data in order to avoid explicitly extracting features. Ref. [27] suggested a novel training criterion for deep neural networks for the maximum interval minimal classification error based on an investigation of the error backpropagation technique. To achieve better results, the loss function and the cross-entropy obtained by gradient analysis are examined concurrently. Two deep learning benchmark databases, MNIST and CIFAR-10, were used for experiments. The proposed CNN was able to achieve effective results in both databases. CNN is incomparable in terms of representing and classifying image features. Nevertheless, CNN is still viewed as a black box model. More research is required to examine the relationship between the different layers of the deep convolutional neural network and the visual nervous system of the human brain, as well as how a deep neural network can be built incrementally, in the same manner that humans do.
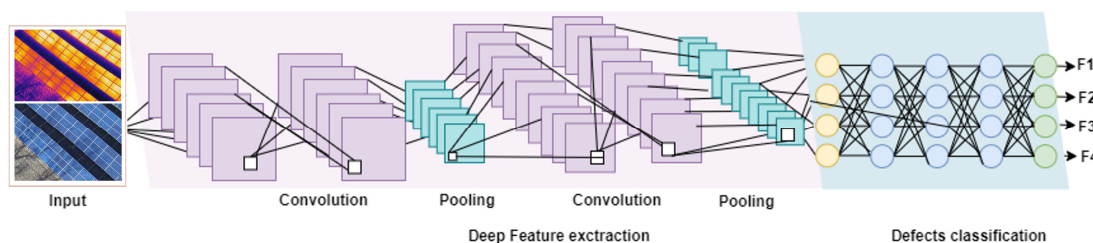

Figure 8. Workflow of CNN-based classification.

2.5. Classification Based on Hybrid Method (CNN-SVM)

Hybrid approaches combine two or more algorithms together in order to be able to take advantage of the features of each of the methods employed. Hybrid approaches have been investigated to address classification problems requiring tremendous computational effort. They have demonstrated superiority compared to traditional approaches. In order to benefit from their surprising ability to solve a wide spectrum of classification problems, we explored hybrid models combining CNN as an automatic feature extractor and SVM as a binary classifier. The CNN-SVM hybrid model proposed exhibited superior features, especially when tackling the categorization of remote sensing images (RSI), which is known to be a complex process characterized by poor pertinence, poor classification accuracy, and low computational efficiency. In [28], a seven-layer CNN structure was created where the ReLU function served as the activation function. Then, remote sensing images were entered into the CNN model, feature maps were retrieved, and the output layer of the CNN was substituted by training the feature maps in the SVM classifier. An empirical study of volcanic ash cloud classification from moderate-resolution image spectroradiometer (MODIS) RSI was performed, and the results assessed. The experimental results demonstrated that the proposed method, when compared to conventional methods, has a lower loss value and better generalization when performing modeling and/or training; the total classification accuracy and Kappa coefficient of the volcanic ash cloud reached 93.5% and 0.8502, respectively, thus indicating preferable volcanic ash cloud identification and visual effects. To increase classification accuracy and computational efficiency, future work is planned that will focus on how to extract more spectral features from various data sources, such as hyperspectral data and data with high spatial and temporal resolution.

2.6. Fault Detection and Diagnosis Procedure

To develop our models, we used the Keras and TensorFlow libraries. The following steps summarize the fault detection and diagnosis procedure for PV modules:

- Step #1: Upload the image and resize it.
- Step #2: Call the fault detection model (class = 0 means healthy, class = 1 means faulty).
- Step #3: Check whether class = 0. If so, this means the PV module is healthy, move to Step #5; otherwise move to Step #4.

- Step #4: Call the fault diagnosis model and identify the type of fault (Fault 1, Fault 2, Fault 3 or Fault 4).
- Step #5: Display the results and move to Step #1.

The procedure consists of two main functions: fault detection and fault diagnosis. Fault detection aims to detect anomalies in the PV modules, while the goal of the fault diagnosis function is to classify the examined faults. This procedure is based on examining the hotspot profile variations; each fault has a specific temperature profile. In order to develop the models, the databases were divided into two sets (80% for training and 20% for testing).

2.7. Evaluation Metrics

To verify the performance of the models, the confusion matrix is computed. The equations used to calculate accuracy, precision, recall, and F1-score are as follows [29]:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{3}$$

$$\text{F1-score} = 2\frac{(\text{Precision} * \text{Recall})}{(\text{Recall} + \text{Precision})} \tag{4}$$

where TP: number of true positives, TN: number of true negatives, FP: number of false positives, FN: number of false negatives.

3. Results and Discussion

The fault detection and diagnosis method developed based on Deep CNN configuration was evaluated using the confusion matrix. The codes were executed using Google Colab Pro with the following specifications: System RAM = 25 GB, GPU RAM T4, 16 GB, Hard Disk = 166 GB, and GPU clock = 585 MHz. More specifically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources, including GPUs. We selected an image size of 224 × 224 pixels due to the limited memory provided by Colab Pro (up to 25 GB) (Figure 9).
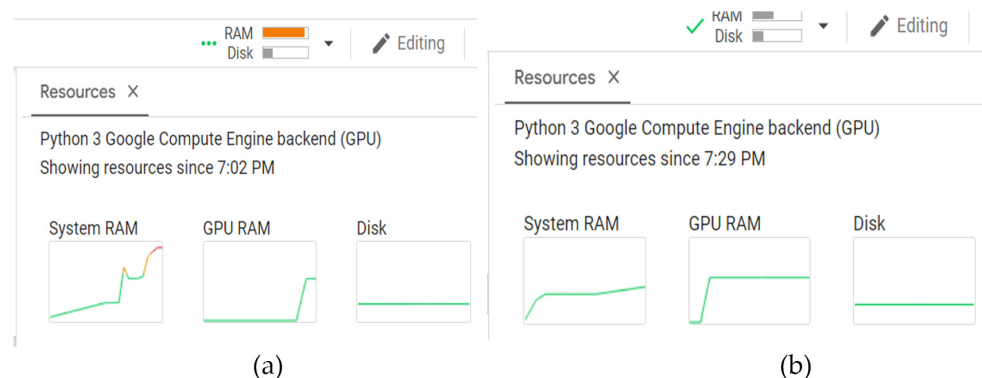

(a)   (b)

Figure 9. Preprocessing step execution using two types of image sizes: (**a**) with image size 480 × 480 pixels; and (**b**) with image size 224 × 224 pixels.

