Pertanika J. Sci. & Technol. 34 (4): 2487 - 2518 (2026)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Photovoltaic Module Defect Classification Analysis Based on Convolutional Neural Network Architecture with Augmentation Dataset

Nurul Atikah Mazlan[1], Shahrani Shahbudin[1]*, and Murizah Kassim[1,2]

[1]*Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[2]*Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

ABSTRACT

As power producers increasingly adopted renewable energy sources, solar energy emerged as a key solution due to its sustainability and efficient energy harvesting. Photovoltaic (PV) modules convert solar irradiance into electricity, but defects in these modules significantly impact power generation efficiency. This study aims to classify PV defect images through deep learning, focusing on Convolutional Neural Networks (CNNs). It explores two different transfer learning architectures: Series Networks, including AlexNet, VGG-16, and VGG-19, and Directed Acyclic Graph (DAG) Networks, such as ResNet-18, ResNet-50, Inception-V3, and GoogLeNet. Series Networks, known for their structured, layer-wise feature extraction, performed well in specificity and precision; meanwhile, DAG Networks, with deeper architectures and residual connections, enhanced feature learning and classification accuracy. The study evaluated augmented and non-augmented datasets to determine the most effective approach. Results showed that the augmented dataset consistently outperformed the non-augmented dataset. Among the architectures, ResNet-50, a DAG Network, achieved the highest classification accuracy (98.96%) and F1 score (97.89%) due to its ability to retain critical features across multiple layers. Meanwhile, AlexNet excelled in specificity (99.53%), and VGG-16 demonstrated superior precision (98.65%), highlighting the strengths of Series Networks in structured, less complex classification tasks.

Keywords: Augmentation dataset, convolutional neural network (CNN), photovoltaic (PV) module defect images

ARTICLE INFO

Article history:
Received: 13 February 2025
Accepted: 29 September 2025
Published: 13 August 2026

DOI: https://doi.org/10.47836/pjst.34.4.26

E-mail addresses:
tikahmazlan01@gmail.com (Nurul Atikah Mazlan)
shahranis@uitm.edu.my (Shahrani Shahbudin)
murizah@uitm.edu.my (Murizah Kassim)
* Corresponding author

e-ISSN: 2231-8526

INTRODUCTION

The usage of renewable energy rather than fossil fuels for the generation of electricity must be taken into consideration due to worries about pollution that causes climate change. This makes solar energy an important and more sought-after power source. Solar power systems, also known as solar energy, are among the most widely used renewable energy sources due to their eco-friendly nature (Hussein et al., 2021). These systems comprise several key components designed to efficiently capture and convert sunlight into electricity. At the heart of the system are photovoltaic (PV) panels, which generate electricity by absorbing sunlight. However, maintaining the efficiency of PV modules is crucial to ensuring optimal power output. Therefore, it is essential for engineers or responsible personnel to regularly inspect and maintain the PV modules at the solar farm (Li et al., 2020). Regular inspection is critical, as human intervention may sometimes overlook defects.

PV systems generate electricity by harnessing sunlight through a series of interconnected PV panels, utilising the photoelectric effect. Each panel consists of multiple PV modules, which are further composed of an array of PV cells. These cells are essential for converting solar energy into electricity. They absorb sunlight and transform it into usable electrical power, enabling the efficient generation of renewable energy (Balani et al., 2021). Monitoring photovoltaic plants and optimising their energy is crucial to ensure these facilities can effectively contribute to meeting the growing electricity demand. While knowing the average power output is important, it is insufficient for assessing the overall performance of a PV plant or detecting potential failures in specific components, such as individual cells or clusters of cells.

Numerous photovoltaic (PV) technologies, such as silicon heterojunctions, perovskites, organic–inorganic hybrids, and dye-sensitised solar cells (DSSCs), have been the subject of intensive research due to the desire for affordable and scalable solar cells. High efficiencies have been quickly attained by perovskite devices, and current efforts are concentrated on improving their long-term stability in ambient conditions. Within silicon-based heterojunctions, the integration of metal–oxide semiconductors onto p-type crystalline silicon (p-Si) has emerged as a simple, low-temperature fabrication route, with SnO_2 offering advantages such as a wide band gap, high carrier mobility, stability, and non-toxicity. Although various deposition techniques have been applied, hydrothermal growth provides a particularly uniform, low-cost, and controllable approach. However, reported SnO_2-based solar cells often involve complex fabrication and lack systematic evaluation of thickness effects. To address this gap, the present study employs hydrothermal deposition to fabricate SnO_2/Si heterojunction solar cells and investigates the influence of SnO_2 layer thickness on charge transport and light-harvesting performance (Özel et al., 2022).

The choice of photovoltaic technology is a key factor influencing both the efficiency and reliability of solar energy systems, and advanced concepts such as dye-sensitised solar cells

(DSSCs) attract considerable interest because of their low cost, easy fabrication and flexible structure. Doping plasmonic nanoparticles (Au, Ag) in (DSSC) has been demonstrated to improve the following phenomena: light absorption, electron injection, and diminish charge recombination at TiO_2/dye/electrolyte. The author in Salimi et al. (2019) identified that plasmonic nanocomposites, including Au@TiO_2 and Ag@TiO_2, can significantly boost power conversion efficiency (PCE), with mesoporous Ag–Au@TiO_2 nanocomposites that offering a cost-effective approach for future PV applications.

All-inorganic perovskite solar cells (PSCs) face challenges of charge carrier recombination that limit their efficiency, according to Faheem et al. (2022). A synergistic strategy using liquid-exfoliated MoS_2 at the $CsPbIBr_2$/HTL interface improves perovskite grain growth, reduces trap density, and stabilises the interface. As a result, the PCE increased from 7.38% to 10.98%, demonstrating a promising route for more efficient and stable PSCs. Various techniques for extracting natural dyes from cornelian cherries (Cornus mas L.) were studied to enhance the performance of dye-sensitised solar cells (DSSCs). The best results were obtained from citric acid-based solvent extraction, which increased short-circuit current and light absorption to increase power conversion efficiency from 0.47% to 0.98%. This demonstrates how bio-based dyes and appropriate extraction techniques can increase the efficiency of DSSCs (Kocak et al., 2019).

A low-cost and simple synthesis route was developed to fabricate carbon-based counter electrodes (CEs) as alternatives to conventional Pt-based ones in DSSCs. The homemade carbon paste CE (H-CE) exhibited a porous structure that enhanced electrolyte diffusion, reduced charge transfer resistance, and improved catalytic activity compared to commercial carbon paste CE (C-CE). As a result, the device achieved a power conversion efficiency of 2.70% and a high fill factor of 73%, surpassing both C-CE and Pt-based electrodes (Altinkaya et al., 2020).

A malfunction in any PV cell can result in reduced power generation, leading to significant losses in output. If a cell is shaded or malfunctioning, the output power decreases (Rao et al., 2021). Not all PV panels can generate output power normally, as faults can happen on the panel because of many factors. According to Nylund and Barbari (2019), a study on defects detected on PV modules obtained by visual inspection, such as cracks, snail trail, degradation, hotspots, delamination and others. Cracks and degradation may happen from the manufacturing process, which can lead to failures in the module. The existence of cracks on a module can contribute to power losses compared to normal PV modules.

