Energy Conversion and Management 314 (2024) 118718

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

An innovative transformer neural network for fault detection and classification for photovoltaic modules

E.A. Ramadan [a], Nada M. Moawad [b,*], Belal A. Abouzalm [a], Ali A. Sakr [b], Wessam F. Abouzaid [b], Ghada M. El-Banby [a]

[a] Department of Industrial Electronics and Control Engineering, Faculty of Electronic Engineering, Menoufia University, 32952 Menouf, Egypt
[b] Department of Electrical Engineering of Computer and Control Systems, Faculty of Engineering, Kafrelsheikh University, 33516 Kafrelsheikh, Egypt

ARTICLE INFO

Keywords:
Fault Detection System
Photovoltaic (PV) systems
Artificial Intelligence
Vision Transformer
Thermography

ABSTRACT

Solar energy from photovoltaic systems (PV) ranks as the third greatest renewable electricity generation resource, expanding quickly through the years as it is free from environmental pollution and has cheap installation costs. Effective performance at high working rates is contingent on the early failure detection of PV modules. This study introduces an innovative deep learning model utilizing a Vision Transformer (ViT) artificial neural network (ANN) for the automatic detection of faults in infrared thermography (IR) images of PV modules. Our approach aims to enhance the accuracy of PV fault detection and classification compared to existing deep learning methods. The proposed framework encompasses three primary stages: (1) image preprocessing, which includes the application of the unsharp mask to sharpen the image's edges or high-frequency components; (2) data augmentation techniques designed to overcome the problem of unbalanced classes that affect the training process, resulting in learning specific majority classes better than others; and (3) implementing a Vision Transformer deep learning model for its precision in digital image analysis. We evaluated the framework using the public Infrared Solar Modules dataset. The performance was quantitatively assessed using several metrics: accuracy, recall, precision, and F1 score. The dataset is classified into eleven different PV anomalies and another class of no-anomaly PV modules. The results show that our proposed approach has 98.23% accuracy for classifying the dataset into two classes, one for the PV anomaly and the other for the no-anomaly. It also has 96.19% accuracy for classifying eleven PV failures and 95.55% for twelve classes, including the no-anomaly class with the eleven types of anomalies. The experimental results underscore the potential of our model for earlier and more precise detection of PV faults. Furthermore, comparative analysis revealed the superior performance of the proposed approach over other deep learning methods.

1. Introduction

Governments and researchers are currently paying the most attention to renewable energy sources (RES), particularly solar systems. According to the International Renewable Energy Agency's (IRENA) report [20], by the end of 2022, the installed solar power capacity had increased to 1,062 GW, having grown at a 50 % annual pace over the previous ten years, which demonstrates that solar energy is preferred over other renewable energy sources for electrical power production systems due to its affordable production. The amount of actual electricity generated by solar energy sources, in 2021 was 1,034 TWh, with a 23 % increase in that year, which contributes 13 % of total electricity generated by RES as shown in Fig. 1.

PV systems that produce solar energy must be safeguarded from malfunctions that would sap their power and pose a number of concerns, including the worst-case scenario of fires [23]. To address these failures, fault detection and diagnosis systems are crucial. These systems can identify the type of defect and its precise position, enabling the correct decision to be made regarding whether to repair or replace the failed PV component [12].

The automatic fault detection and diagnosis (FDD) models can be divided into these categories: quantitative method, which is based on

* Corresponding author.
E-mail addresses: Ebrahim_ramadan@el-eng.menofia.edu.eg (E.A. Ramadan), Nada_Hanfy2011@eng.kfs.edu.eg (N.M. Moawad), Drbelalabozalam@yahoo.com, belal.abou2015@el-eng.menofia.edu.eg (B.A. Abouzalm), ali_asakr@eng.kfs.edu.eg (A.A. Sakr), wesam.abozaid@eng.kfs.edu.eg (W.F. Abouzaid), ghadaelbanby75@gmail.com, ghada.elbanby@el-eng.menofia.edu.eg (G.M. El-Banby).

https://doi.org/10.1016/j.enconman.2024.118718
Received 6 January 2024; Received in revised form 19 June 2024; Accepted 19 June 2024
Available online 26 June 2024

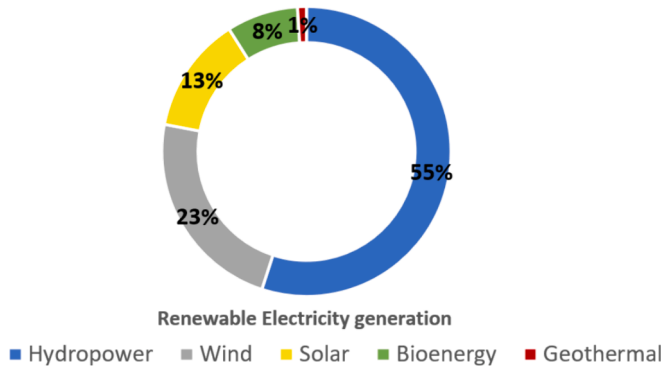


Fig. 1. IRENA report.

mathematical model fundamental principles; the qualitative method, which is based on if-then rules and decision trees; and the data-based model, which depends on process data history.

Artificial intelligence (AI) approaches enable computers to learn from prior experience. It has been used extensively in a wide range of industries, including engineering, robotics, and many more. In several aspects of RES, especially in PV system studies, including prediction, it is a potent and significant instrument [38,18,28]. Different methods, like statistical techniques or machine learning and deep learning, which can handle complicated and non-linear issues, can be employed in data-driven defect detection for PV systems. ANN [17,13], fuzzy logic [4], support vector machine (SVM) [1] are some examples of AI systems applied in PV FDD systems.

In model-based FDD, the system's mathematical models are constructed based on an understanding of the physical design principles under typical operating conditions. Signal analysis uses the input–output data from the PV panel model and solves the nonlinear equations. The presence of a system problem is identified using the variations between measurements of the real system and model projections. An example of a model-based FDD approach called single-diode model [3] it is accurate under high irradiance conditions but less accurate under low irradiance conditions.

A priori knowledge is necessary for the data-based approach because it draws its information from a vast amount of process history data [8,44]. The data-driven FDD technique makes use of training data from various operating circumstances and failed scenarios to determine the link between a set of input and output signals.

In [19] digital twin virtual or digital representation that replicates the behavior of a real object is used to perform model-based DC power generation of PV arrays and establish a technique of PV fault detection as a hybrid data-driven and model-based FDD system. This paper presents the convolutional mixer (Conv-Mixers) for classifying PV faults that builds on patch embedding and combines depth-wise and point-wise convolutions. Conv-Mixer receives as input 2D pictures created by applying a Markov transition field transform to data on PV DC array power. A long-range notification system (LoRa) is used in the paper that is suitable for usage in PV farms, which are low-power wide-area networks.

There are many types of PV failures, according to their effect on the system or the reason for the failure from the beginning. Delamination, bubbles, yellowing, burns, deterioration, hot spots, scratches, or crack faults are examples of long-lasting failures according to the state of the cells. Open circuit, closed circuit, potential-induced degradation, junction j-box failures, diode failures, and inverter failures are a few examples of electrical connection failures in a PV system. As a result, permanently faulty modules can be easily taken out and changed. According to received light or partial shade effects, temporal defects develop. Without removing the modules, temporary issues like shadowing and soiling of dirt or snow can be quickly fixed [15].

The authors in [4], have used a fuzzy classification algorithm. According to three index values: healthy condition, EVA fault, and delamination fault, in this work, failure can be identified using the pixel counting approach for thermal imaging to detect the discoloration of EVA and delamination failures. However, it doesn't diagnose other kinds of defects; it simply concentrates on where the hot area is.

A hybrid features-based support vector machine model for hot spot detection and PV panel categorization is introduced in [1]. A data fusion strategy is used to create color histograms, also using a 2nd-order co-occurrence matrix and a local binary pattern to get the characteristics of images.

The authors [30] have examined various convolutional neural network models using thermal images that were captured by ground-based and unmanned aerial vehicle (UAV) operators. They have a high-performance classifier of PV photos as an operational, or hotspot PV module, using pre-processing methods such as normalization, gray scaling, thresholding, Sobel Feldman and, box blur filtering.

In [17], an intelligent system is put out that automatically locates and allocates relative hot solar panel locations by fusing telemetry data with region-based convolutional neural networks (R-CNN).

A new system using a dataset of electro-luminescence images for finding solar cell faults was designed by [13], which is compatible with mobile and low-power computing hardware. K-means, Mobile-Net-V2, and linear discriminant algorithms are used to group solar cell photos into clusters and create a detection model for each cluster that is created. It can clear up confusion between various cell shapes and extract the characteristics that distinguish faulty solar cell types from non-defective solar cells.

1.1. Motivations and contributions

The requirement for finding a method that makes that task simple and accurate is motivated by the need for an existing PV FDD system that can detect defects in an automated manner with no need for interference from experts to assist PV system maintenance and protect the system from power losses or other different risks. By proposing a new contactless methodology for detecting and classifying PV module failures depending on thermal dataset images and deep learning using transformer neural networks, whose modeling accuracy has been demonstrated to be comparable to that of convolutional neural networks (CNN), it can outperform many CNN models. In our paper, we introduce a novel method of PV FDD that is applied to 20,000 images of the public infrared solar module dataset that was collected from actual large-scale PV fields in 25 countries on six continents. The contributions of the proposed methodology are:

- Preprocessing the raw dataset using image filters to enhance their appearance by sharpening the edges, which makes the image details clear and helps distinguish between different faults. Increasing the model globality performance by applying the offline oversampling augmentation technique also solves the problem of unbalanced classes of datasets and doubles the number of images.
- A novel PV fault detection model is applied using a modified multiscale vision transformer, ViT ANN, which basically depends on the self-attention technique that produces relationships between a certain part of an image and the rest of the same image. The distinct PV fault type depends on the whole image characteristics and their relations and importance between all parts of the image, not only specific local spatial features as in CNN methodology [9,46]. That globality of transformers makes our model more flexible and scalable and increases its accuracy. Also, modifying the traditional ViT architecture by adding a transformer encoder layer in parallel as a multilayer makes our proposed multiscale ViT model obtain more dependencies or relations with different knowledge from images using different multilayer decoders and combines this information to encode the correct type of PV anomaly, which improves the