It can be seen from Figure 9a that the system RAM was fully used, and it was not possible to run our code correctly in the case of images with a size of 480 × 480, while in Figure 9b, it can be seen that in the case of images with a size of 224 × 224, there is enough memory to run the code. This also depends on the size of the dataset.

3.1. Fault Detection Based on Binary Classification

The configuration of the developed fault detection model (DeepCNN) consisted of 15 layers (Figure 10): the first nine layers (four conv2D layers, four MaxPooling layers and one dropout layer) were used for feature extraction; the second six layers (one flatten layer, four dense layers, and a final dense layer containing the Softmax function), with total of 56,210 trained parameters.
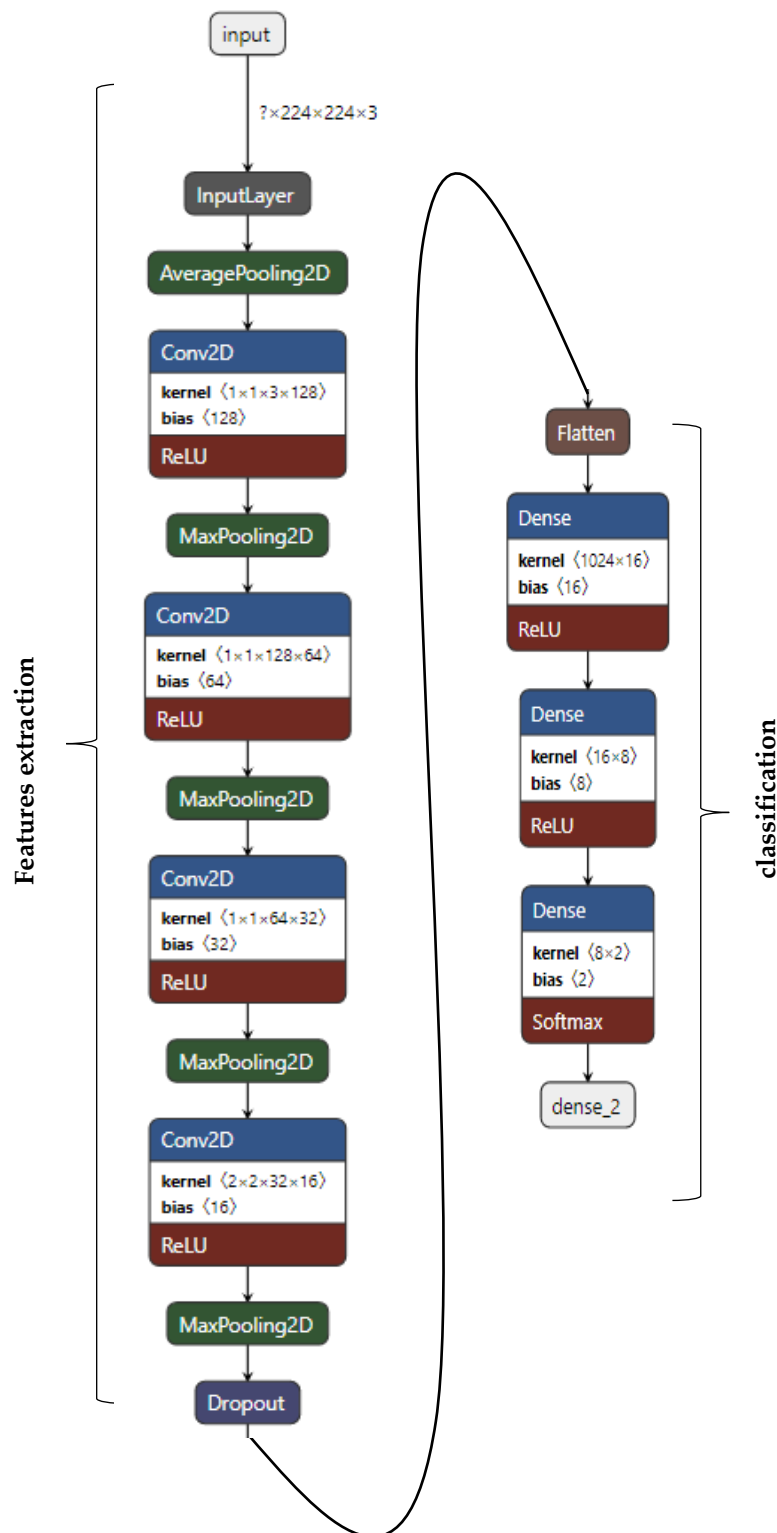

Figure 10. Configuration of the developed DeepCNN model for fault detection.

Figure 11 depicts the system memory, GPU memory, and disk capacity requirements. As can be seen, the fault detection code required about half (12 GB) of the available memory (25 GB). However, it consumed most of the available GPU RAM (16 GB). A disk with a capacity of 128 GB was large enough to support the size of database #1.
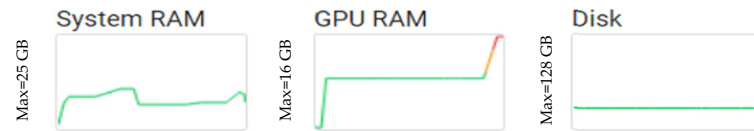

Figure 11. Assigned hardware resources (RAM and disk capacity) during the training process (multiclass classification).