Testing each individually for solar farms with thousands of PV panels can be time-consuming and cumbersome. Maintaining the quality and reliability of these panels is crucial for efficient power generation, long-term durability, and maximising returns on investment. Instead of manual testing, it would be highly beneficial to have real-time information on the energy output of each panel in the plant. This would enable quick identification of underperforming panels, allowing for timely repairs or replacements. Achieving this requires individual monitoring of each PV panel and transmitting the data to the control system. Given the high costs associated with solar-generated electricity, optimising system performance, energy production, and reliability is particularly important.

System failures in PV modules can occur due to various defects that impact their efficiency. According to Nylund and Barbari (2019), PV module failures typically progress through three stages: infant failure, midlife failure, and wear-out failure, all of which reduce efficiency. Common issues such as partial shading, hotspots, and cracking can further compromise performance. In this study, defects are categorised into fractures, delamination, encapsulation discolouration, glass breakage, and gridline corrosion. Cracks may develop during handling or manufacturing processes like soldering, while delamination often results from low-quality materials or improper processing. When the encapsulating material yellows, the PV module's ability to transmit light is diminished. Improper clamp geometry can result in glass breakage, which may lead to hotspot issues. The PV module may overheat and stop producing electricity because of the temperature increase. According to Li et al. (2020), gridline corrosion can be triggered by environmental factors such as heavy snow, which causes the degradation of components over time.

Quickly identifying and resolving faults is crucial for maintaining the efficiency of photovoltaic (PV) systems. Commercial power plants prioritise maximising energy production, minimising losses and maintenance costs, and ensuring safe operation (Appiah et al., 2019; Basnet et al., 2020). Since PV systems are prone to faults, early detection is key to enhancing their safety, performance, and longevity. To strengthen protection measures, the U.S. National Electric Code requires the installation of Overcurrent Protection Devices (OCPDs) and Ground Fault Detection Interrupters (GFDIs) in PV systems (Pierdicca et al., 2018).

One of the aspects that must be stressed to ensure that problems are found as quickly as possible is the inspection process. These problems may be resolved with the use of technological advancements that can reduce the need for manpower. Deep learning algorithms are a branch of artificial intelligence designed to simulate human brain function and decision-making (Meng et al., 2023). The neural network architecture used in deep learning models has hidden layers for a variety of purposes, including feature extraction. This technique was largely used in current PV module inspection methods to categorise PV module defects. Convolutional Neural Networks (CNNs) are the most widely utilised deep learning algorithm for image classification. This is due to CNN's ability to extract patterns, classify, and even detect defects in images of PV modules (Henrique et al., 2021). CNN, also known as a feedforward network, processes input data through convolutional and pooling layers before passing it to fully connected layers for image classification (Cipriani et al., 2020). CNNs are extensively applied in computer vision and pattern recognition tasks such as image classification, object detection, facial recognition, natural language processing, and style transfer (Zyout & Oatawneh, 2020). In the field of photovoltaics, CNNs have been employed as classifiers to detect defects in PV modules, with various architectures trained to achieve high accuracy in defect identification. Based on Azis et al. (2022), DenseNet-201 was utilised by the researcher for the classification of images. The results show that DenseNet-201 outperforms other CNN architectures, such as DCNN and VGG-16. Additionally, Zyout and Oatawneh (2020) utilised a modified version of the AlexNet-based CNN architecture and gave outperformed results. This proves that CNN performs better than other deep learning methods.

Transfer learning is a deep learning technique that leverages a pre-trained model from one task as a foundation for solving a related task (Salih et al., 2023). It is typically significantly quicker and simpler than training a network from the start for updating and retraining. Application areas where the method is frequently utilised include object identification, picture recognition, and speech recognition. Transfer learning has two types, which are transfer learning with a series network and transfer learning with a DAG network. A series network is a type of deep learning neural network where layers are arranged sequentially, with a single input layer and a single output layer. Examples of series networks include AlexNet, VGG-16, and VGG-19. In contrast, a Directed Acyclic Graph (DAG) network has a more complex structure, allowing layers to receive inputs from multiple layers and pass data to others. This design enables greater flexibility and deeper architectures. Examples of DAG networks include GoogLeNet, ResNet-50, Inception-V3, and ResNet-101 (Yu et al., 2023).

The model might experience overfitting issues due to the limited and imbalanced number of images in the training dataset. Overfitting of the model happens when a model can learn the training dataset well, but not for a new dataset (Pavlitskaya et al., 2022). The technique known as augmentation is used on the small dataset to increase the network architecture's accuracy (Zulkarnain et al., 2023). According to Bakhshayesh et al. (2022), the augmentation of the dataset is the process of increasing the number of images by applying the augmentation technique. This is because overfitting issues experienced by machine learning can compromise the model's accuracy in the classification task. The augmented dataset will transform such geometric transformations that are usually used in image augmentation. Some types of geometric transformation, including annotation, blurring, flipping, and rotating, are applied during the augmentation process (Ahmad et al., 2022; Zulkarnain et al., 2023).

Many researchers have explored deep learning-based methods for PV defect image classification. However, most studies have only compared their models against one or two other architectures (Rahman et al., 2019), lacking a comprehensive evaluation of PV classification using Convolutional Neural Networks (CNNs) for both Directed Acyclic

Graph (DAG) and Series Network architectures. The objective of this study is to determine the most effective CNN architecture for classifying PV defect images using both augmented and non-augmented datasets within DAG and Series Network frameworks, to assess CNN performance in PV module defect classification using key performance metrics, and to determine the architecture that achieves the best overall classification performance.

LITERATURE REVIEW

Several previous studies have been reviewed to analyse the methods applied for classifying defects in PV modules. Table 1 summarises these works according to the deep learning networks employed for PV defect classification.

In Tella et al. (2025), an ensemble-based deep learning framework was applied for defect detection using the ELPV dataset of 2,624 EL images. The ensemble combined eight models, which are AlexNet, SENet, GoogLeNet, Xception, Vision Transformer (ViT), Darknet53, ResNet18, and SqueezeNet, to improve prediction accuracy. The ensemble achieved accuracies of 68.36% and 68.31%, outperforming a previous hybrid model (61.15%). However, the study also showed that individual models can sometimes perform better, with ResNet18 achieving 73.02% accuracy in a binary classification task.

The authors in Drir et al. (2024) proposed an ensemble CNN that employed weighted feature fusion by integrating three architectures: VGG-16, ResNet, and MobileNet. A comparative analysis with the individual baseline models revealed that the ensemble

Table 1
PV defect classification method

Reference	Method Used	Result
Tella et al., 2025	AlexNet SENet GoogLeNet (Inception V1) Xception Vision Transformer (ViT) Darknet53 ResNet18 SqueezeNet	A 2624 sample of EL images was classified based on their defect by using 8 ensemble models. ResNet-18 obtained the best accuracy result with 73.02%, demonstrating that the network outperforms the ensemble under specific conditions.
Drir et al., 2025	VGG-16 ResNet MobileNet	12 PV defect images were identified by using the new model, which combined deep features from the CNN architecture's layer that was involved in this study. The ensemble CNN achieved an average accuracy of 96% compared to individual CNN models.
Kellil et al., 2023	DCNN VGG-16	Fine-tuned VGG-16 outperformed in terms of accuracy compared to DCNN for both binary-class and multiclass analysis of PV defect images
Azis et al., 2022	DenseNet-201 VGG-16	DenseNet-201 outperformed in performance metrics compared to VGG-16

achieved superior performance, attaining an average accuracy of 96%. This approach demonstrated higher accuracy and precision in classifying 12 types of PV defects from thermographic images.