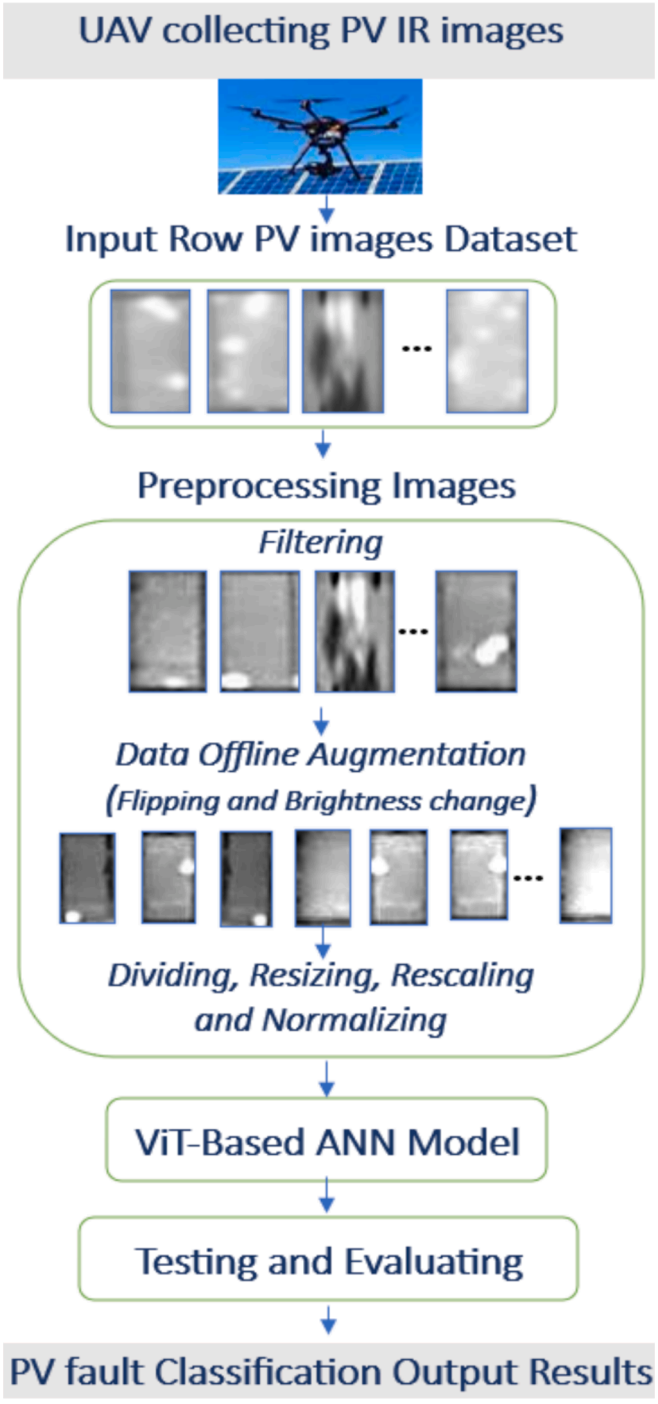


Fig. 2. Phases of the proposed FDD Method.

performance, robustness and globality of the model during the training process.

- Training and evaluating our model for identifying and categorizing PV IR images into anomaly class and no-anomaly class, as well as training the model and proving its performance to classify eleven different and difficult failures of PV modules like soiling, cracking, and hotspots. The results ensure the model's capability for identifying and classifying PV anomalies with superior accuracy compared to other CNN models and also previous papers.

The paper is organized as follows: section 2 discusses the stages of dataset preprocessing before training the model to illustrate image features and increase system performance; it also presents the detailed proposed AI transformer model for fault detection and classification; section 3 presents the simulation results; and section 4 concludes our work.

2. Methodology

To mitigate any potential risks to the photovoltaic system, we suggest a deep learning technique that enhances fault diagnosing systems through the use of a data-driven fault detection system that has been successful in classifying various PV module failures of thermographic images. This allows the solar system to be protected without requiring expert intervention or destructive methods.

Data acquisition of images for real-time diagnosis can be made online by a standard drone equipped with a high-resolution IR camera. According to the resolution and the number of images required to cover the entire photovoltaic modules, the drone can capture images, at a rate of approximately one image per second for a 20-megapixel resolution camera. For a photovoltaic farm of 1 MW (around 1.5:2 ha), it would take about 15–20 min [41,37]. Most drones now are equipped with Wi-Fi or cellular capabilities for transferring captured images to be processed. For example, an average image size of 5 MB and a transfer rate of 10 MB/s could take one image to be transmitted for about 0.5 s (1000 images take around 8–9 min) [48,47]. Depending on the criticality of the real-time fault detection system, flights can be scheduled at regular intervals, such as every hour for highly important systems or daily flights for less critical ones. In case a fault occurs after a drone returns, a suggested implementation for a hybrid monitoring approach to overcome delay using regular drone flights besides fixed cameras to provide immediate fault detection in critical areas [22,45]. Another suggestion for predictive maintenance algorithms is to utilize historical data and environmental conditions to predict potential faults and then schedule drone flights proactively [24,33].

Although our suggested work is applied offline at the public infrared solar module dataset, we can enhance our work by implementing it online in the future.

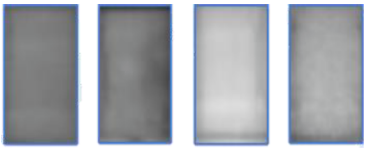
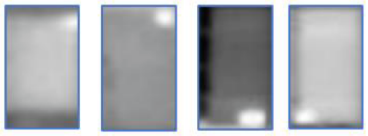
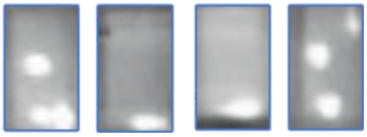
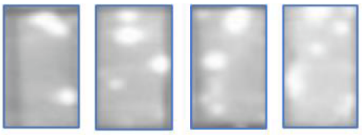
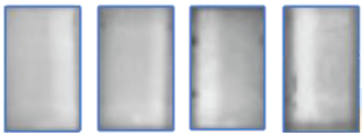
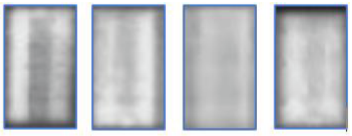
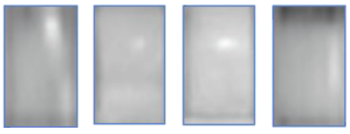
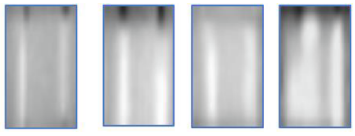
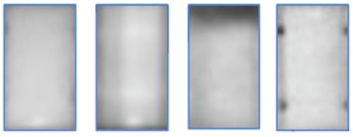
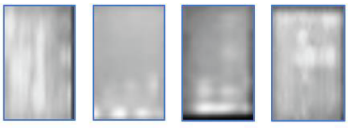
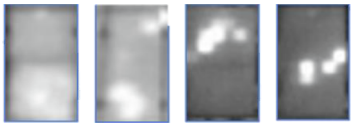
The first phase of the proposed methodology after collecting data images, as shown in Fig. 2, is preprocessing by applying image filters (sharpening filter) to enhance image edges and using offline data augmentation alongside resizing, rescaling, and normalizing the image to increase the system performance and speed up the training process. Secondly, applying AI and deep learning of vision transformer ViT-based ANN [9] for modeling the complex and nonlinear PV fault detection system FDD, which can detect and classify the faulty module remotely without interfering with or disturbing the normal operation of the PV system, based on learning the machine to do that job instead of humans. Finally, training our method, evaluating, and testing the results.

2.1. Dataset images of infrared solar modules

There are eleven different types of PV module defects that are included in those public dataset images [31], which benefit our fault detection model by getting information that helps it be more generalized according to varieties of faults. The dataset includes 20,000 PV module images from twelve classes, which are not balanced in numbers. Raptor Maps Inc. uses mid-wave and long-wave IR cameras (3–13.5 μm) in a piloted aircraft and also UAV systems to collect the images of solar panel modules with 24 × 40 × 1 resolution and 8-bit depth representing the temperature values in grayscale, with 3.0 to 15.0 cm/pixel spatial resolution due to the variation of distances between the IR camera and the modules.

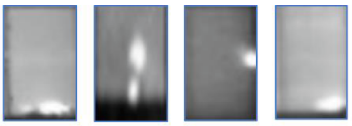
Dataset images are divided into 10,000 images for the no-anomaly class, in which their PV modules have not experienced any type of failure or nominal operation. The rest of the images are not evenly distributed into eleven different PV anomalies, e.g., diode, diode-multi, hot spot, and shadowing. Owing to the practical findings, the anomalous

Table 1
Detailed distribution, description, and some samples of dataset anomalies.

No.	Class Name	Description	No. of images	Samples
1	No Anomaly	Normal-operated module without any anomalies	10,000	
2	Cell	Only one cell has a square hot spot	1,877	
3	Cell Multi	Multiple cells with square hot spots	1,288	
4	Cracking	Cracked surface of the module	940	
5	Diode	A typical activated bypass diode makes up one-third of the module	1,499	
6	Diode Multi	Many active bypass diodes, usually affecting two-thirds of the module	175	
7	Hot Spot	A thin-film module has one pointed hot spot	249	
8	Hot Spot Multi	A thin-film module or snail trail has multiple pointed hot spots	246	
9	Offline Module	The whole module is disconnected and heated	827	
10	Shadowing	Plants, man-made structures, or contiguous rows that block the sun	1,056	
11	Soiling	Debris on the surface of the module, such as dirt or dust	204	

(*continued on next page*)

Table 1 (*continued*)

No.	Class Name	Description	No. of images	Samples
12	Vegetation	Vegetation, e.g., trees blocking PV panels	1,639	

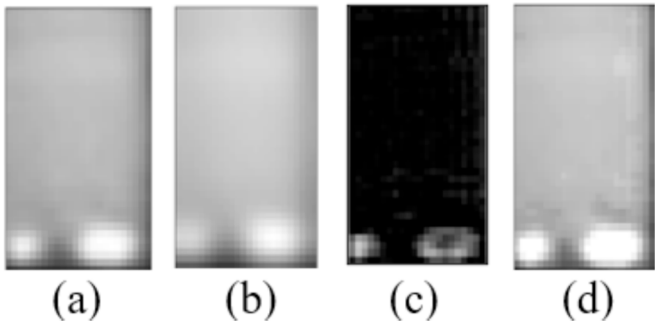


Fig. 3. Unsharp filter steps : (a) Original image, (b) Smoothed image, (c) Edged image, (d) Sharpened output image.

PV modules in the dataset impacted 21,793 KW_{DC} of power production overall, or 16.93 % of all anomalies [25]. The detailed distribution, description of those anomalies, and some samples of dataset images are illustrated in Table 1.

2.2. Pre-processing images

In the proposed methodology, we prepare the thermography dataset images to extract the most helpful information from the PV anomalies that were displayed on them. The preprocessing process is optimized through two stages: the first is the filtering of images using a sharp filter, and the second is the offline augmentation using the oversampling technique.

2.2.1. Filtering

The resolution of the PV module images in the dataset is small, and its details need to be enhanced and cleared enough using filtering to make the classification task more accurate and improve performance and testing results. The appropriate filter must be chosen carefully due to the circumstances of the image details, which might be distorted using the incorrect filter.

The first preprocessing technique applied to the dataset image is the unsharp filter, which is preferred when low noise levels are present in the images (due to their poor resolution). It is chosen to sharpen the image's edges or high-frequency elements. A sharpening filter functions by making areas of transition in an image more contrasty. This filter uses a low-pass filter as an embedded mid-step to get the edged masked image when subtracting the blurred image from the original image, and then adds the original image to its edged image to get the final sharper image (Eqs. (1) and (2)). Real sharpening steps can be more explained as follows (see example in Fig. 3):

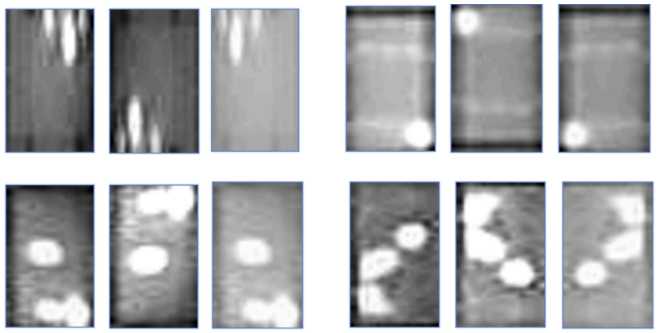


Fig. 4. Samples of the new dataset images N_d after augmentation.