Figure 12a shows the variation in training loss and accuracy functions. As can be seen, the loss reached 0.04 and the accuracy was very close to 1. Figure 12b displays the confusion matrix. It can be observed that three Faulty images were misclassified into the Normal class, and nine images belonging to the Normal class were misclassified into the Faulty class. The ROC curve shows that all of the examined classes were very close to 1.
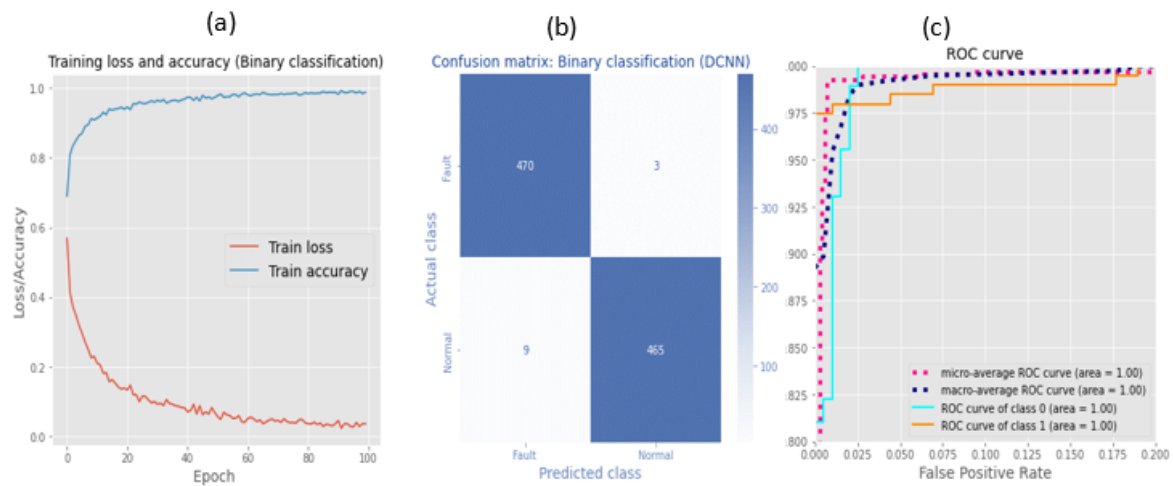

Figure 12. Binary classification (fault detection): (**a**) training loss and accuracy functions; (**b**) confusion matrix; and (**c**) ROC curve.

The calculated error metrics, including precision, recall, F1-score, and accuracy, during the testing process, are listed in Table 2, as well as the required running time during the training process.

Table 2. Calculated error metrics (precision, recall, F1-score and accuracy) and running time (during the training process).

Class	Precision	Recall	F1-Score	Support	Training Time (min)	Accuracy (%)
Faulty {1}	0.98	0.99	0.99	473		
Norma {2}	0.99	0.98	0.99	474	9	
Accuracy (%)	-	-	-	-		98.73
Macro average	0.99	0.99	0.99			
Weighted average	0.99	0.99	0.99			

From Table 2, it can be seen that the accuracy was 98.73% and the running time was about 9 min. Precision, recall and F1-score ranged between 98% and 99%. Therefore, it can be concluded that the deepCNN model developed for fault detection is able to detect anomalies with very good accuracy.