In Kellil et al. (2023), both binary and multiclass classifications of PV defect images were analysed using a small DCNN and a fine-tuned VGG-16 model. The results indicated that the fine-tuned VGG-16 outperformed the small DCNN, achieving higher accuracy in both classification tasks. The limitations of ResNet in image classification, particularly due to the identity shortcut, were highlighted in Azis et al. (2022), where the authors introduced a more efficient architecture, DenseNet-201. The identity shortcut was found to cause a collapsing domain problem, reducing the network's learning capacity. To address this, the study employed DenseNet-201 and VGG-16 for PV defect classification. The results demonstrated that DenseNet-201 achieved superior accuracy and overall performance metrics compared to VGG-16.

Meanwhile, in this work, CNN was selected as the classifier for PV defect classification, with seven architectures applied through transfer learning, covering both Series and DAG Networks. The Series Networks included AlexNet, VGG-16, and VGG-19, while the DAG Networks consisted of ResNet-18, ResNet-50, GoogLeNet, and Inception-V3. Unlike ensemble approaches, such as the one reported in Drir et al. (2024), which achieved high accuracy but required greater complexity and computational cost, this work focused on standard CNN architectures that are simpler and less resource-intensive. Previous studies have also utilised standard CNNs; however, most were limited to only two or three models, offering less comprehensive evaluation.

METHODOLOGY

Figure 1 illustrates the workflow of the proposed approach, which is structured into three phases. The initial phase entails gathering PV defect images from the solar farm, covering five types of defects: glass breakage, delamination, encapsulation discolouration, cracks, and gridline corrosion. Next, the dataset is split into two parts: one with augmentation and one without augmentation. In the final phase, the images are classified using a Convolutional Neural Network (CNN) with seven different architectures: AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, GoogLeNet, and Inception-V3. The performance of each

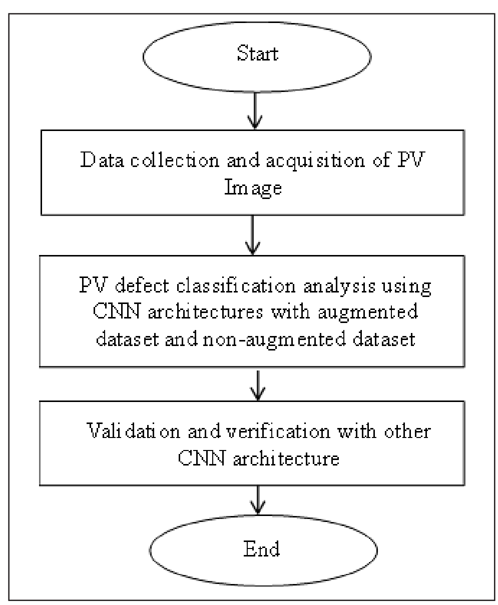

Figure 1. System flowchart

architecture is then evaluated using performance metrics for both the augmented and non-augmented datasets.

Data Collection and Acquisition of PV Image

According to Nylund and Barbari (2019), many defect occur in PV modules have been analysed. Classification of PV defect images in this work only covered 5 types of defects, which are cracks, delamination, glass breakage, encapsulation discolouring, and gridline corrosion. The type of images used in this work is RGB, which consists of various sizes and angles. The dataset comprises 1,394 images of varying sizes. Figure 2 displays PV defect images collected from a solar farm, categorised into five classes. The dataset includes 76 images of cracks, 109 images of delamination, 114 images of encapsulation discolouration, 64 images of gridline corrosion, and 1,031 images of glass breakage.

Due to the various sizes of images, it resizes the input image for network requirements and has a specific input size to train the model. This research uses CNN networks such as AlexNet, VGG, ResNet, GoogLeNet, and Inception-V3. Every network has its image input size. VGG (VGG-16 and VGG-19), ResNet (ResNet-18 and ResNet-50), and GoogLeNet have the same image input size, which requires images to be in 224-by-224 dimensions (Pangilinan et al., 2022). Alexnet requires 227 × 227 pixels of input size, and Inception-V3 requires 299 × 299 pixels of image input size (Anilkumar et al., 2021).

The training dataset does not have the same quantity for every class. Data augmentation is applied for original dataset as the number of images in every class is imbalanced and to reduce overfitting of the model. The distribution of classes, however, varies significantly; where 64 images displaying gridline corrosion and 1,031 images showing glass breakage. Since the dataset was gathered from a single location, this may have reduced its diversity in terms of lighting or the severity of defects. Future work should therefore focus on developing a larger and more diverse dataset that incorporates multi-site data, varied environmental conditions, and a more balanced distribution across defect types.

The dataset is divided into 70% for training and 30% for validation, yielding 976 training images and 418 validation images. To enhance the dataset, augmentation techniques such as rotation and X, Y translation are applied. Figure 3 displays the images after the augmentation process.

PV Defect Classification using Convolutional Neural Network (CNN)

Transfer learning is classified into two distinct types: transfer learning using a series network and transfer learning using a Directed Acyclic Graph (DAG) network. A series network is a deep learning model where layers are structured in sequential order, consisting of one input layer and one output layer. Examples of series networks include AlexNet, VGG-16, and VGG-19. On the other hand, a DAG network features a more complex design, enabling