- Smooth the original image $\boldsymbol{I(x,y)}$ (remove high-frequency edges) by applying low-pass filters like median or gaussian filters, so the resulted blurred image $\boldsymbol{I_{smooth}(x,y)}$ is the smoothed image of the original one.
- Subtract the smoothed image version from the original one to get the edges of the original image as an edge image $\boldsymbol{I_{edge}(x,y)}$ in Eq. (1). (like output of Laplacian edge detection filter).
- Add the original image with the edged image after multiplying it by a scaling factor $\boldsymbol{S}$ that may vary between 0.2 and 0.7 to get the sharpened image $\boldsymbol{I_{sharp}(x,y)}$ in Eq. (2).

$$I_{edge}(x,y) = I(x,y) - I_{smooth}(x,y) \tag{1}$$

$$I_{sharp}(x,y) = I(x,y) + S^{*} I_{edge}(x,y) \tag{2}$$

We applied a digital unsharp masking using ImageFilter.UnsharpMask() method of Pillow Python Imaging Library, with parameters of 2 Blur Radius and 150 % of Unsharp Strength.

The effect of filtering helps to increase the accuracy of the proposed methodology as it sharpens the shape of different fault edges, like squared hot spots in the cell anomaly class or pointed one of the thin-film modules in the hot-spot anomaly class, which improves the sensitivity and F-score for classifying the eleven faults of PV modules.

2.2.2. Offline augmentation oversampling

The second preprocessing technique implemented in our work is offline oversampling augmentation. Oversampling is one approach that is used to artificially increase the number of dataset images, which gives the classification process the ability to be generalized and increase accuracy and convergence without getting overfitting during training or missing learning the significant features of classes [2,26,21]. Using data augmentation is a promising solution to overcome the problem of unbalanced classes that affect the training process, which results in learning specific majority classes more than others. Augmentation can sometimes be used as under-sampling by deleting some of the dataset images in the classes to match the least class, as in [29]. But reducing the dataset often affects the accuracy, especially for a large number of classes, which will dramatically reduce system performance.

By applying appropriate oversampling to not lose essential features in images, we made the number of images in all classes equal and doubled the total number to reach more than 40,000. This augmentation is done offline, and then the new dataset is divided into training, validation, and testing sets [21]. We apply horizontal and vertical flipping of images by 180 degrees, as well as brightness changes in the 0.2 and 1.2 range for the images using the Image Data Generator in Keras library that will randomly make rotations, reversals, and reductions or increases in brightness for all the filtered dataset images F_d, which produces new artificial versions A_d of those images. The new dataset N_d is composed of real and artificial images as $N_d = A_d \cup F_d$. Some samples of the resulted dataset are illustrated in Fig. 4.

Choosing flipping (mirroring) when making different image copies helps the algorithm to be more generalized in that it will ensure learning

Table 2
Summary of pre-processing steps.

No	Preprocessing Steps	Technique	Programming Method	Parameters and details	Reason
1	*Filtering*	Unsharp filter (Offline)	Image-Filter. Unsharp-Mask () (Pillow Imaging Library)	• Blur Radius = 2 • Unsharp Strength = 150 %	Enhance the shape of fault edges
2	*Data Augmentation*	Oversampling (Offline)	Image Data Generator (Keras library)	• Horizontal flipping • Vertical flipping • Brightness change range = [0.2:1.2]	• Increase system's globality • Help to balance dataset classes
3	*Dividing Dataset*	Holdout Validation	Manually	• Training data = 80 % • Validation and test data = 10 %	Simple way for unbiased evaluation process
4	*Resizing and Normalization of Image*	Beginning of training	Image Data Generator (Keras library)	• Image size = 160 × 160 × 3 • Normalized image values range = [0:1]	• Increase performance, Speed up training

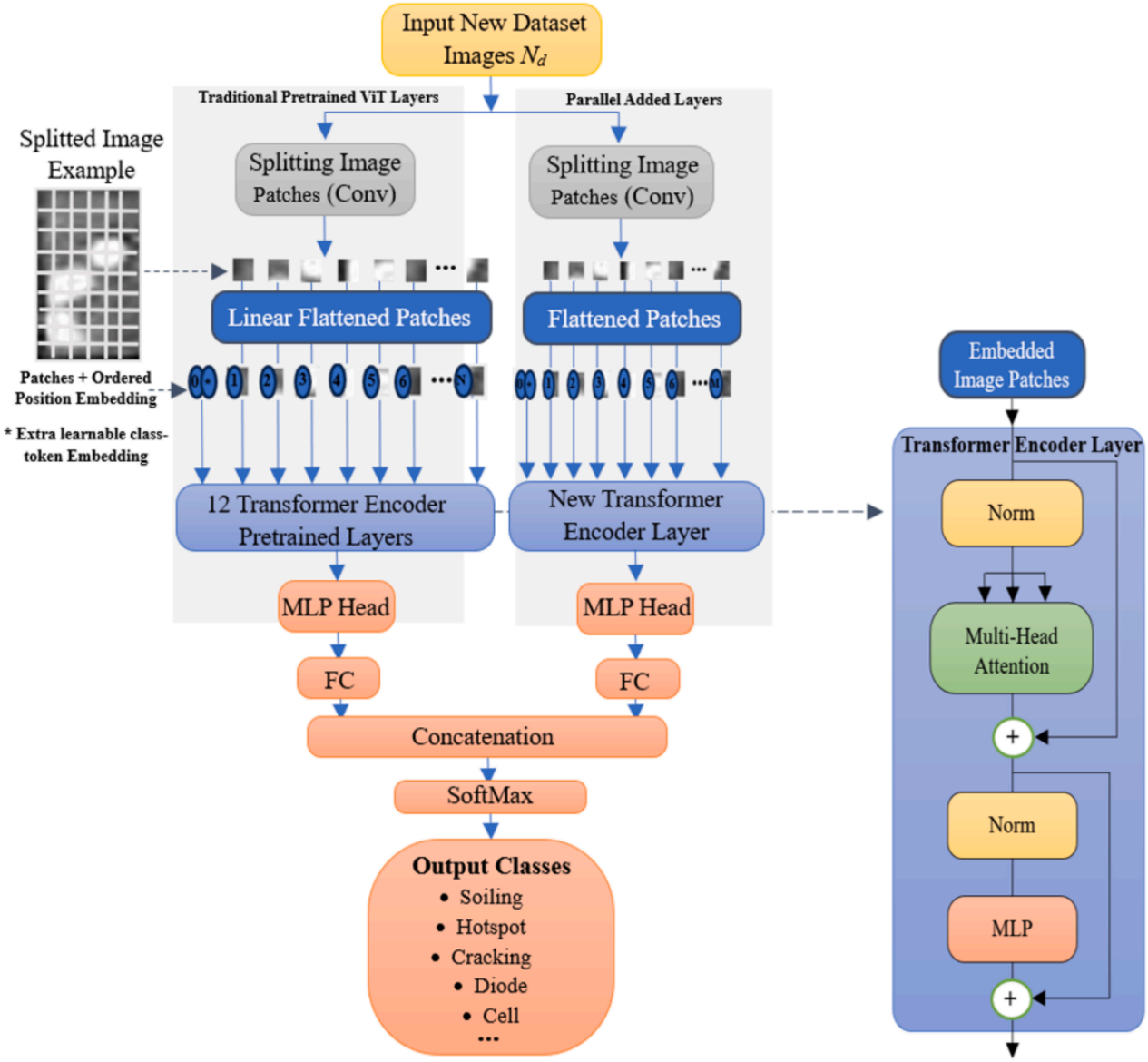


Fig. 5. Proposed ViT-based ANN model architecture.

the correct class from the eleven PV faults, wherever the location of the faulty areas (or hot spots) is, and will only focus on distinguishing the size, shape, or number of those areas. Increasing our system's globality improves performance, accuracy, stability, and convergence. Table 2 summarizes the preprocessing stages and their importance.

2.3. Proposed ViT- based ANN model

Deep learning is the best choice for modeling an automated PV FDD system that can handle nonlinear systems and process huge amounts of data to help machine learning using neural networks. Sequential data

processing (e.g., natural language processing (NLP) and time series forecasting) is an important issue in deep learning, which needs feedback loops (recurrent connections) in neural network models to deal with sequential dependencies, as in recurrent neural networks (RNN). After that, long-short-term memory NN (LSTM) is designed to solve the RNN problem of gradient vanishing while dealing with long dependencies using a gating mechanism. LSTM models are presented powerfully as one of the various frameworks that are used for PV fault detection systems [18,12], for example, the LSTM NN classifier with the aid of discrete wavelet transform analysis (signal processing techniques) for energy value feature extraction, which was introduced in [43], for detecting PV high impedance failure from a simulation of an IEEE 13-bus system with 91.21 % accuracy. On the other hand, transformer NN [42] was initially developed for NLP, and it was regarded as a revolution in sequential processing because it can process sequential input in parallel to manage complicated dependencies for large tasks using attention mechanisms. Since then, transformer NN models have seen considerable implementation, and in 2020, the authors of the vision transformer model (ViT) [9] modified the design of transformer NN to be more convenient for applications of computer vision like image recognition and classification processes.

Training model from scratch needs very large dataset to give superior accuracy for training and validating, and it is difficult to collect such a large PV dataset [21], so transfer learning is a solution in many cases to help improve system performance. For models in [21,29] and [6], knowledge is transferred from the pretrained model and then used to learn from the target dataset by retraining the model. Unfortunately, the ViT model performs inadequately on training small datasets because of a lack of inductive bias, which prevents it from being generalized [9]. So, thanks to pretraining on a huge quantity of data (the large-scale visual recognition dataset of ImageNet) and transfer learning on smaller datasets, the ViT model outperforms many famous models [46] which is considered a creative, novel method of image recognition that has shown good accuracy and promising performance in learning tests. On image recognition benchmarks, this technique outperformed a number of modern algorithms and offered higher interpretability thanks to the attention mechanism.

The use of ViT has significantly advanced biomedical science, including the classification of oral cancer [36,7], interpretation of radiography of the chest [40], and identification of cardiovascular illness [32]. ViT has also demonstrated success in the detection of different earthquakes [34], the evaluation of metal 3-D printing quality [49], the detection of smoke caused by fires [50] and also the detection of damaged parts (faults) in solar and wind power plants [11].

As shown in Fig. 5, a suggested ViT-based model for simulating an automated PV defect detection system is built using transfer learning and customized extra layers running alongside the conventional ViT-B32 model in the Keras library, then concatenating the two outputs of them for final classification.

First, the new dataset image is split into patches x_N, with a hybrid architecture that uses a convolution (Conv) layer instead of dealing with pixels to minimize complexity (two Convs in parallel with different information). Then linearizing embedded patches to 1D sequence as X and adding a positional embedding (PE) vector as standard 1D (not advanced 2D) for the embedded patch with X and PE dimensions of d_{model} Eqs. (3)–(5).

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{3}$$

$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{4}$$

where pos and i are the patch position and its current dimension.