3.2. Fault Diagnosis Based on Multiclass Classification

The configuration of the developed fault diagnosis model consisted (deepCNN) of 17 layers (Figure 13): the first 11 layers (five conv2D layers, five MaxPooling layers, and one dropout layer) were used for feature extraction; the second six layers (one flatten layer, four dense layers, and a final dense layer containing the Softmax function), with total of 110,196 trained parameters.
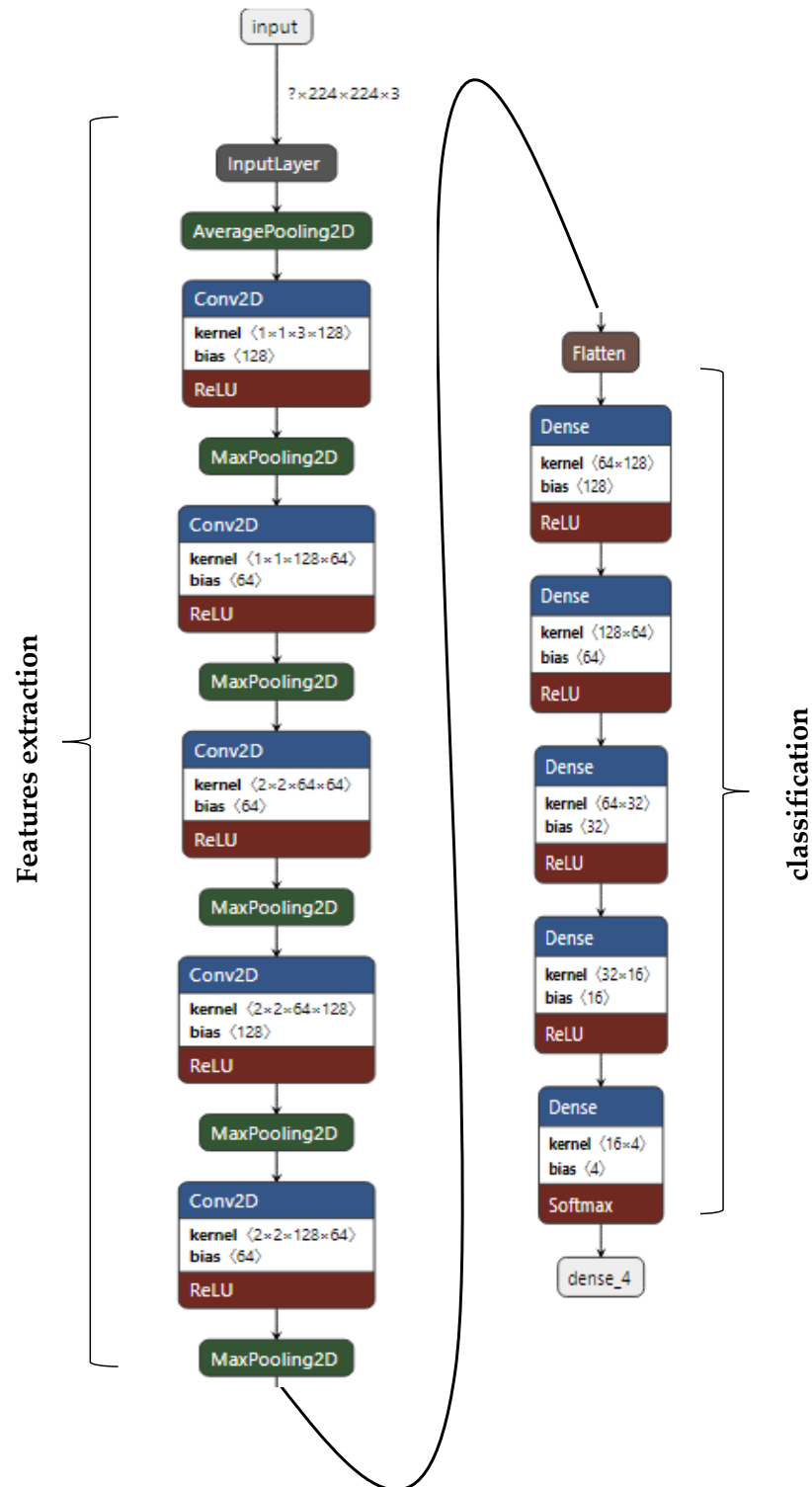

Figure 13. Configuration of the developed DeepCNN model for fault diagnosis.

Figure 14 depicts the system memory, GPU memory, and disk capacity requirements. As can be seen, the code required about 80% (about 20 GB) of the available system memory capacity (25 GB). However, it consumed most of the available GPU RAM (16 GB). In this case, too, the disk with a capacity of 128 GB was large enough to support the size of database #2.
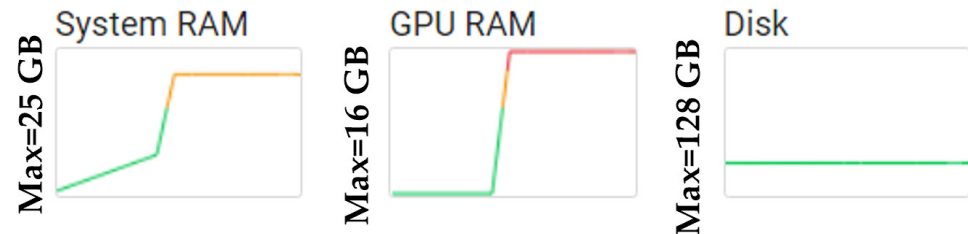

Figure 14. RAM and disk consumption during the training process (multiclass classification).

Figure 15a shows the variations in the training loss and accuracy functions. As can be seen, the loss has reached 0.01 and the accuracy was very close to 1. Figure 15b displays the obtained confusion matrix. The ROC curve shows that all the examined classes were very close to 1 (Figure 15c). From the confusion matrix, it can be seen that in the first row, the classes were correctly classified; however, in the second row, 15 images were misclassified. In addition, in the third row, five images were misclassified, and nine images are misclassified in the fourth row.
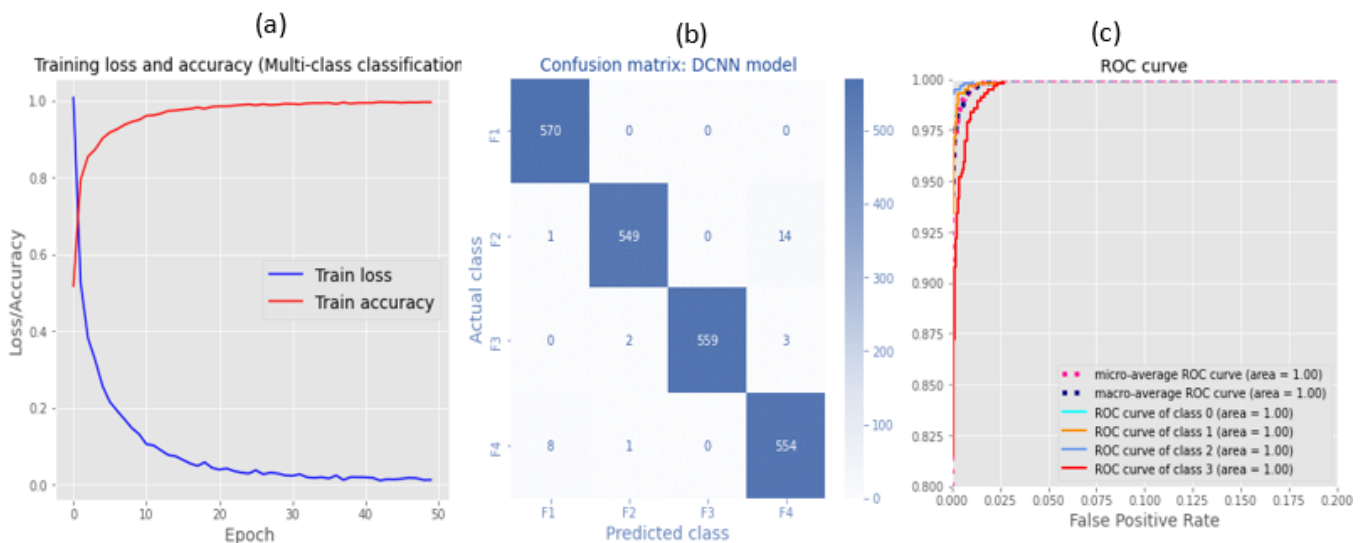

Figure 15. Multiclass classification (fault diagnosis): (**a**) training loss and accuracy functions; (**b**) confusion matrix; and (**c**) ROC curve.