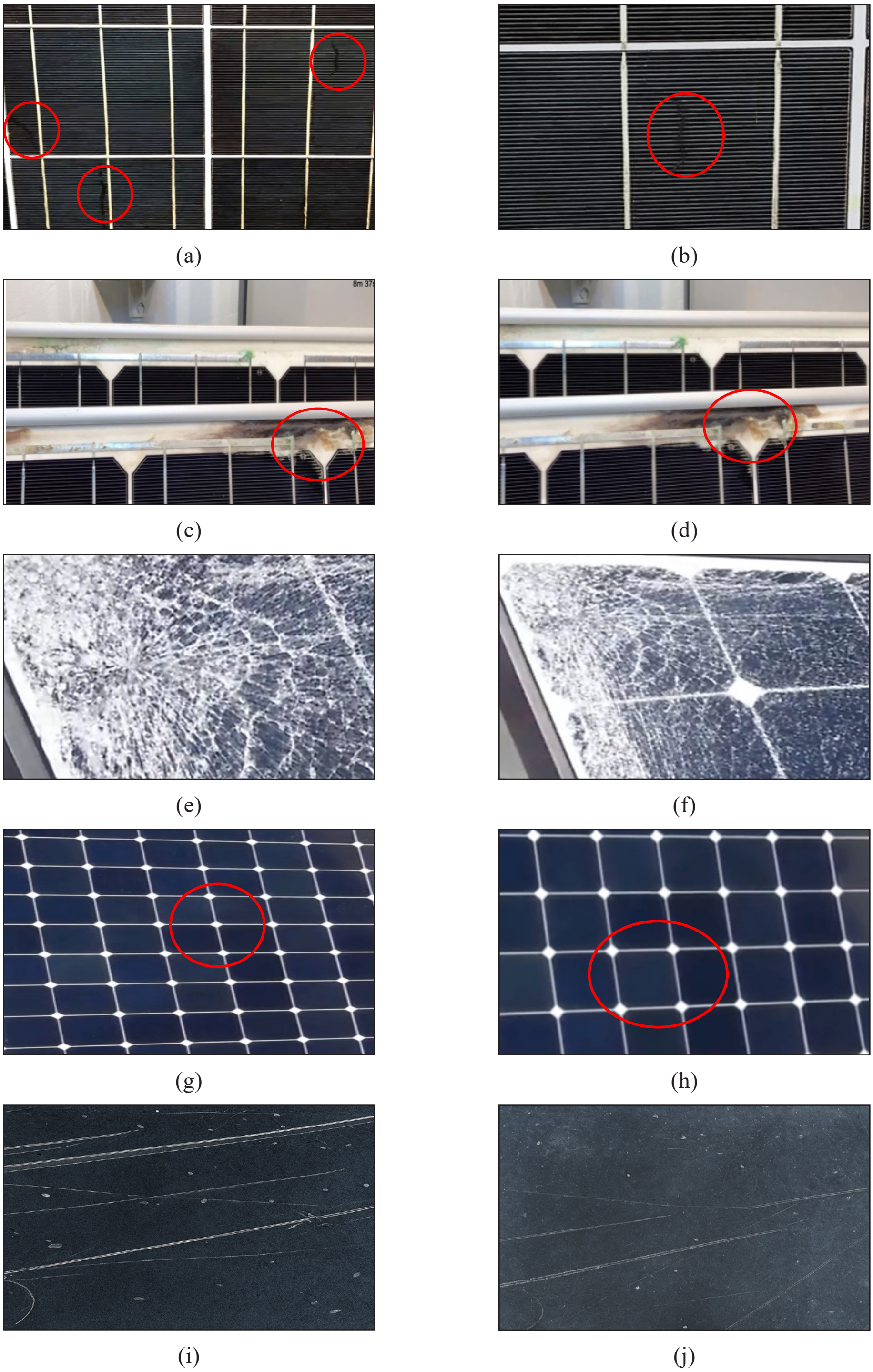

Figure 2. Example of PV module defect images (a-b) Cracks; (c-d) Delamination; (e-f) Glass breakage; (g-h) Encapsulation discolouring and (i-j) Gridline corrosion

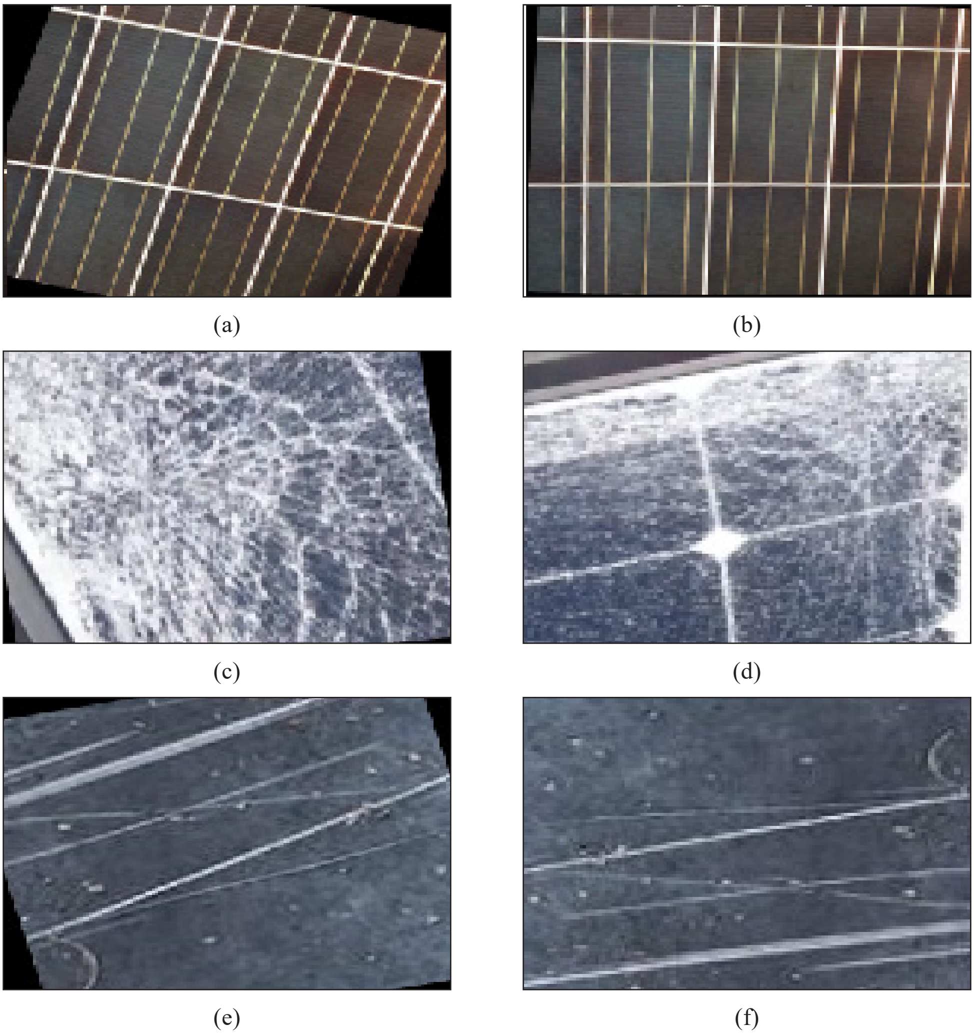

Figure 3. Sample of augmentation dataset with (a, c, e) X, Y translation and (b, d, f) rotation

layers to receive inputs from multiple sources and transmit data to others, providing enhanced flexibility in deep learning models. The design of a DAG network can be more complex, with layers creating data for other layers and accepting input from other layers. Some networks that fall within the DAG network category are GoogLeNet, ResNet-50, Inception-V3, and ResNet-101 (Anilkumar et al., 2021).

AlexNet

AlexNet is an 8-layer deep network that is also one of the CNN architectures. The layers consist of 3 fully connected layers and 5 convolutional layers (Xu et al., 2022). The details of the layer for the AlexNet architecture can be referred to in Table 2.