An extra learnable embedding x_{class} is added at the beginning of patches to be updated during training and has the resultant classification.

$$X = [x_{class}; x_1 E_1; x_2 E_2; \cdots] + PE_{pos} \tag{5}$$

where, E is a trainable embedding linear projection for each patch x with $E \in \mathbb{R}^{P.P.C \times d_{model}}$, and $PE_{pos} \in \mathbb{R}^{N+1 \times d_{model}}$ (P and C are the patch size and its channels, and N is the number of patches)

The embedded patches are fed to 12 transformer encoder layers in series and one added layer in parallel as a multiscale model. The encoder of a transformer consists of a multi-head attention layer, skip or residual connection, and normalization layer, as illustrated in Eq. (6) (X is the input), and then the output is fed to a multilayer perceptron (MLP) NN, which contains two dense linear layers (feedforward network (FFN) or fully connected (FC)) in Eq. (7), with a nonlinear activation layer in between.

$$out = LayerNorm(X + Attention(X)) \tag{6}$$

$$FFN(X) = W_2\sigma(W_1 X) \tag{7}$$

where W_1 and W_2 are the matrices of the dense layers, and σ denotes the Gaussian error linear unit (GELU) activation function.

The self-attention layer mechanism (in the multi-head layer) transforms the input vector into 3 vectors: a query vector **q** that represents an image patch, keys vector **k** that represent all other image patches, and a value vector **v** that is equal to **q** (in self-attention) with dimensions ($d_k = d_v = d_q = d_{model}$). All these sets of vectors are stacked in their matrices (Q, K, and V) in Eq. (8) to calculate the relative importance or dependencies (attention) between all patches compared to a particular patch associated with the same image, using the scaled dot product of queries Q and keys K to get the degree of importance as scores, and before applying the SoftMax function to get scores probabilities, dividing scores by $\sqrt{d_{model}}$ (as a scaling factor) for normalization and stabilization, then multiplying with values V to finally get weighted values.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_{model}}}\right)V \tag{8}$$

The multi-head attention layer, which is actually implemented in the architecture, can benefit from linearly inputting various patches h times (h number of heads) in parallel. The output of one head represents one self-attention layer of matrices projected with its trainable weights as shown in Eq. (9), and to get dependencies or attentions at the same time from different important positions of image patches randomly, concatenate the outputs of all heads as in Eq. (10) [42,14].

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{9}$$

where, $W_i^Q \in \mathbb{R}^{d_q \times d_{model}}$, $W_i^K \in \mathbb{R}^{d_k \times d_{model}}$, $and W_i^V \in \mathbb{R}^{d_v \times d_{model}}$ are the projection parameter weights matrices while $d_k = d_v = d_q = d_{model}/h$.

$$MultiHead(Q, K, V) = Concat(head_1, \cdots, head_h)W^O \tag{10}$$

where, $W^O \in \mathbb{R}^{d_{model} \times d_{model}}$.

The attention mechanism in transformers allows the model to weigh the importance of different parts of the input data dynamically. Attention allows the model to consider the entire input of the PV anomaly image holistically, rather than just local patches. This global perspective helps in identifying long-range dependencies or complex patterns and correlations that might be missed by models focusing only on local features. Transformers might correlate different hot spots' locations, shapes, numbers, sizes, and exact intensity of brightness in different locations with the correct classification of PV anomaly type. Also, the model can adaptively focus on different regions of the image simultaneously, potentially highlighting the areas that are more likely to contain faults, thus improving the accuracy of detection and providing insights into which parts of the image the model considers important for making decisions. Unlike 2D-CNNs, which have some limitations

Table 3
Details of proposed architecture parameters.

Parameters	Traditional ViT-B32	Parallel Added Layers	Proposed ViT-based ANN Model
Patch size = Kernel size (Conv)	32	8	Multiscale 32 and 8
No. of Patches (Image size/Patch size)^2	25	400	25 and 400
Hidden size (Dimension d_{model}) = No. of Filters	768	100	768 and 100
Key, Value and Query dimensions ($d_k = d_v = d_q = d_{model}/h$) (Multi-Head Attention)	64	16	64 and 16
MLP size (Transformer encoder)	3072	500	3072 and 500
No. of Heads (h) (Multi-Head Attention)	12	6	12 and 6
Dropout rate (Transformer encoder)	0.1	0.5	0.1 and 0.5
No. of Transformer encoder Layers (Depth)	12	1	13
Trainable Transformer Layers	Last 6	1	7
Trainable Params	42,528,768	159,088	Total = 42,687,856
Total Params	87,436,800	159,088	Total = 87,595,888

Table 4
Comparison of the proposed model with different deep learning models.

Model	Depth (Layers)	Parameters (Millions)	Size (MB)	Computational Cost (GFLOPs)
ResNet50	50	~25.6	~98	~3.8
Xception	71	~22.9	~88	~8.4
EfficientNet-B0	237 (including compound scaling of depth)	~5.3	~20	~0.39
VGG16	16	~138	~528	~15.5
Proposed ViT-based model	13	More than 87	More than 330	More than 17.6

Table 5
Hyperparameters details.

Hyperparameters of deep learning and proposed models	**Value**
Input shape	160 × 160 × 3
Output layer activation function	SoftMax
Dense layers activation function	ReLU, or GELU (Transformer encoder only)
Normalization parameters	Epsilon = 10^{-6} (Transformer encoder only)
Epochs	Less than 60 due to Early-Stopping with patience 9
Batch size	8
Optimizer	Rectified Adam
Learning rate	10^{-4}
learning rate decay factor	0.2
Patience	5
Validation split	10 % for both validation and test
Loss function	Categorical Cross Entropy (label smoothing = 0.2)

compared with transformers, they primarily focus on local or spatial features due to their convolutional nature. While deeper layers can capture more abstract representations, they might still miss global context as effectively as the Transformer's attention mechanism. Also, the convolutional filters in CNNs are fixed after training and might not adapt dynamically to different parts of the input as the attention mechanism does. CNNs do not inherently provide the same level of interpretability as attention mechanisms, making it harder to understand which parts of the image influenced the model's decisions. In order to extract additional dependencies or relations with distinct information from images utilizing various multiscale layers, we have adjusted the conventional ViT model by including parallel layers which improves the ViT performance.

Although ViT model complexity is due to the self-attention mechanism, which has a quadratic complexity in terms of the number of image patches and the number of layers (the model's depth) that increase the number of parameters, using frozen layers helps to overcome complexity problems. The architecture of our model's depth has overall, 13 transformer encoder layers, with the first six layers' parameters being frozen or fixed to prior knowledge to speed up training and reduce the overall number of trainable parameters from 87 M to 42 M. The dropout layer can be used in model configuration to avoid overfitting; here, it is used in the internal ViT model of the hidden transformer encoder layers after every dense layer. During implementation, adding it to the last FC layer gives less accuracy, so it can be dispensed with in the last FC layer because it might degrade the overall performance of the proposed ViT transformer model. The detailed architecture parameters of the proposed model are summarized in Table 3, which illustrates our model's complexity and the total number of parameters.

3. Experiments and results

The PC specifications and hardware platforms used in the experiments of the proposed model are Intel Core i7-11800H at 2.30 GHz, 16 GB of RAM, and NVIDIA GeForce RTX 3050 with 4 GB of GPU. The used software platform operating system was Windows 10 Pro × 64-bit, and the used framework was Keras Frame on the Tensor-Flow 2.10.1 backend (GPU support with CUDA toolkit = 11.2 and CUDNN = 8.1.0) of Python programming deep learning. There have been many experiments to train and evaluate the validity of the proposed approach. First, we train and test our model with the new dataset, which results from the two stages of preprocessing the raw dataset. Next, we compare our suggested evaluation results with the outcomes of training the traditional ViT model and four distinct well-known pretrained CNN models using our preprocessed dataset, as well as with state-of-the-art results obtained from using the same published dataset.

The deep-learning CNN networks used for comparison with the proposed model are ResNet50, Xception, Efficient-NetB0, and VGG16. These networks were pretrained on the Image Net dataset with 1000 different classes for the extraction of deep fundamental features, letting weights in all layers be updated during training (not frozen) to increase accuracy. The overall accuracy significantly decreases when attempting to train frozen deep learning model layers due to the big differences between the nature of the Image Net dataset and IR images, indicating the critical need for retraining all layers. ResNet50 has 50 layers total, which develop a network by stacking residual blocks on top of one another [16]. In order to connect the activation layer by skipping over intermediate levels, it implements the "Skip Connection" idea, found at the core of residual blocks. Efficient-NetB0 uses the compound scaling method to scale networks' dimensions (resolution, depth and width) equally in order to maximize their accuracy [39]. Xception is made up of 36 convolutional layers for feature extraction that are arranged as a linear stack of residual depth-wise separable convolution layers [5]. VGG16 consists of 3x3 convolution layers with the same padding on stride 1 and a max pool layer with a 2x2 filter on stride 2[35], then two completely connected layers and softmax for output are added after these layers. The detailed data of the compared deep learning models is shown in Table 4. For experiments, the appropriate fully connected or dense layers are added at the top of all models for classifying dataset PV anomalies.

Cross-validation techniques are essential in machine learning and

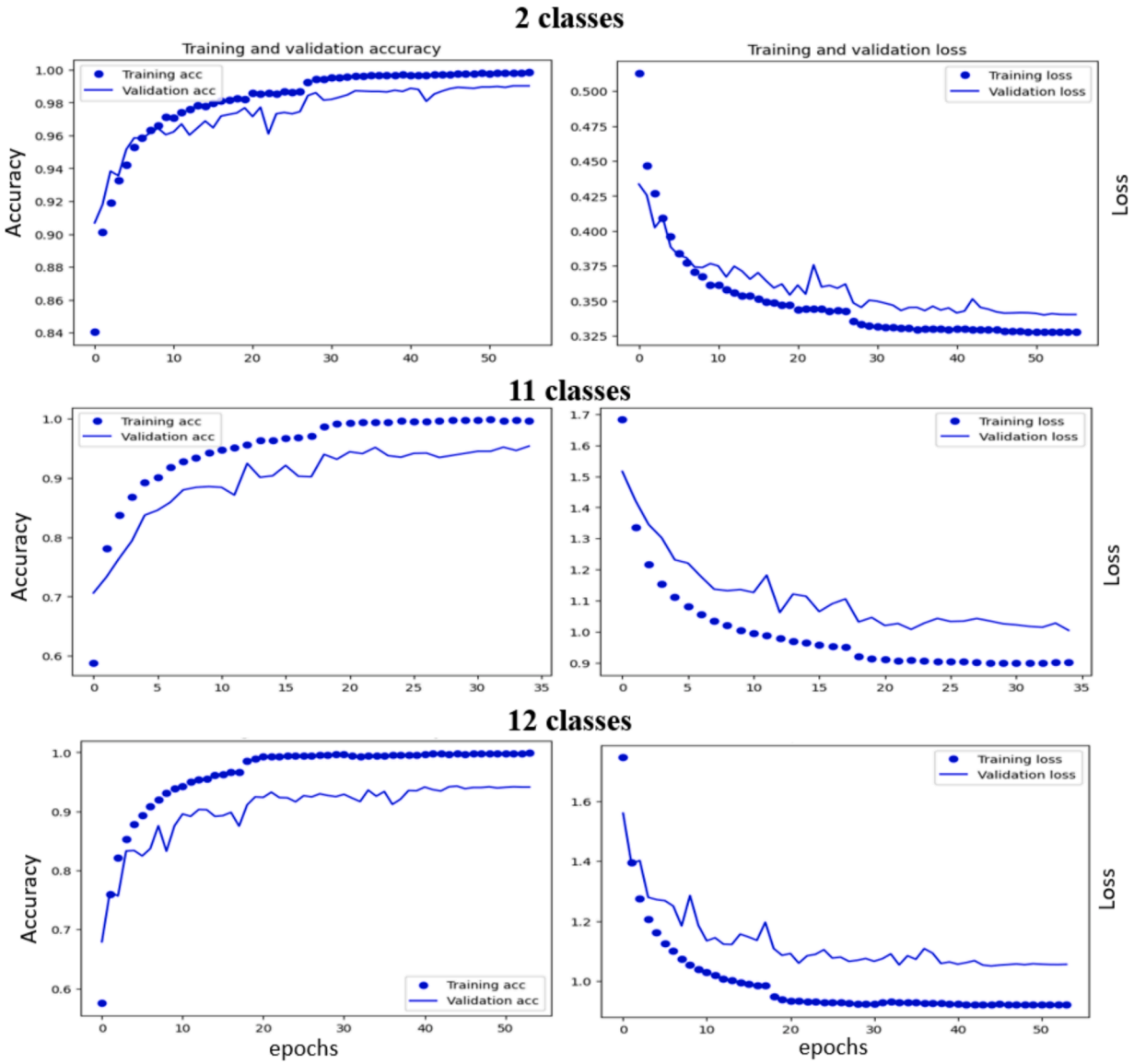


Fig. 6. Model performance (accuracy and loss) of classifying: 2 classes, 11 classes, and 12 classes.