F1 (short-circuited PV module), F2 (dust accumulation), F3 (shading effect) and F4 (damaged bypass diode) achieved classification accuracies of 100% (570 out of 570, 0 errors), 97.34% (549 out of 564, 15 errors), 99.11% (559 out of 564, five errors) and 98.40% (554 out of 563, five errors), respectively.

For more details, the calculated error metrics, including precision, recall, F1-score and accuracy, are listed in Table 3, as well as the required running time.

The accuracy was 98.71%, and the precision, recall, and F1-score ranged between 97% and 100%, clearly showing the accuracy of the developed fault diagnosis model. The fault diagnosis model required more time (15 min) than the fault detection model.

Table 3. Calculated error metrics (precision, recall, F1-score and accuracy), and running time (during the training process).

Class	Precision	Recall	F1-Score	Support	Training Time (min)	Accuracy (%)
Fault 1	0.98	1.00	0.99	570		
Fault 2	0.99	0.97	0.98	564		
Fault 3	1.00	0.99	1.00	564		
Fault 4	0.97	0.98	0.98	563	15 min	98.71
Accuracy (%)	-	-	-	-		
Macro average	0.99	0.99	0.99	2261		
Weighted average	0.99	0.99	0.99	2261		

4. Comparative Study

To check the effectiveness of the proposed fault detection and diagnosis method, we conducted a comparative study with other machine learning (e.g., k-NN, SVM, and CatBoost), deep learning (VGG-16), and hybrid (CNN-SVM) methods. Both databases were used to design fault detection and diagnosis models. We used the same procedure as that shown in Section 2.6. Figure 16 shows the simulated confusion matrices for different methods of fault detection (binary classification).
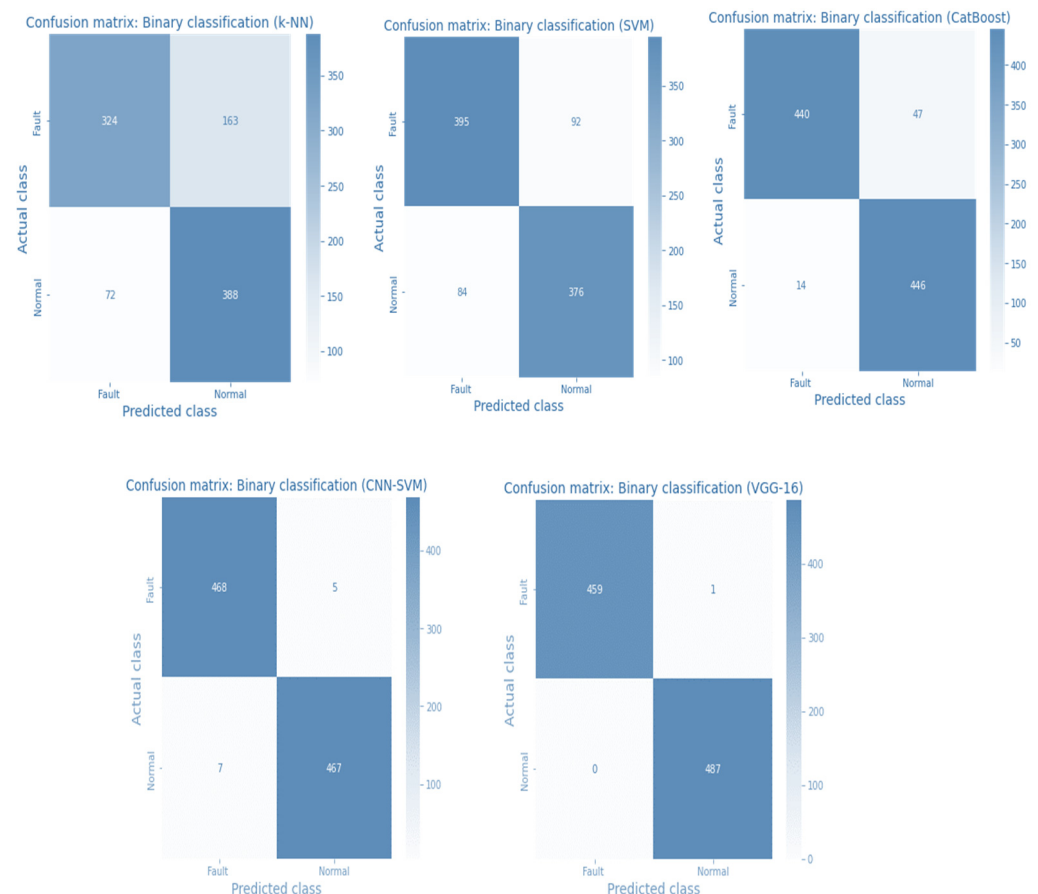
Figure 16. Binary classification (fault detection): confusion matrix using k-NN, SVM, CatBoost, CNN-SVM and VGG-16.