Table 2
AlexNet architecture

Layer		Feature Map	Size	Kernel Size	Stride	Activation
Input	Image	1	227 × 227 × 3	-	-	-
1	Convolution	96	55 × 55 × 96	11 × 11	4	relu
	Max Pooling	96	27 × 27 × 96	3 × 3	2	relu
2	Convolution	256	13 × 13 × 256	5 × 5	1	relu
	Max Pooling	256	13 × 13 × 256	3 × 3	2	relu
3	Convolution	384	13 × 13 × 384	3 × 3	1	relu
4	Convolution	384	13 × 13 × 384	3 × 3	1	relu
5	Convolution	256	13 × 13 × 256	3 × 3	1	relu
	Max Pooling	256	6 × 6 × 256	3 × 3	2	relu
6	FC	-	9216	-	-	relu
7	FC	-	4096	-	-	relu
8	FC	-	4096	-	-	relu
Output	FC	-	100	-	-	Softmax

Note. Adapted from (Li et al., 2022)

This network consists of 8 layers with 5 total convolution layers, three max-pooling layers, and the remaining fully connected layers. The input image size required by this network is 227 × 227 × 3. First, the images will undergo the first convolution layer with a feature map of 96, making the images' size 55 × 55 × 96. The AlexNet adds a maximum pooling layer or sub-sampling layer with a filter size of 33 and a stride of 2. After reduction, 27 × 27 × 96 will be the final image's size. The final fully connected layer's output is fed back into a 1000-way Softmax, which generates a distribution over the 1000 class labels (Xu et al., 2022).

This network was selected for its deeper architecture and strong capability in learning complex features, which is crucial for detecting and classifying PV module defects such as cracks. In addition, AlexNet incorporates local response normalisation, which enhances the model's generalisation by standardising responses across multiple channels (Siuly et al.,2024).

Visual Geometry Group (VGG)

The Computer Vision Group from Oxford University discovered another CNN, namely the Visual Geometry Group (VGG), which is a complete deep-learning network (Tao, 2022). Image style transfer based on the VGG neural network model. According to Bagaskara and Suryanegara (2021), VGG uses fewer parameters while the network's performance improves. VGG-16's 3 × 3 filter size represents an improvement over AlexNet's version. The design of VGG-16 is composed of three Fully Connected Layers, three Max Pooling

Operations, and a Softmax Activation Layer for each block. Known for its efficiency and ease of use, VGG-16 excels in a variety of computer vision tasks, including image classification and object recognition. Max-pooling layers appear after a stack of convolutional layers with progressively deeper layers in the model's architecture (Zou et al., 2020). Using this method, the model can acquire intricate hierarchical representations of visual qualities, resulting in accurate and consistent predictions (Arvinth, 2024).

Meanwhile, the VGG-19 architecture is relatively simple, making it easier to understand and implement. With many available pre-trained models, it is well-suited for transfer learning in computer vision tasks. Its strong feature extraction capability is particularly important for capturing PV module defects in greater detail (Arvinth, 2024).

Figure 4 illustrates the layer structure of the VGG architectures. Column D displays the layers of VGG-16, while column E shows the layers of VGG-19. Columns A, B, and C highlight earlier versions of VGG, including modifications made before the development of VGG-16 and VGG-19 (Xia et al., 2020). The VGG architecture requires input images of

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 4. VGG architecture
Note. Adapted from (Bagaskara & Suryanegara, 2021)

224 × 224 pixels. It features convolution layers with 3 × 3 and 1 × 1 filters, which function as linear input transformations. A stride of 1 pixel is used for subsampling. The hidden layers use ReLU activation functions, which play a key role in the VGG architecture. As shown in Figure 4, VGG-19 contains 19 weight layers, whereas VGG-16 has 16. Both networks incorporate a max pooling layer along with multiple deep convolutional layers, and trainable convolutional layers of 19 in VGG-19 and 16 in VGG-16 (Triyadi et al., 2022).

Residual Neural Network

Unlike other CNNs, where convolutional layers can be stacked indefinitely without significant performance degradation, Residual Neural Networks (ResNet) introduce a unique residual block. This block uses skip connections to bypass certain layers, helping to mitigate issues like vanishing or exploding gradients during training (Rahman et al., 2021). ResNet has evolved, with architectures such as ResNet-18 and ResNet-50 being commonly compared. The primary difference between them is network depth, where ResNet-50 includes additional convolutional layers to enhance learning capability (Chindarkkar et al., 2020). ResNet-18, in contrast, is a smaller network with two-layer-deep residual blocks. The ResNet-18 architecture is a compact network featuring two layers of ResNet blocks. Compared to traditional CNN architecture, ResNet demonstrates superior performance in extracting relevant features that may encounter vanishing gradient problems or reduced representation capacity. ResNet's usage of skip connections enhances training convergence and feature extraction while also enabling deeper model creation (Luu et al.,2025).

Table 3 illustrates the layer structure of both ResNet-18 and ResNet-50 designs. Five convolution stages make up each ResNet architecture, and their purpose is to extract an image's features. 224 × 224-pixel images were required for this architecture's input size. Convolution is 3 × 3 for each layer with a 3 × 3 kernel size. Subsampling in ResNet-50 is achieved using convolution with a stride of two. This deep architecture, with 50 layers, provides significant advantages for image classification by enabling more complex feature extraction. Figure 5 illustrates the ResNet-50 architecture, highlighting its greater depth in comparison to ResNet-18. The network features global average pooling, a fully connected layer, and a Softmax layer at the final stage for classification (Li & He, 2018)

Inception-V3

Inception-V3 is a CNN architecture that consists of 48 layers deep. This network required images to have dimensions of 299 by 299 as requirement to use this network. The improvement towards GoogLeNet produces Inception-V3, where this network is the enhanced version of the GoogLeNet network. Inception-V3's layers are shown in Figure 5, where this network consists of 13 total layers.

Differing from LeNet and VGG, the uniqueness of this network can be seen in the layer where there are special layers called the Inception Module. This module uses receptive

Table 3
ResNet architecture

Layer name	Output size	ResNet-18	ResNet-50
Conv1	112 × 112	7 × 7, 64, stride 2	7 × 7, 64, stride 2
Conv2_x	56 × 56	3 × 3 max pool, stride 2	3 × 3 max pool, stride 2
		$\begin{bmatrix}3x3,64\\3x3,64\end{bmatrix}x2$	$\begin{bmatrix}1x1,64\\3x3,64\\1x1,256\end{bmatrix}x3$
Conv3_x	28 × 28	$\begin{bmatrix}3x3,128\\3x3,128\end{bmatrix}x2$	$\begin{bmatrix}1x1,64\\3x3,64\\1x1,256\end{bmatrix}x3$
Conv4_x	14 × 14	$\begin{bmatrix}3x3,256\\3x3,256\end{bmatrix}x2$	$\begin{bmatrix}1x1,256\\3x3,256\\1x1,1024\end{bmatrix}x6$
Conv5_x	7 × 7	$\begin{bmatrix}3x3,512\\3x3,512\end{bmatrix}x2$	$\begin{bmatrix}1x1,512\\3x3,512\\3x3,2048\end{bmatrix}x3$
	1 × 1	Average pool, 1000-d fc, softmax	
FLOPs		1.8×10^9	4.1×10^9

Note. Adapted from (Pangilinan et al., 2022)