Table 6
Testing metrics results for classifying the two classes.

Model	Accuracy	Precision	Recall	F1-score
ResNet50	0.9380	0.9400	0.9400	0.9400
VGG16	0.9465	0.9500	0.9500	0.9500
Efficient-NetB0	0.9620	0.9621	0.9620	0.9620
Xception	0.9677	0.9700	0.9700	0.9700
ViT-B32	0.9756	0.9750	0.9760	0.9755
Proposed model	**0.9823**	**0.9823**	**0.9823**	**0.9823**

Table 7
Testing metrics results for classifying the 11 classes.

Model	Accuracy	Precision	Recall	F1-score
ResNet50	0.8896	0.8900	0.8900	0.8900
VGG16	0.8948	0.9000	0.8900	0.8900
Xception	0.9212	0.9200	0.9200	0.9200
Efficient-NetB0	0.9299	0.9300	0.9300	0.9300
ViT-B32	0.9524	0.9500	0.9500	0.9500
Proposed model	**0.9619**	**0.9600**	**0.9600**	**0.9600**

Table 8
Testing metrics results for classifying the 12 classes.

Model	Accuracy	Precision	Recall	F1-score
ResNet50	0.8869	0.8900	0.8900	0.8900
VGG16	0.8917	0.8900	0.8900	0.8900
Xception	0.9123	0.9100	0.9100	0.9100
ViT-B32	0.9135	0.9200	0.9100	0.9100
Efficient-NetB0	0.9361	0.9400	0.9400	0.9400
Proposed model	**0.9555**	**0.9600**	**0.9600**	**0.9500**

data science to evaluate the performance of the proposed model and ensure its generalizability to unseen data. Here we use holdout validation for its simplicity by splitting the dataset into three groups: a training group that contains 80 % of the dataset and 10 % for the validation and test groups. Then we train our model on the training set while unbiasedly evaluating its performance on the validation set to help prevent overfitting. After that, the testing process is made using the unseen data of the holdout test set.

The dataset grows to about 40,000 images after augmentation; we utilize 20,000 images for the No-Anomaly class and the same number for the Anomaly class for classifying the two classes because this helps

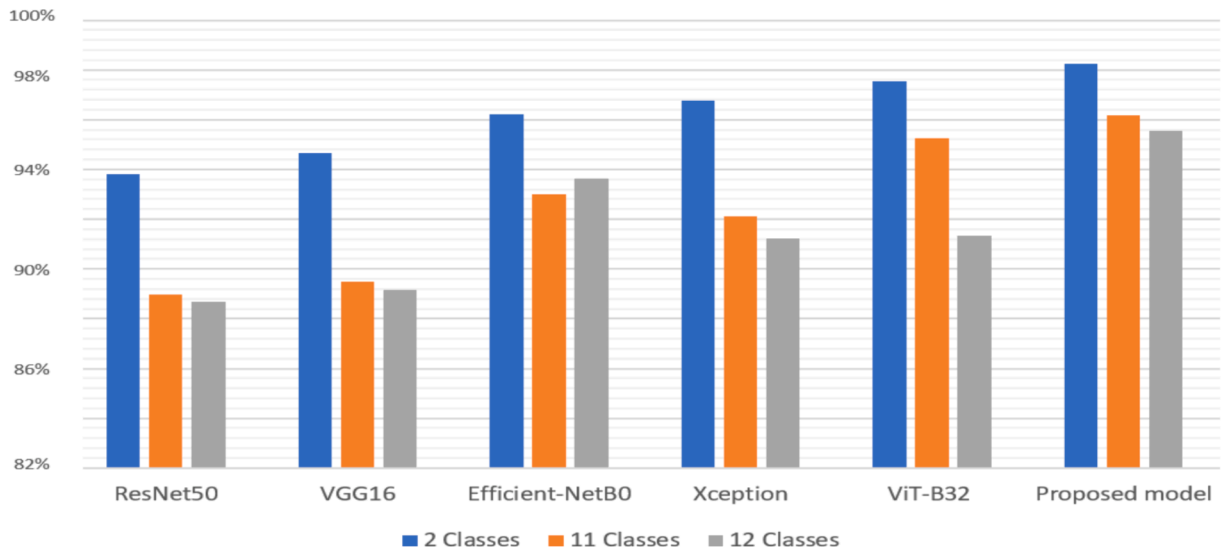


Fig. 7. Summary of the accuracy of the compared models.

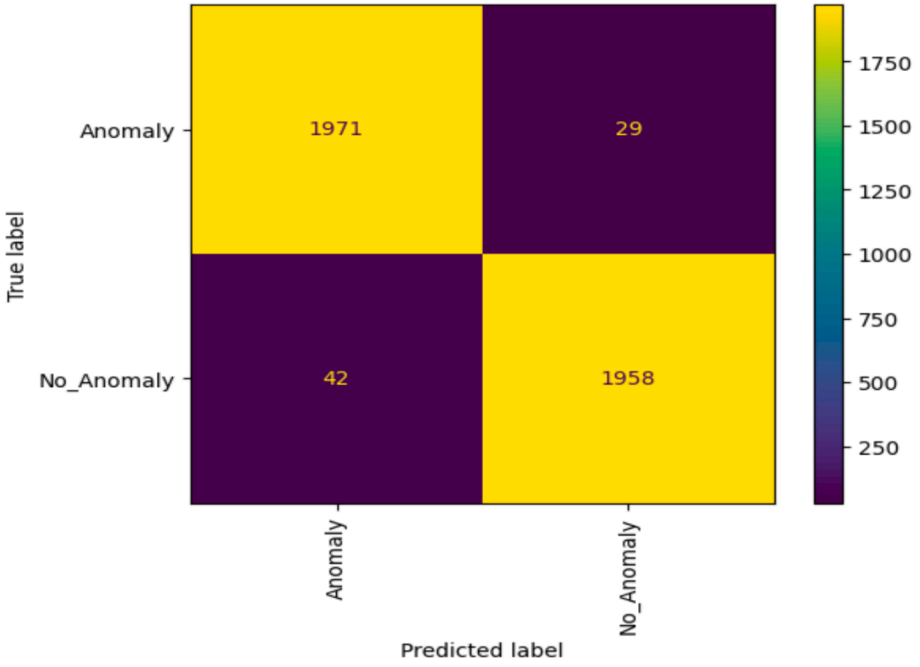


Fig. 8. Confusion matrix of the proposed model for classifying the two classes.

Table 9
Detailed testing metrics of the proposed method for classifying anomaly and no anomaly classes.

Class Name	*Precision*	*Recall*	*F1-score*
Anomaly	0.9791	0.9855	0.9823
No Anomaly	0.9854	0.9790	0.9822

maintain balance during the training phase. Additionally, when classifying eleven distinct types of PV anomalies using 2100 images per class, a total of 23,100 images are obtained; when classifying twelve classes, which include the No-Anomaly class with the eleven PV anomaly classes, a total of 25,200 images are obtained.

The images are resized to 160 × 160 × 3 pixels, and the training is run with batch sizes of 8 and approximately 50 epochs. We start training

Table 10
Detailed testing metrics of the proposed method for classifying 11 classes.

Class Name	*Precision*	*Recall*	*F1-score*
Cell	1.00	0.95	0.97
Cell Multi	0.95	0.94	0.94
Cracking	0.94	0.98	0.96
Diode	1.00	1.00	1.00
Diode Multi	1.00	0.98	0.99
Hot Spot	0.96	0.99	0.97
Hot Spot Multi	0.98	0.98	0.98
Offline Module	0.88	1.00	0.94
Shadowing	0.99	0.79	0.88
Soiling	0.98	1.00	0.99
Vegetation	0.94	0.98	0.96

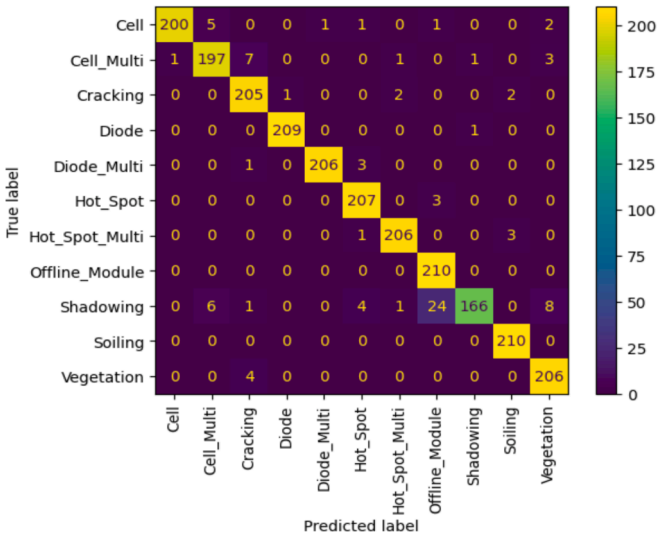


Fig. 9. Confusion matrix of the proposed model for classifying 11 classes.

with a learning rate value of $1e^{-4}$ and then reduce it by a factor of 0.2 every 5 consecutive unsuccessful epochs (i.e., when validation accuracy stops increasing), using "ReduceLROnPlateau" while training proceeds, and the training stops finally after 9 consecutive unsuccessful epochs, using "EarlyStopping" in the callbacks function of Keras, to avoid overfitting during learning in the three classification cases. The rectified Adam optimizer is used in optimizing our model, which adds an idea to correct the adaptive learning rate's variance, and reduce the chosen loss function which is the categorical cross entropy with label smoothing of 0.2 which acts as a form of regularization by preventing the model from becoming too confident and overly relying on certain training examples. All the hyperparameters are summarized in Table 5.

Although the training process may take a long time, according to model complexity and the platform specification, our methodology can be verified in real-time as we can capture and classify the required IR test image during PV system operation to help support maintenance and early fault detection. Our PC hardware experiments took an average of 690, 388, and 446 s per epoch in the training and validation phase for the classification of 32000 + 4000 images in two classes, 18480 + 2310 images in eleven classes and 20160 + 2520 images in twelve classes,

Table 11
Detailed testing metrics of the proposed method for classifying 12 classes.