The calculated error metrics, including precision, recall, F1-score, and accuracy, are provided in Table 4, as well as the required running time.

Table 4. Calculated error metrics (precision, recall, F1-score and accuracy), and running time (during the training process): k-NN, SVM, CatBoost, VGG-16 and hybrid CNN-SVM.

Class	Precision	Recall	F1-Score	Training Time (min)	Accuracy (%)
k-NN Model: (n_neighbors = 5, n_jobs = 1, algorithm = 'kd_tree')					
Faulty	0.82	0.67	0.73	13+	75.18
Normal	0.70	0.84	0.77		
SVM model: (kernel = rbf, probability = True, random_state = 42)					
Faulty	0.82	0.81	0.82	13+	81.41
Normal	0.80	0.82	0.81		
CatBoost Model: (Iterations = 500, max_depth = 5, n_estimators = 100, learning_rate = 0.99)					
Faulty	0.97	0.90	0.94	13+	93.56
Normal	0.90	0.97	0.94		
Hybrid CNN-SVM model: (11 layers: 4Con2D, 4 MaxPooling, 1 dropout, 1 flatten layer and 1 dense layer)					
Faulty	0.99	0.99	0.99	7	98.73
Normal	0.99	0.99	0.99		
VGG-16					
Faulty	1.0000	0.9978	0.9989	25	99.89
Normal	0.998	1.000	0.999		

Figure 17 shows the simulated confusion matrices for different methods of fault diagnosis (multiclass classification) for k-NN, SVM, CatBoost, VGG-16 and CNN-SVM.

With reference to the results presented in Tables 4 and 5, Figures 13 and 14, the following points can be highlighted:

- In the case of machine learning, it takes about 13 min to process the first database (binary classification) and 32 min to process the second database (multiclass classification) to extract features from images.
- The worst classification accuracy is obtained by the k-NN (70.10%) and SVM (81.41%), while acceptable accuracy is given by the CatBoost classifier (93.56%).
- Deep learning methods perform better than the investigated machine learning models (k-NN, SVM and CatBoost).
- VGG-16 based on the transfer learning approach provides very good results (more than 99% in both binary and multiclass classifications), but the main issue with this approach is the complexity of its configuration and the fact that it takes too much time compared to the other examined configurations.
- Hybrid approaches that use the CNN learning process to extract features and the ML classifier also did well (98.73% for binary classification and 98% for multiclass classification).

The proposed deepCNN configuration shows good accuracy and comprises a small configuration that could be implemented easily.
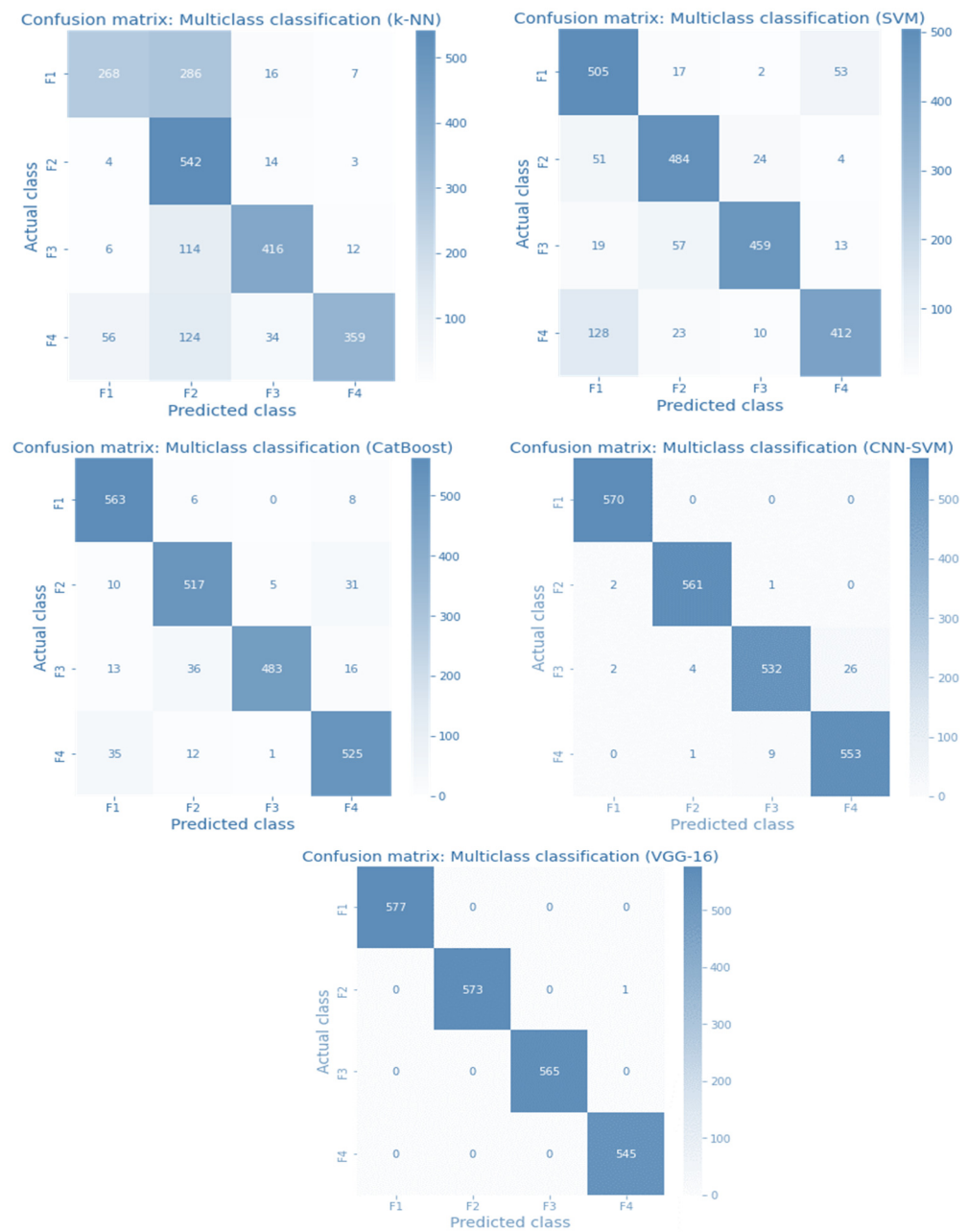
Figure 17. Multiclass classification (fault diagnosis): confusion matrix using k-NN, SVM, CatBoost, CNN-SVM and (VGG-16).