Type	Kernel size/stride	Input size
Convolution	3 × 3/2	299 × 299 × 3
Convolution	3 × 3/1	149 × 149 × 32
Convolution	3 × 3/1	147 × 147 × 32
Pooling	3 × 3/2	147 × 147 × 64
Convolution	3 × 3/1	73 × 73 × 64
Convolution	3 × 3/2	71 × 71 × 80
Convolution	3 × 3/1	35 × 35 × 192
Inception module	Three modules	35 × 35 × 288
Inception module	Five modules	17 × 17 × 768
Inception module	Two modules	8 × 8 × 1,280
Pooling	8×8	8 × 8 × 2,048
Linear	Logits	1 × 1 × 2,048
Softmax	Output	1 × 1 × 1,000

Figure 5. Inception-V3 architecture
Note. Adapted from (Zhang et al., 2022)

kernels of various sizes to capture different levels of detail. Zero padding is applied to ensure consistency in the output size of convolution operations. The final feature maps are generated by concatenating the filtered outputs. The inception operation enhances feature extraction, allowing the network to capture richer details from the input image (Feng et al., 2019). The efficiency of the network is increased by Inception modules' usage of 1×1 convolutions, which lower the dimensionality and complexity of feature maps. Furthermore, this CNN architecture's auxiliary classifiers at intermediate layers improve performance by offering additional supervision during training, which helps address the vanishing gradient issue (Arnob et al., 2024)

GoogLeNet

GoogLeNet is a deep Convolutional Neural Network (CNN) model consisting of 22 layers (Swarup et al., 2023). It requires input images of size 224×224; the detailed layer structure can be seen in Figure 6. Designed with computational efficiency, GoogLeNet incorporates two extra classifier layers, which are linked to the outputs of the Inception (4a) and Inception (4d) layers (Zhang et al., 2022). This network consists of two auxiliary classifiers to conduct the gradient forward for gradient vanishing purposes. By combining feature information from various scales, convolutional kernels of various sizes broaden the network's breadth and increase its scalability, which improves the network's capacity to identify features during training (Wang & Feng, 2024).

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Figure 6. GoogLeNet architecture
Note. Adapted from (Zhang et al., 2022)

EffecientNet

EfficientNet is a family of convolutional neural networks developed to optimise both accuracy and computational efficiency through a novel compound scaling method. Unlike traditional CNNs that scale depth, width, or resolution independently, EfficientNet uniformly scales these dimensions using a set of fixed scaling coefficients, resulting in a more balanced architecture (Tan & Le, 2019). The EfficientNetV2 series further improves upon this by introducing faster training speed, reduced memory usage, and improved parameter efficiency, making it suitable for large-scale image classification tasks (Tan & Le, 2021). EfficientNetV2 represents an enhanced version of the original EfficientNet, developed to deliver higher accuracy while minimising computational demands. One of its key improvements lies in training efficiency, with reports indicating that EfficientNetV2 can train up to 11 times faster than its predecessor, EfficientNetV1 (Kamal et al., 2023).

Validation and Verification with Other CNN Architectures

Performance is evaluated using multiple metrics, with Figure 7 illustrating a sample confusion matrix for binary classification. This matrix highlights key values: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which help measure the model's accuracy and overall effectiveness in classification.

The confusion matrix consists of four key variables. True Positive (TP) represents the number of instances where the classifier correctly identifies the positive class, while True Negative (TN) indicates the cases where the classifier correctly identifies the negative class. False Negative (FN) refers to instances where the classifier incorrectly predicts an actual positive case as negative, whereas False Positive (FP) refers to cases where the classifier incorrectly predicts an actual negative case as positive (Alves et al., 2021). From these values, evaluation metrics such as accuracy, F1-score, sensitivity, specificity, and precision can be derived (Akşehir & Kiliç, 2022).

		Predicted Class	
		Negative	**Positive**
Actual Class	**Negative**	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Figure 7. Confusion matrix
Note. Adapted from (Akşehir & Kiliç, 2022)

Evaluation metrics, including accuracy, F1-score, sensitivity, specificity, and precision, are obtained from the confusion matrix, with their formulas listed in Equations 1 to 5. These metrics are used to evaluate the effectiveness of PV classification.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad [1]$$

$$\text{F1-Score} = 2 * \frac{\text{Precision}*\text{Senstivity}}{\text{Precision}+\text{Sensitivity}} \quad [2]$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad [3]$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad [4]$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad [5]$$

RESULTS AND DISCUSSIONS

The training process of all networks was performed on a system featuring an Intel (R) Core (TM) i3-5005U CPU operating at 2.00GHz, 8GB of RAM, and Windows 10 64-bit; the simulation and performance evaluation were carried out for the experimental configuration using MATLAB R2020a. The training time was considerably longer than with GPU-based implementations because all training was done on a CPU without GPU acceleration. Compared to shallower models like AlexNet, deeper architectures like ResNet-50 and Inception-V3 required noticeably longer convergence times during training. GPU acceleration would be necessary to deploy these models in real-time PV monitoring systems to increase computational efficiency and decrease latency, even though CPU-based training was adequate for experimental evaluation. Additionally, since deeper networks require more processing and storage power, memory usage is a crucial factor. This emphasises how, when thinking about real-world deployment, model accuracy must be balanced with scalability and resource requirements.

The performance of different CNN architectures, namely AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, Inception-V3, and GoogLeNet were evaluated through testing and comparison to determine the best-performing model. The data are split into two groups: training with 70% (976 images) and validation of 30% (418 images). The training settings were set to 20 iterations per epoch and 32 for batch size. Stochastic Gradient Descent with Momentum (SGDM) is used as a training solver, and the learning rate applied is 0.0001.

Comparison of Accuracy Performance

The architecture analysis is performed by comparing the accuracy of augmented and non-augmented data classifications. A non-augmented dataset is the raw data that is not going through augmentation techniques of images, such as rotation or flipping. The dataset that has been augmented will undergo an augmentation process, which includes translation and rotation. The results for both analyses will be discussed in this part to identify which technique shows the best accuracy.

Without Data Augmentation

Figure 8 presents each network's accuracy when trained on the original dataset. VGG-19 exhibited outstanding performance, attaining an impressive accuracy of 92.58%. This result indicates that VGG-19 outperforms the other networks in accurately classifying non-augmented images. Despite performing relatively less favourably than the other networks, AlexNet still achieved a commendable accuracy above 90%. A comparative analysis between VGG-19 and AlexNet was conducted by the authors in Yu et al. (2023). VGG-19 exhibited lower training losses, suggesting a lower error rate than AlexNet during the training process.

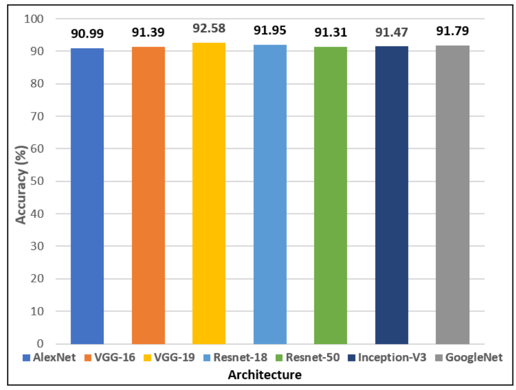

Figure 8. Comparison of accuracy between AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, Inception-V3 and GoogLeNet architectures (Without data augmentation)

A comparison analysis of training progress was made between the highest and lowest accuracy scores achieved. Figures 9 and 10 illustrate the training progress for both VGG-19 and AlexNet. The training cycle was set to 20 epochs with 640 iterations. Figure 9 illustrates the training progress of VGG-19 after the network has been trained. The accuracy initially starts at 30% and gradually increases, reaching 90% by the 100th iteration. The model maintains stability from the 230th iteration until it achieves a final accuracy of 92.58%. The corresponding training loss for VGG-19 is 0.2.