Class Name	*Precision*	*Recall*	*F1-score*
Cell	0.96	0.95	0.95
Cell Multi	0.93	0.97	0.95
Cracking	0.96	0.96	0.96
Diode	0.96	0.99	0.97
Diode Multi	0.99	1.00	0.99
Hot Spot	0.97	1.00	0.98
Hot Spot Multi	0.95	1.00	0.98
No Anomaly	0.99	0.95	0.97
Offline Module	0.87	0.94	0.90
Shadowing	0.98	0.77	0.86
Soiling	0.97	0.99	0.98
Vegetation	0.96	0.97	0.97

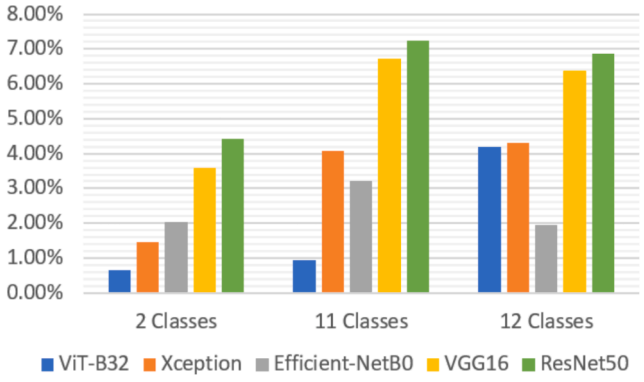


Fig. 11. The accuracy improvement of the proposed model versus the compared models.

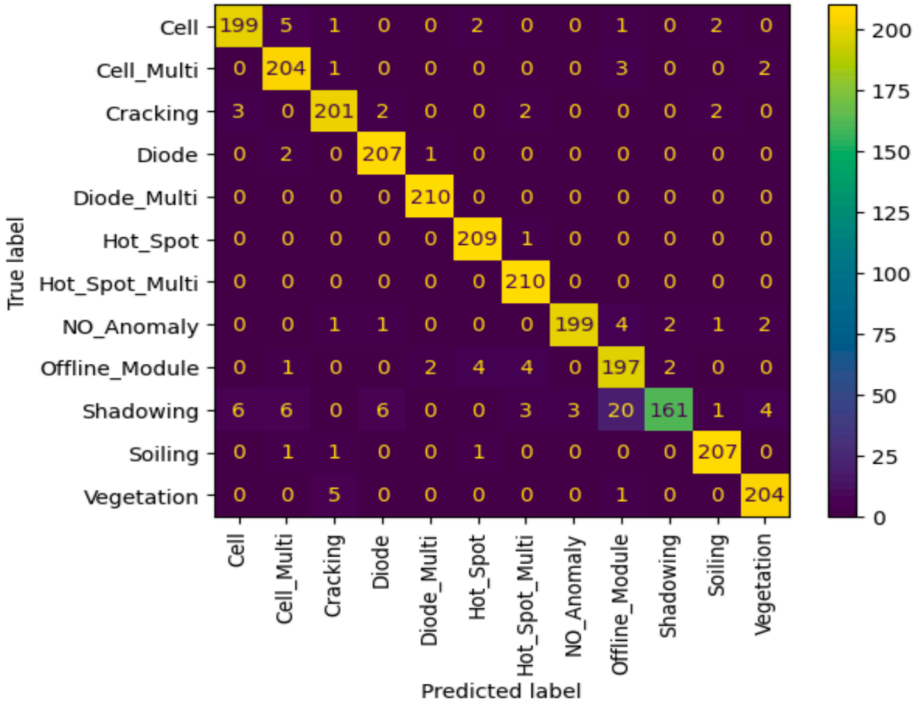


Fig. 10. Confusion matrix of the proposed model for classifying 12 classes.

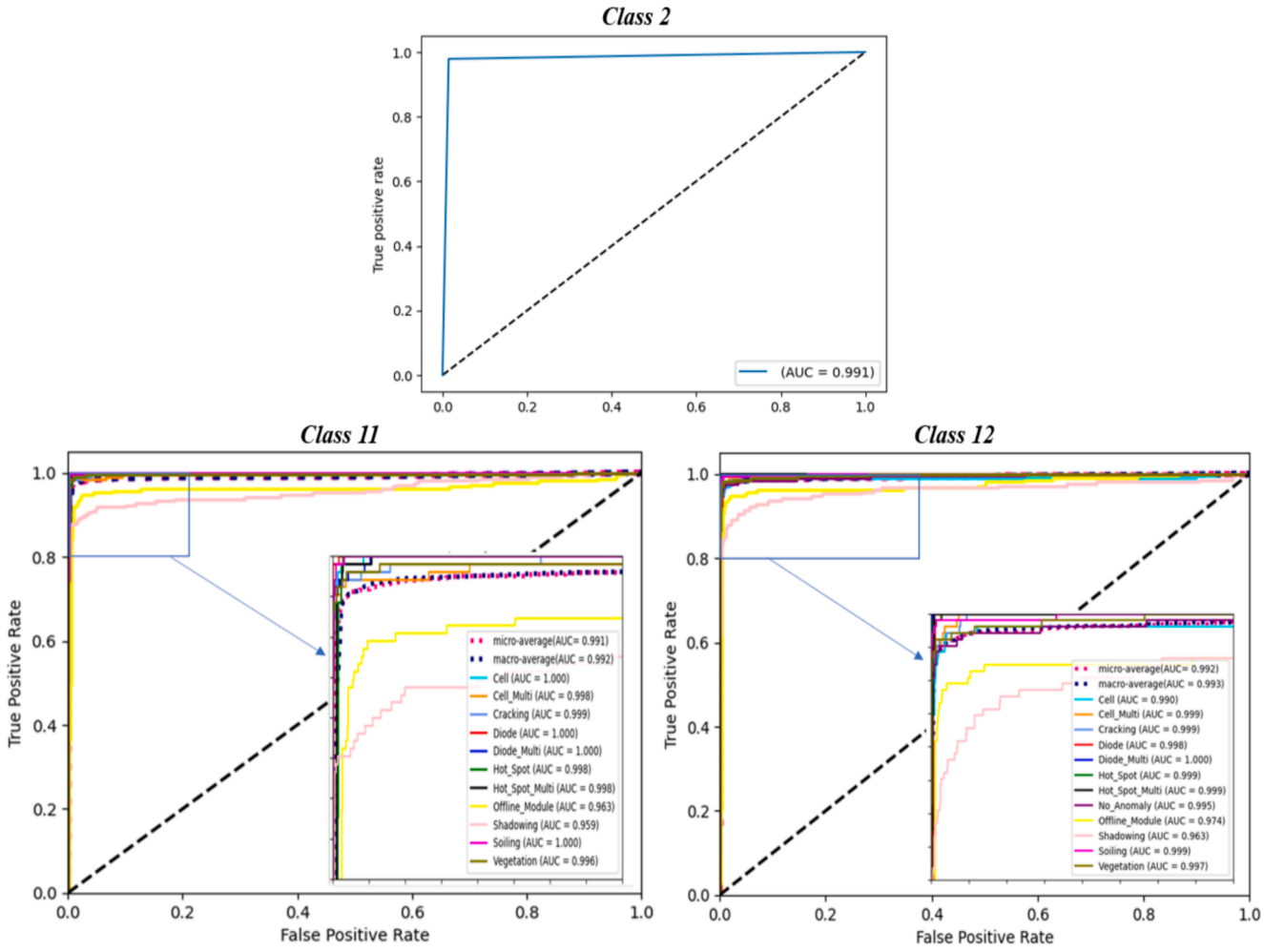


Fig. 12. ROC curves and AUC of the proposed model classification results.

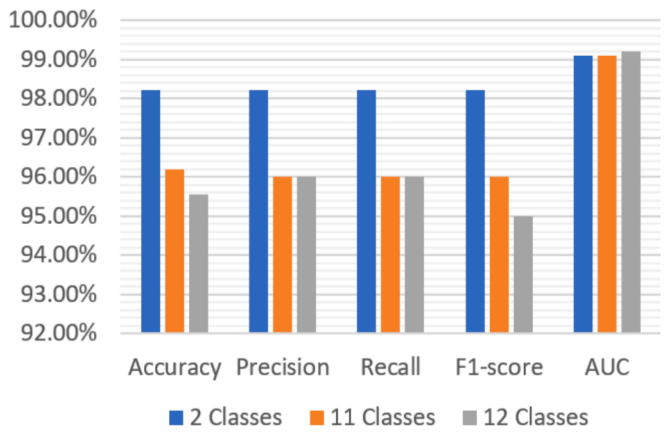


Fig. 13. Proposed method average statistics summery.

respectively. In the testing phase, it has an average time response of 6.5, 8.23, and 9.13 ms per image while processing and classifying 184, 122, and 110 images per second using VRAM of 4 GB of GPU.

Evaluation metrics are used to assess the performance of the proposed methodology. Accuracy, precision, recall (sensitivity), and F1-score metrics, along with the confusion matrix, are used to evaluate the classification errors and predictions of testing the proposed model. The accuracy metric measures how frequently a model for classification is overall right, as shown in Eq. (11), and precision measures how often a model correctly predicts the target class in Eq. (12), while recall in Eq. (13) demonstrates whether a model is able to locate every object in the target class. A substitute for accuracy metrics, F1-score in Eq. (14), is a model performance metric that equally weights precision and recall (harmonic mean).

$$Accuracy = \frac{True\,Positives + True\,Negatives}{True\,Positives + True\,Negatives + False\,Positives + False\,Negatives} \tag{11}$$

$$Precision = \frac{True\,Positives}{True\,Positives + False\,Positives} \tag{12}$$

$$Recall = \frac{True\,Positives}{True\,Positives + False\,Negatives} \tag{13}$$

$$F1\,score = \frac{2\,Precision * Recall}{Precision + Recall} \tag{14}$$

where, true positive means that the model classifies the anomaly class truly, and true negative identifies that the model classifies the opposite anomalies correctly. On the other hand, false positive means the model classifies the anomaly class falsely (incorrectly), and false negative defines the incorrect classification of the opposite anomalies.

To handle fault detection modeling using various AI techniques, training and validation procedures are created individually for each of the compared models. This shows how each model performs and guarantees the validity of the suggested model. Fig. 6 illustrates our training

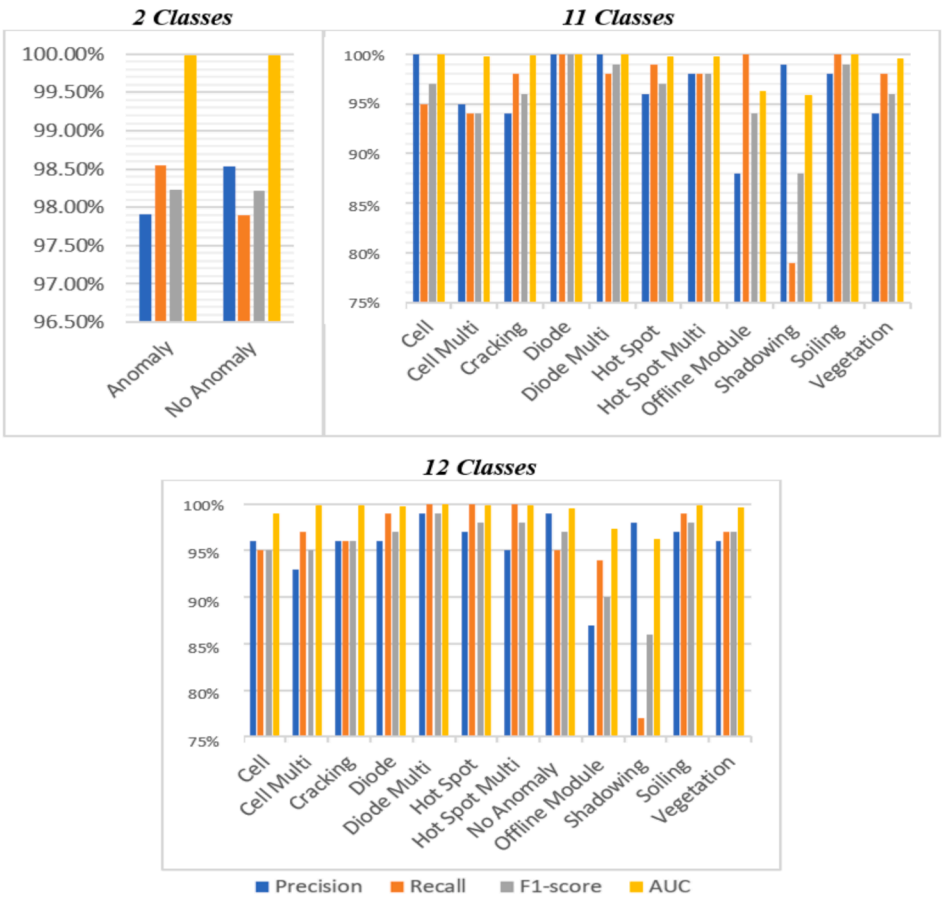


Fig. 14. Summary of statistics of the proposed method.