Table 5. Calculated error metrics (precision, recall, F1-score and accuracy), and running time (during the training process): k-NN, SVM, CatBoost, VGG-16 and hybrid CNN-SVM.

Classes	Precision	Recall	F1-Score	Training Time (min)	Accuracy (%)
k-NN Model: (n_neighbors = 5,n_jobs = 1,algorithm = 'kd_tree')					
Fault 1	0.80	0.46	0.59	37	70.10
Fault 2	0.51	0.96	0.67		
Fault 3	0.87	0.76	0.81		
Fault 4	0.94	0.63	0.75		
SVM model: (kernel = rbf, probability = True, random_state = 42)					
Fault 1	0.72	0.88	0.79	49	81.41

Table 5. *Cont.*

Classes	Precision	Recall	F1-Score	Training Time (min)	Accuracy (%)
Fault 2	0.83	0.86	0.85		
Fault 3	0.93	0.84	0.88		
Fault 4	0.85	0.72	0.78		
CatBoost Model: (Iterations = 500, max_depth = 5, n_estimators = 100, learning_rate = 0.99)					
Fault 1	0.91	0.98	0.94	41	93.56
Fault 2	0.91	0.92	0.91		
Fault 3	0.99	0.88	0.93		
Fault 4	0.91	0.92	0.91		
Hybrid CNN-SVM model: (11 layers: 4Con2D, 4 MaxPooling, 1 dropout, 1 flatten layer and 1 dense layer)					
Fault 1	0.99	1.00	1.00	7	98.00
Fault 2	0.99	0.99	0.99		
Fault 3	0.98	0.94	0.96		
Fault 4	0.96	0.98	0.97		
VGG-16					
Fault 1	1.0000	1.0000	1.0000	45	99.95
Fault 2	1.0000	0.9983	0.9991		
Fault 3	1.0000	1.0000	1.0000		
Fault 4	0.9982	1.0000	0.9991		

5. Conclusions

This research work presents a comprehensive comparative study between machine learning, deep learning, and hybrid methods for fault detection and diagnosis of PV modules by means of infrared thermography images. A deepCNN configuration was developed from scratch for binary and multiclass classification. The results showed the ability of the developed deepCNN to detect and classify four major faults that can occur in PV modules with very good accuracy (98.70% for both binary and multiclass classification cases).

A comparative study demonstrated that fault detection and diagnosis based on deep learning presented the highest accuracy versus machine learning methods. The VGG-16 provides more accurate results. However, the proposed configuration is the most appropriate with regard to simplicity of implementation.

As a continuation of the present work, and in order to experimentally verify and generalize the developed model, our next work aims to incorporate the deepCNN model into a low-power edge device for real-world application. The IRT images collected by drones will also be used to test how well the model works for other PV installations and locations.

Author Contributions: Conceptualization, S.B. and N.G.; Methodology, A.M. and S.K.; software, validation, S.B., N.G. and A.M.; formal analysis, A.M.; investigation, S.K.; writing—original draft preparation, A.M., N.G. and S.B.; writing—review and editing, S.B., A.M, S.K. and N.G.; visualization, S.K.; supervision, S.B.; funding acquisition, S.B.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest in this work.

Abbreviations

AE-LSTM	Auto-Encoder LSTM
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
DL	Deep Learning
ICNM	Isolated Convolutional Neural Model
IR	InfraRed Image
IRT	IR Thermography
KNN	k-Nearest Neighbor
LSTM	Long Short-Term Memory
LWL	Locally weighted learning
ML	Machine Learning
NN	Neural Networks
PVM	PV module
ReLU	Rectified Linear Unit
RoI	Region of Interest
SHM	Structural Health Monitoring
SVM	Support Vector Machine
UAV	Unmanned aerial vehicle
VGG	Visual Geometry Group