Figure 10 illustrates the training progress of AlexNet. The accuracy begins at 48% and undergoes a significant increase, reaching 90% by the 40th iteration. Afterwards, the model remains stable, resulting in a final accuracy of 90.99%. The training loss for AlexNet is slightly higher than that of VGG-19, with a difference of 0.6. Comparing the training progress of the two models, VGG-19 demonstrates a smoother increase in accuracy and lower training loss compared to AlexNet. These findings align with the analysis conducted by the authors in Tao (2022), which highlights VGG-19 as the top-performing model in terms of accuracy and training loss.

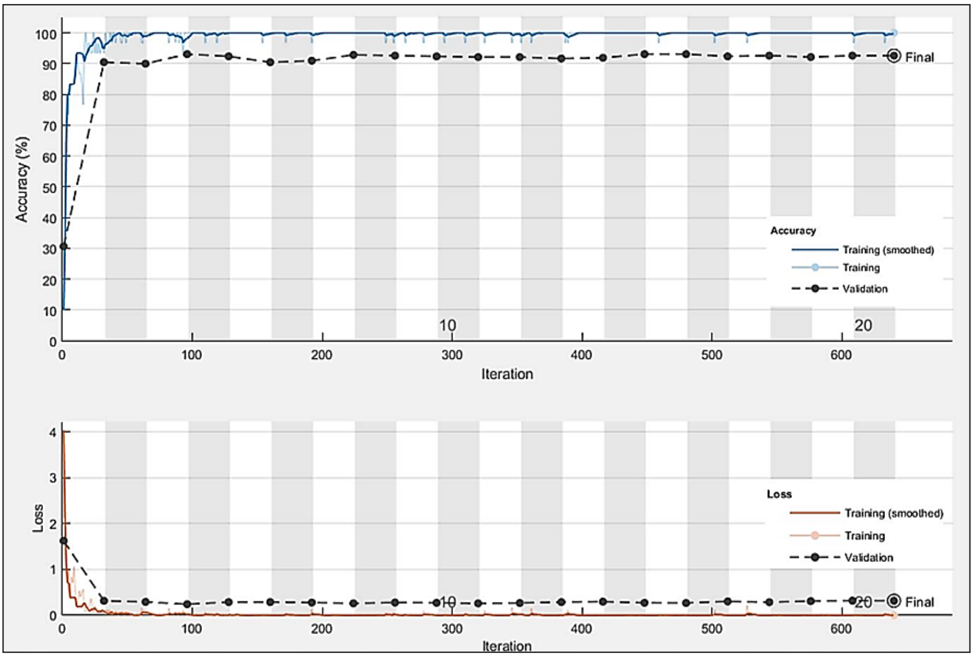

Figure 9. The training progress of VGG-19

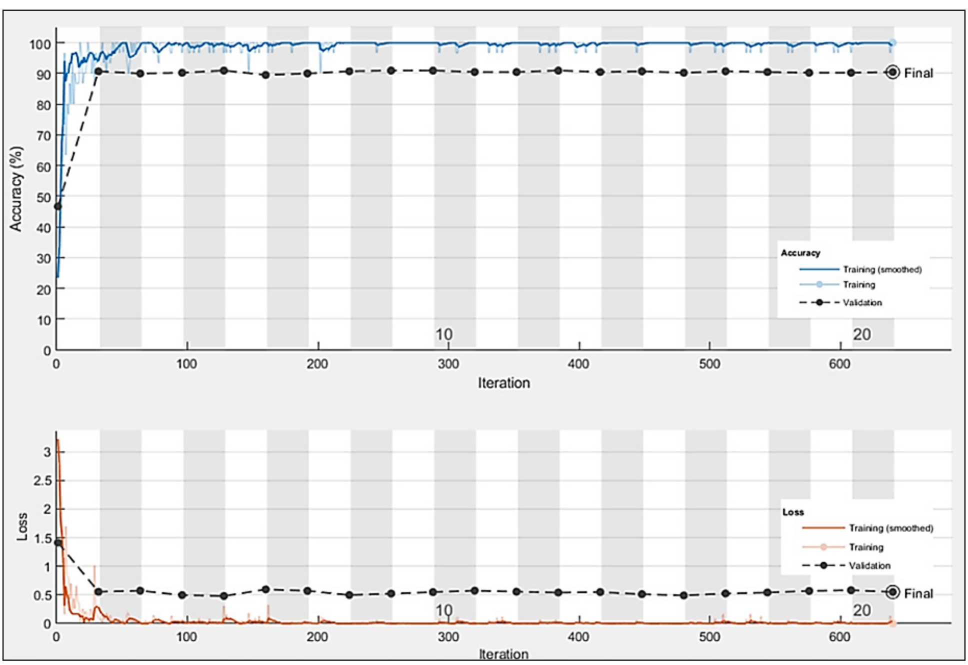

Figure 10. The training progress of AlexNet

With Data Augmentation

Figure 11 illustrates the accuracy comparison among different CNN architectures after dataset augmentation. ResNet-50, a DAG network, achieves the highest accuracy, demonstrating its superior deep feature extraction capability. Although GoogLeNet records the lowest accuracy at 98.33%, it still performs well, exceeding the 98% threshold. Notably, VGG-19, ResNet-18, and Inception-V3 represent both series and DAG networks that achieve identical accuracy levels, as shown in Figure 11.

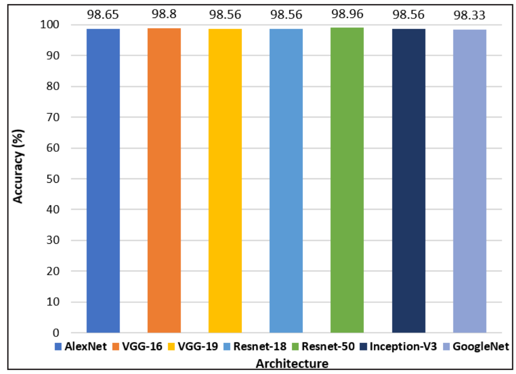

Figure 11. Comparison of accuracy between AlexNet, VGG-16, VGG-19, Resnet-18, Resnet-50, Inception-V3 and GoogLeNet architectures (With data augmentation)

Dataset augmentation significantly enhances model performance across all architectures, reinforcing its effectiveness in improving classification accuracy. Abu Salim et al. (2022) similarly reported accuracy improvements when the dataset size increased from 551 to 850 images. Additionally, in a Faster R-CNN implementation, ResNet-50 exhibited the highest performance, with accuracy rising from 63.2% (without augmentation) to 72.4% (with augmentation). These findings emphasise the benefits of data augmentation in both series and DAG networks, ensuring improved learning and robustness in PV defect classification. Figure 12 and Figure 14 depict the training progress of accuracy for the augmented dataset. ResNet-50's highest accuracy is achieved, starting at 33% and experiencing a rapid increase to 90% in less than 50 iterations. From the 170th iteration onwards, the accuracy stabilises at 98%, eventually reaching a final accuracy of 98.96%. Remarkably, the training loss is close to 0, indicating the outstanding performance of ResNet-50 in accurately classifying the images.