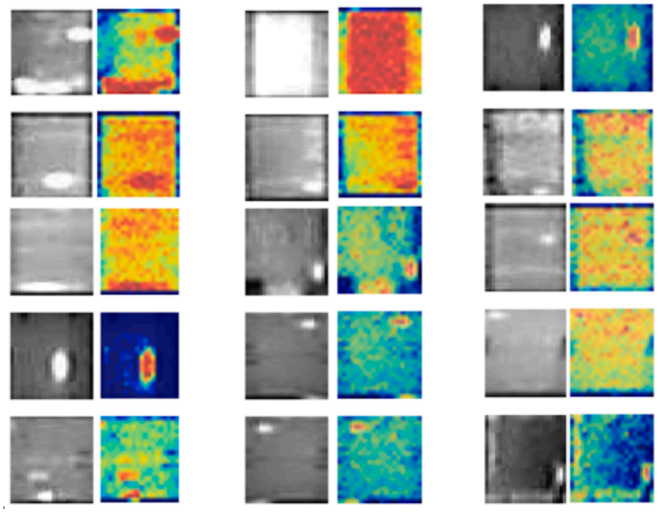


Fig. 15. Heat map visualization of the proposed model for some images produced from the parallel transformer layer.

and validation model's performance for the three classification cases. During training our model, the added parallel layers make the model more flexible in re-training the old transferred layers and also in training the new ones, which makes a remarkable improvement in the performance in the three cases, especially the third case.

The evaluating or testing processes for classifying data are made three times: the first one, which identifies two classes and produces metrics results displayed in Table 6, the second, which identifies 11 classes and produces results illustrated in Table 7, and the third, which identifies and classifies 12 classes and yields results summarized in Table 8.

The testing phase of our PV fault detection and classification system model reveals that the suggested method outperforms the compared deep learning models, which have an accuracy of 98.23 %, 96.19 %, and 95.55 % for the classification of two, eleven, and twelve classes, respectively, as shown in Fig. 7. Not only does the model we propose precede all CNN models, demonstrating that "attention is all you need"[42], but it also improves the classic ViT-B32 model in terms of precision, recall, and F1 score when compared to the results of the testing models, especially the 12 classes case.

The confusion matrix and testing metrics for classifying the 2 classes in Fig. 8 and Table 9 illustrate that the number of wrongly predicted images is only 71 from a total of 4000 test images, with a precision of 97.91 % and 98.54 % for the Anomaly and No Anomaly classes, respectively. The proposed methodology has the ability to distinguish between different and difficult eleven PV anomalies; as we can see in Table 10 and Fig. 9, the average precision is 96 % for detecting the 11 classes, and the best result is the Diode class, which has an F1-score of approximately 100 %, and the worst one is Shadowing with 88 %, which might be because the shadowing effect in IR PV module images covers major areas in those images, making a slight similarity with the offline module effect in images. Also, as we can see from the results of the 12 classes in Fig. 10 and Table 11, our model has the Diode Multi as the best class performance with an F1-score of 99 %, and the overall model performance with precision and recall of 96 % and the F1-score of 95 %.

The compared CNN models have the Efficient-NetB0 and Xception models as the second-best performers, and the worst is the ResNet50 model. According to the accuracy, the improvement of our model versus the ResNet50 CNN model was about 4.43 %, 7.23 %, and 6.86 %, and the

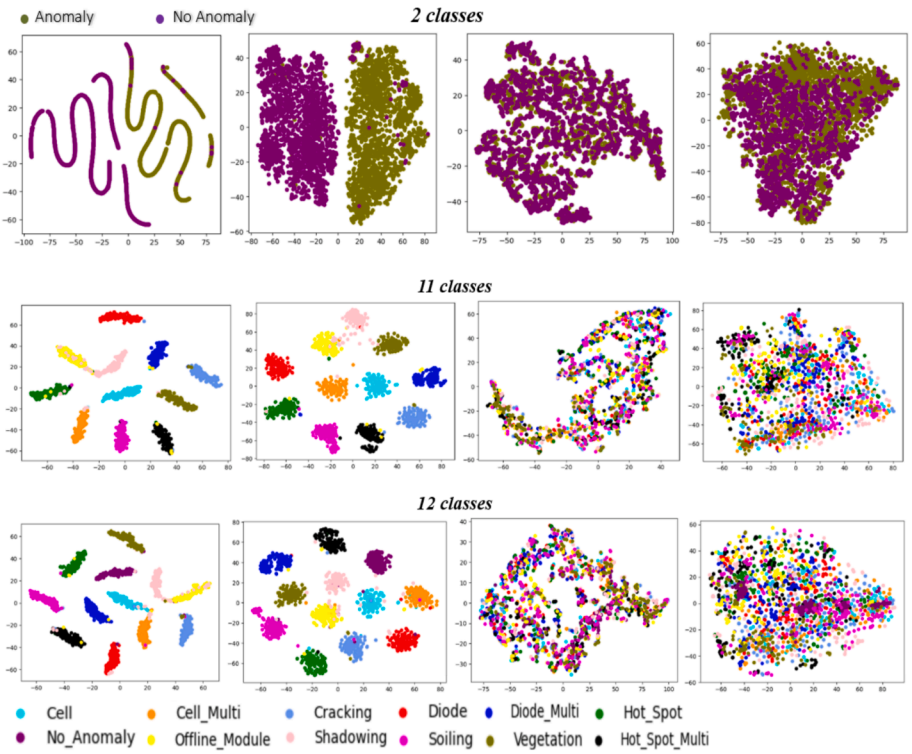


Fig. 16. The t-SNE diagram of the proposed model after (12 transformer layers, the parallel transformer layer, the concatenation layer, and the final fully connected layer).

Table 12
Comparison of the proposed model with previous models for the same dataset.

Model/Ref.	No. of Classes	Accuracy %	Precision %	Recall %	F1%
CNN [2]	2	92.50	92.00	92.00	92.00
	8	78.85	–	–	–
	11	66.43	–	–	–
Residual Ensemble [26]	2	94.40	–	–	–
	12	85.90			
Alex-Net Multiscale [21]	2	97.32	97.63	97.00	97.32
	11	93.51	93.52	93.51	93.49
CNN-Edge devices [25]	12	85.40	–	–	–
K-means & Inception & Residual [27]	8	89.00	72.00	70.00	69.00
Efficientb0 & NCA & SVM [10]	12	93.93	91.50	88.28	89.82
Proposed model	2	**98.23**	**98.23**	**98.23**	**98.23**
	11	**96.19**	**96.00**	**96.00**	**96.00**
	12	**95.55**	**96.00**	**96.00**	**95.00**

NCA: Neighborhood component analysis

progress from the traditional ViT model was about 0.67 %, 0.95 %, and 4.2 %, all for 2, 11, and 12 class cases, respectively, as shown in Fig. 11.

The Receiver Operating Characteristic Curve (ROC) and area under the ROC curve (AUC) are calculated as shown in Fig. 12 to measure the performance of the three classification cases, with an AUC average over 99 %, which indicates that the proposed model has a high degree of separability and discrimination. Figs. 13 and 14 summarize the proposed method statistics for all measured metrics for the three classification cases.

To measure the sensitivity of our model to the input images, Fig. 15 shows the Gradient-Class Activation Map (Grad-CAM) visualization or heat maps for some tested images, which are overlayed with the original images, showing the dependencies from the output classes of the parallel transformer layer.

For visualizing high-dimensional data as classes, the t-distributed stochastic neighbor embedding (t-SNE) method is used in Fig. 16 for the three classification cases, which shows that in the early layers, the clusters are not grouped correctly yet, and the error is reduced while the layers get deeper, and finally, in the last layer, the classes are correctly well separated.

We conducted a comparison with recent published literature that categorizes the same data set of the 11 classes of PV module defects for a more comprehensive evaluation of the proposed model. The compared models used different deep learning models, as illustrated in Table 12.

In order to assess the real-time inspection capabilities, the authors in [25] put the CNN model into practice on an edge device with an accuracy of 85.4 % for classifying the 12 classes. In [27], the CNN model with an inception module and a residual block is introduced to increase accuracy after refining the dataset using K-means clustering, which allowed the CNN model to classify eight faults with an average accuracy of 89 % and a precision of 72 %. Using the NCA feature selector, 17,000 features of the example Efficient-b0 model were chosen for the technique in [10], and then an SVM classifier was used to perform classification for the 12 classes, with an accuracy average of 93.93 %. Another potent technique is put out in [21] using a multi-scale CNN visual perception level kernel based on the transfer learning strategy, utilizing Alex-Net's pre-trained knowledge, which results in 97.32 % and 93.51 % accuracy for the 2 and 11 classes, respectively. All these results compared to ours ensure that our transformer model is competing strongly with the CNN models and other machine learning methods and giving the best performance results.

The remarkable results of our proposed FDD methodology can improve PV system operation and efficiency, which will help continuously detect failures and classify variant anomaly types, protect the system from risks, and reduce its power losses. The proposed system also aids in the maintenance of the PV system by making the right decision

with those faulty components according to anomaly type, like removing the damaged one as in cracks or just needing to be cleaned as in soiling.

4. Conclusion and Recommendations

This research presents a novel deep learning methodology that leverages data-driven methodologies for fault detection systems. The Infrared Solar Modules dataset of thermographic photos is used to identify PV system anomalies with the use of the transformer ANN approach. First, as a preprocessing stage, a sharpening filter is applied to the dataset. Next, oversampling augmentation is used to enlarge and balance the images. Finally, a ViT-based model is used to classify different PV anomalies, yielding excellent prediction results when compared to other strategies. Our suggested method proves that transformer accuracy can reach 98.23 % for two classes (Anomaly and No-Anomaly), 96.19 % for eleven PV faults, and 95.55 % for twelve classes (No-Anomaly and the eleven anomalies such as Cell, Cell-Multi, Shadowing, Hot-Spot, etc.). Applying our proposed model demonstrates that transformer models are competitive with CNN models, and a strong classification methodology for FDD systems can be obtained by extracting the key features from images through an attention mechanism only.