References

1. Pvps, I.; Masson, G.; Kaizuka, I.; Detollenaere, A.; Lindahl, J. *Snapshot of Global PV Markets*; Report IEA PVPS T1-35:2019; IEA: Paris, France, 2021; p. 23.
2. Mellit, A.; Tina, G.M.; Kalogirou, S.A. Fault Detection and Diagnosis Methods for Photovoltaic Systems: A Review. *Renew. Sustain. Energy Rev.* **2018**, *91*, 1–17. [CrossRef]
3. Gaabour, A.; Metatla, A.; Kelaiaia, R.; Bourennani, F.; Kerboua, A. Recent bibliography on the optimization of multi-source energy systems. *Arch. Comput. Methods Eng.* **2019**, *26*, 809–830. [CrossRef]
4. Mellit, A.; Kalogirou, S. Artificial Intelligence and Internet of Things to Improve Efficacy of Diagnosis and Remote Sensing of Solar Photovoltaic Systems: Challenges, Recommendations and Future Directions. *Renew. Sustain. Energy Rev.* **2021**, *143*, 110889. [CrossRef]
5. Venkatesh, S.N.; Sugumaran, V. A Combined Approach of Convolutional Neural Networks and Machine Learning for Visual Fault Classification in Photovoltaic Modules. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2022**, *236*, 148–159. [CrossRef]
6. Venkatesh, S.N.; Sugumaran, V. Machine Vision Based Fault Diagnosis of Photovoltaic Modules Using Lazy Learning Approach. *Measurement* **2022**, *191*, 110786. [CrossRef]
7. Ibrahim, M.; Alsheikh, A.; Awaysheh, F.M.; Alshehri, M.D. Machine Learning Schemes for Anomaly Detection in Solar Power Plants. *Energies* **2022**, *15*, 1082. [CrossRef]
8. Kurukuru, V.B.; Haque, A.; Khan, M.A.; Tripathy, A.K. Fault Classification for Photovoltaic Modules Using Thermography and Machine Learning Techniques. In Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS), Baltimore, MD, USA, 29 September–3 October 2019; pp. 1–6.
9. Kurukuru, V.S.; Haque, A.; Tripathy, A.K.; Khan, M.A. Machine Learning Framework for Photovoltaic Module Defect Detection with Infrared Images. *Int. J. Syst. Assur. Eng. Manag.* **2022**, *23*, 1771–1787. [CrossRef]
10. Jumaboev, S.; Jurakuziev, D.; Lee, M. Photovoltaics Plant Fault Detection Using Deep Learning Techniques. *Remote Sens.* **2022**, *14*, 3728. [CrossRef]
11. Li, X.; Yang, Q.; Lou, Z.; Yan, W. Deep Learning Based Module Defect Analysis for Large-Scale Photovoltaic Farms. *IEEE Trans. Energy Convers.* **2018**, *34*, 520–529. [CrossRef]
12. Akram, M.W.; Li, G.; Jin, Y.; Chen, X.; Zhu, C.; Ahmad, A. Automatic Detection of Photovoltaic Module Defects in Infrared Images with Isolated and Develop-Model Transfer Deep Learning. *Sol. Energy* **2020**, *198*, 175–186. [CrossRef]
13. Lu, F.; Niu, R.; Zhang, Z.; Guo, L.; Chen, J. A Generative Adversarial Network-Based Fault Detection Approach for Photovoltaic Panel. *Appl. Sci.* **2022**, *12*, 1789. [CrossRef]
14. Ahmed, W.; Hanif, A.; Kallu, K.D.; Kouzani, A.Z.; Ali, M.U.; Zafar, A. Photovoltaic Panels Classification Using Isolated and Transfer Learned Deep Neural Models Using Infrared Thermographic Images. *Sensors* **2021**, *21*, 5668. [CrossRef] [PubMed]

15. Su, Y.; Tao, F.; Jin, J.; Zhang, C. Automated Overheated Region Object Detection of Photovoltaic Module with Thermography Image. *IEEE J. Photovolt.* **2021**, *11*, 535–544. [CrossRef]

16. Kumar, G.; Bhatia, P.K. A Detailed Review of Feature Extraction in Image Processing Systems. In Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, India, 8–9 February 2014; pp. 5–12.

17. Medjahed, S.A. A Comparative Study of Feature Extraction Methods in Images Classification. *Int. J. Image Graph. Signal Process.* **2015**, *7*, 16. [CrossRef]

18. Kellil, N.; Aissat, A.; Mellit, A. Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions. *Energy* **2023**, *263*, 125902. [CrossRef]

19. Pierdicca, R.; Paolanti, M.; Felicetti, A.; Piccinini, F.; Zingaretti, P. Automatic faults detection of photovoltaic farms: SolAIr, a deep learning-based system for thermal images. *Energies* **2020**, *13*, 6496. [CrossRef]

20. Cui, F.; Tu, Y.; Gao, W. A Photovoltaic System Fault Identification Method Based on Improved Deep Residual Shrinkage Networks. *Energies* **2022**, *15*, 3961. [CrossRef]

21. Mellit, A. An embedded solution for fault detection and diagnosis of photovoltaic modules using thermographic images and deep convolutional neural networks. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105459. [CrossRef]

22. Wang, M.H.; Lin, Z.H.; Lu, S.D. A fault detection method based on cnn and symmetrized dot pattern for pv modules. *Energies* **2022**, *15*, 6449. [CrossRef]

23. Ma, L.; Crawford, M.M.; Tian, J. Local Manifold Learning-Based k-Nearest-Neighbor for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [CrossRef]

24. Lo, C.-S.; Wang, C.-M. Support Vector Machine for Breast MR Image Classification. *Comput. Math. Appl.* **2012**, *64*, 1153–1162. [CrossRef]

25. Samat, A.; Li, E.; Du, P.; Liu, S.; Miao, Z.; Zhang, W. CatBoost for RS Image Classification with Pseudo Label Support from Neighbor Patches-Based Clustering. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 8004105. [CrossRef]

26. Kaur, T.; Gandhi, T.K. Automated Brain Image Classification Based on VGG-16 and Transfer Learning. In Proceedings of the 2019 International Conference on Information Technology (ICIT), Bhubaneswar, India, 19–21 December 2019; pp. 94–98.

27. Xin, M.; Wang, Y. Research on Image Classification Model Based on Deep Convolution Neural Network. *J. Image Video Proc.* **2019**, *2019*, 40. [CrossRef]

28. Sun, X.; Liu, L.; Li, C.; Yin, J.; Zhao, J.; Si, W. Classification for Remote Sensing Data With Improved CNN-SVM Method. *IEEE Access* **2019**, *7*, 164507–164516. [CrossRef]

29. Machart, P.; Ralaivola, L. Confusion matrix stability bounds for multiclass classification. *arXiv* **2012**, arXiv:1202.6221.