Researchers and the PV manufacturing sector can use this research effort as a unique reference to improve the possibilities for defect detection in solar PV systems. Recommendations for possible limitations of our method are to try the proposed methodology on different huge datasets with different capturing conditions (environmental or equipment configuration) and different methods like visual or electroluminescence images to improve its robustness and increase its stability and globality for classifying more fault classes. Also trying to handle the computational complexity of the model using different architectures or techniques, like distillation or pruning methods. For future work, using an online and real-time FDD system in the field with the aid of edge technology and the Internet of Things (IoT) would improve the PV system's performance. A hybrid approach that combines the privileges of transformer ANN and CNN approaches may be utilized to localize or categorize other PV faults to support maintenance and early detection efforts aimed at preventing damage to PV systems.

CRediT authorship contribution statement

E.A. Ramadan: Writing – review & editing, Supervision, Resources. **Nada M. Moawad:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Belal A. Abouzalm:** Writing – review & editing, Supervision, Project administration, Formal analysis. **Ali A. Sakr:** Supervision, Conceptualization. **Wessam F. Abouzaid:** Supervision, Formal analysis, Data curation, Conceptualization. **Ghada M. El-Banby:** Writing – review & editing, Visualization, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Infrared Solar Modules dataset is available online.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] Ali MU, Khan HF, Masud M, Kallu KD, Zafar A. A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography. Sol Energy 2020;208:643–51. https://doi.org/10.1016/j.solener.2020.08.027.
[2] Alves RHF, de Deus JGA, Marra EG, Lemos RP. Automatic fault classification in photovoltaic modules using Convolutional Neural Networks. Renew Energy 2021; 179:502–16. https://doi.org/10.1016/j.renene.2021.07.070.
[3] Arabshahi M, Torkaman H, Keyhani A. A method for hybrid extraction of single-diode model parameters of photovoltaics. Renew Energy 2020;158:236e252.
[4] Balasubramani G, Thangavelu V, Chinnusamy M, Subramaniam U, Padmanaban S, Mihet-Popa L. Infrared thermography based defects testing of solar photovoltaic panel with fuzzy rule-based evaluation. Energies 2020;13:1343. https://doi.org/10.3390/en13061343.
[5] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–8.
[6] Deepak S, Ameer PM. Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med 2019;111:103345. https://doi.org/10.1016/j.compbiomed.2019.103345.
[7] Deo BS, Pal M, Panigrahi PK, Pradhan A. An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification. medRxiv; 2022: 2022–11. doi: 10.1101/2022.11.13.22282266.
[8] Ding SX. Model-based Fault Diagnosis Techniques Design Schemes, Algorithms, and Tools ISBN 978-3-540-76303-1 e-ISBN 978-3-540-76304-8. doi: 10.1007/978-3-540-76304-8 Library of Congress Control Number: 2008921126, Springer-Verlag Berlin Heidelbe; 2008.
[9] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: transformers for image recognition at scale; 2020. doi: 10.48550/arXiv.2010.11929.
[10] Duranay ZB. Fault detection in solar energy systems: a deep learning approach. Electronics 2023;12(21):4397. https://doi.org/10.3390/electronics12214397.
[11] Dwivedi D, Babu KVSM, Yemula PK, Chakraborty P, Pal M. Identification of surface defects on solar pv panels and wind turbine blades using attention based deep learning model. Eng Appl Artif Intel 2024;131:107836. https://doi.org/10.1016/j.engappai.2023.107836.
[12] El-Banby GM, Moawad NM, Abouzalm BA, Abouzaid WF, Ramadan EA. Photovoltaic system fault detection techniques: a review. Neural Comput Appl 2023;1–14. https://doi.org/10.1007/s00521-023-09041-7.
[13] El-Rashidy MA. An efficient and portable solar cell defect detection system. Neural Comput Appl 2022;34:18497–509. https://doi.org/10.1007/s00521-022-07464-2.
[14] Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, et al. A survey on vision transformer. IEEE Trans Pattern Anal Mach Intell 2022;45(1):87–110. https://doi.org/10.1109/TPAMI.2022.3152247.
[15] Haque A, Bharath K, Khan M, Khan I, Jaffery Z. Fault diagnosis of photovoltaic modules. Energy Sci Eng 2019;7(3):622–44. https://doi.org/10.1002/ese3.255.
[16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
[17] Herraiz AH, Marugan AP, Marquez FPG. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renew Energy 2020;153:334e348. https://doi.org/10.1016/j.renene.2020.01.148.
[18] Hong YY, Pula RA. Methods of photovoltaic fault detection and classification: a review. Energy Rep 2022;8:5898–929. https://doi.org/10.1016/j.egyr.2022.04.043.
[19] Hong YY, Pula RA. Diagnosis of PV faults using digital twin and convolutional mixer with LoRa notification system. Energy Rep 2023;9:1963–76. https://doi.org/10.1016/j.egyr.2023.01.011.
[20] International Renewable Energy Agency's report; 2023. https://www.irena.org/Publications.
[21] Korkmaz D, Acikgoz H. An efficient fault classification method in solar photovoltaic modules using transfer learning and multi-scale convolutional neural network. Eng Appl Artif Intel 2022;113:104959. https://doi.org/10.1016/j.engappai.2022.104959.
[22] Koubiais E, Konstantopoulos GC. A combined drone and fixed-camera monitoring system for real-time photovoltaic inspection. J Renew Energy 2020;12(3):450–61. https://doi.org/10.1016/j.jre.2020.05.012.
[23] Kurukuru VSB, Blaabjerg F, Khan MA, Haque A. A novel fault classification approach for photovoltaic systems. Energies 2020;13(2):308. https://doi.org/10.3390/en13020308.
[24] Kusiak A, Li M. The prediction and diagnosis of wind turbine faults. Renew Energy 2010;36(1):9–19. https://doi.org/10.1016/j.renene.2010.05.014 (While this is about wind turbines, the principles of predictive maintenance are similar and can be adapted to PV systems.).
[25] Le M, Le D, Vu HHT. Thermal inspection of photovoltaic modules with deep convolutional neural networks on edge devices in AUV. Measurement 2023;218: 113135. https://doi.org/10.1016/j.measurement.2023.113135.
[26] Le M, Nguyen DK, Dao VD, Vu NH, Vu HHT. Remote anomaly detection and classification of solar photovoltaic modules based on deep neural network. Sustainable Energy Technol Assess 2021;48:101545. https://doi.org/10.1016/j.seta.2021.101545.
[27] Lee SH, Yan LC, Yang CS. LIRNet: a lightweight inception residual convolutional network for solar panel defect classification. Energies 2023;16(5):2112. https://doi.org/10.3390/en16052112.

[28] Li B, Delpha C, Diallo D, Migan-Dubois A. Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: a review. Renew Sustain Energy Rev 2021;138:110512. https://doi.org/10.1016/j.rser.2020.110512.
[29] Liu J, Guo F, Gao H, Huang Z, Zhang Y, Zhou H. Image classification method on class imbalance datasets using multi-scale CNN and two-stage transfer learning. Neural Comput Appl 2021;33:14179–97. https://doi.org/10.1007/s00521-021-06066-8.
[30] Manno D, Cipriani G, Ciulla G, Di Dio V, Guarino S, Lo Brano V. Deep learning strategies for automatic fault diagnosis in photovoltaic systems by thermographic images. Energ Conver Manage 2021;241:114315. https://doi.org/10.1016/j.enconman.2021.114315.
[31] Millendorf M, Obropta E, Vadhavkar N. Infrared solar module dataset for anomaly detection. In: The International Conference on Learning Representations (ICLR); 2020.
[32] Ning Y, Zhang S, Xi X, Guo J, Liu P, Zhang C. Cacemvt: Efficient coronary artery calcium segmentation with multi-scale vision transformers. In: 2021 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE; 2021. p. 1462–7.
[33] Rinaldi S, Ferrero R, Ponci F. Predictive maintenance strategy for photovoltaic systems. Renew Energy 2020;155:319–30. https://doi.org/10.1016/j.renene.2020.03.127.
[34] Silva B, Sousa JJ, Cunha A. Detecting earthquakes in SAR interferogram with vision transformer. In: IGARSS 2022–2022 IEEE international geoscience and remote sensing symposium. IEEE; 2022. p. 739–42. https://doi.org/10.1109/IGARSS46834.2022.9883523.
[35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition; 2014. arXiv preprint arXiv:1409.1556.
[36] Singha Deo B, Pal M, Panigrahi PK, Pradhan A. Supremacy of attention-based convolution neural network in classification of oral cancer using histopathological images. medRxiv; 2022:2022-11. doi: 10.1101/2022.11.13.22282265.
[37] Tahar KN, Ahmad A, Mohd Saman M. Assessing the use of aerial imagery and aerial mapping for solar photovoltaic farm planning and monitoring. Renew Sustain Energy Rev 2017;77:433–8. https://doi.org/10.1016/j.rser.2017.03.132.
[38] Talaat M, Elkholy MH, Alblawi A, et al. Artificial intelligence applications for microgrids integration and management of hybrid renewable energy sources. Artif Intell Rev 2023;56:10557–611. https://doi.org/10.1007/s10462-023-10410-w.
[39] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR; 2019, p. 6105–14. arXiv preprint arXiv:1905.11946.
[40] Usman M, Zia T, Tariq A. Analyzing transfer learning of vision transformers for interpreting chest radiography. J Digit Imaging 2022;35(6):1445–62.
[41] Valavanis KP, Vachtsevanos GJ. Handbook of unmanned aerial vehicles. Springer; 2015.
[42] Vaswani A, Shazeer NM, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need, 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA; NIPS 2017. ArXiv abs/1706.03762, doi: 10.48550/arXiv.1706.03762.
[43] Veerasamy V, et al. LSTM recurrent neural network classifier for high impedance fault detection in solar PV integrated power system. IEEE Access 2021;9:32672–87. https://doi.org/10.1109/ACCESS.2021.3060800.
[44] Wang J, Zhou J, Chen X. Data-driven fault detection and reasoning for industrial monitoring. Springer Nature, third edition 2022 (eBook); 2022, p. 264. doi: 10.1007/978-981-16-8044-1.
[45] Wang W, Lee HY. Enhancing real-time fault detection in photovoltaic systems using hybrid monitoring methods. IEEE Trans Sustainable Energy 2021;12(4):2165–73. https://doi.org/10.1109/TSTE.2021.3058972.
[46] Xu Y, Wei H, Lin M, et al. Transformers in computational visual media: a survey. Comp Visual Media 2022;8:33–62. https://doi.org/10.1007/s41095-021-0247-3.
[47] Zeng C, Shi H, Wu W, Li H, Liang W. A real-time fault monitoring system for large-scale photovoltaic power plants based on wireless communication. Renew Energy 2019;136:640–9. https://doi.org/10.1016/j.renene.2019.01.065.
[48] Zhang D, Liu X, Chen M. Real-time data acquisition and monitoring system for photovoltaic arrays. IEEE Access 2018;6:70116–23. https://doi.org/10.1109/ACCESS.2018.2880342.
[49] Zhang W, Wang J, Ma H, Zhang Q, Fan S. A transformer-based approach for metal 3d printing quality recognition. In: 2022 IEEE international conference on multimedia and expo workshops (ICMEW). IEEE; 2022. p. 1–4.
[50] Zhou Y, Wang J, Han T, Cai X. Fire smoke detection based on vision transformer. In: 2022 4th international conference on natural language processing (ICNLP). IEEE; 2022. p. 39–43.