ResNet-50 demonstrated strong generalisation capability and was less prone to overfitting, largely due to its skip connections and batch normalisation layers (Negara et al., 2021). The skip connections enable direct feature transfer across layers, effectively addressing the vanishing gradient problem and allowing very deep networks to be trained without significant performance loss (Verma et al., 2024) The training loss approaching zero further confirms ResNet-50's effectiveness in accurately classifying PV defect images, while the convergence of training and validation losses at low values indicates minimal overfitting and stable learning behaviour (Zhu et al., 2020). Previous studies have also shown that ResNet-50 surpasses other architectures in PV module fault detection, achieving a maximum accuracy of 98.22%. Owing to its superior performance in image classification and pattern recognition tasks, ResNet-50 is considered one of the most suitable architectures for PV defect detection applications (Alsudi & Al-Aubidy, 2023).

The results obtained from the augmented dataset were used as a reference to analyse misclassification patterns, and a confusion matrix analysis was performed to evaluate class-wise performance and to identify misclassifications among the PV defect classes. Figure 13 shows the confusion matrix of ResNet-50 for 5 types of defects. Referring to Figure 13, the Cracks class achieved a perfect score, with all 23 images correctly classified and no misclassifications. The Delamination class recorded 32 correct predictions out of 33, with 1 image misclassified as Glass Breakage. Similarly, both Encapsulation Discolouring and Gridline Corrosion each had 1 misclassified image. The Glass Breakage class demonstrated strong performance with 306 correctly classified images, while 3 images were misclassified, distributed between Glass Breakage and Gridline Corrosion. These results reflect the model's high accuracy overall, with only minor misclassification occurring across a few classes.

Figure 14 illustrates the training progress of GoogLeNet, which starts with a relatively low initial accuracy of 68% but gradually improves, reaching a final accuracy of 98.33%. The training process stabilises around the 190th iteration, marking the saturation point. Similarly, ResNet-50 exhibits a strong learning curve, with its training loss nearing zero, highlighting the positive impact of data augmentation on model performance.

Cracks and Delamination classes in Figure 15 achieved excellent results in their respective defect prediction tasks. Cracks class achieved a perfect score of 23, while Delamination recorded 33 correct predictions, with no misclassifications in either class. However, Encapsulation Discolouring, Glass Breakage, and Gridline Corrosion classes experienced some misclassification issues. Encapsulation Discolouring had 5 images misclassified as Glass Breakage, indicating the model was unable to distinguish between these two defects. The Glass Breakage category contains one misidentification because the model mistakenly labelled a Gridline Corrosion image as Glass Breakage. The Gridline Corrosion category received a score of 18 out of 19 because one image was misclassified. GoogLeNet model achieved high accuracy in detecting Cracks and Delamination, but struggled to identify Encapsulation Discolouring correctly, which indicates the model had a problem distinguishing between visually similar defect types.

The small performance gap observed between certain architectures, where ResNet-50 at 98.96% and GoogLeNet at 98.33%, may partly be attributed to the evaluation approach, where results were averaged across three runs. This averaging tends to smooth out minor fluctuations between runs, leading to reduced variation in the final reported scores. As a result, the differences in accuracy appear smaller and may not fully reflect run-to-run variability. Although statistical significance tests such as standard deviation or confidence intervals were not reported, the models were run three times and showed highly consistent results with minimal variation. Therefore, the reported averages are considered reliable for comparative analysis, which aligns with the objective of this work.

These findings underscore the effectiveness of augmentation in enhancing classification accuracy across different architectures. While both GoogLeNet and ResNet-50, representing DAG networks, show significant improvements, ResNet-50 outperforms with higher accuracy and lower training loss, confirming its superior deep feature extraction capability. Despite GoogLeNet achieving a slightly lower accuracy, it still demonstrates robust performance, reinforcing the advantages of DAG networks in transfer learning for PV defect classification.

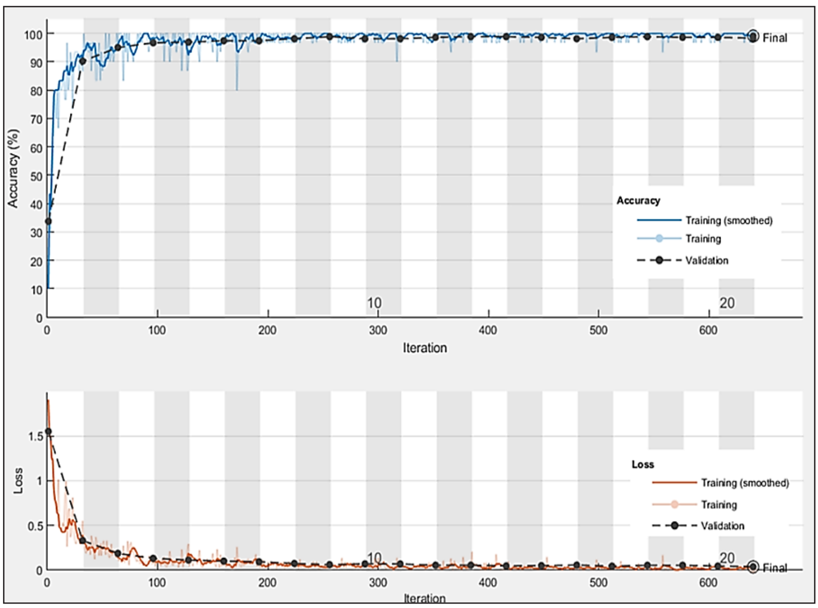

Figure 12. The training progress of ResNet-50

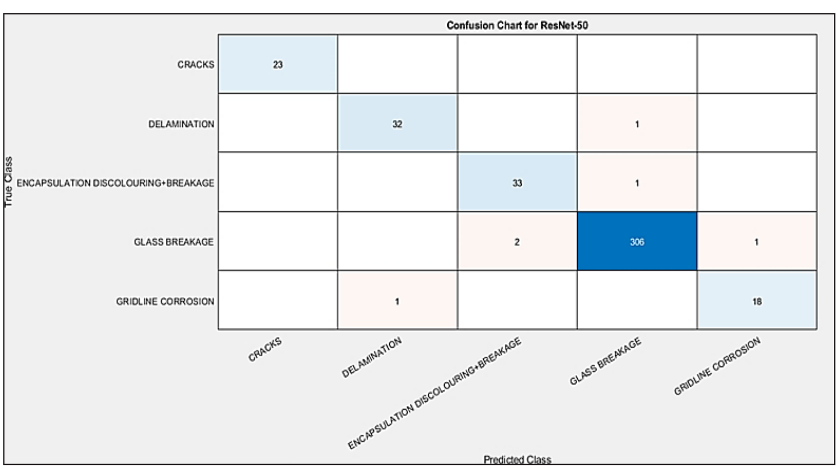

Figure 13. Confusion Matrix of ResNet-50

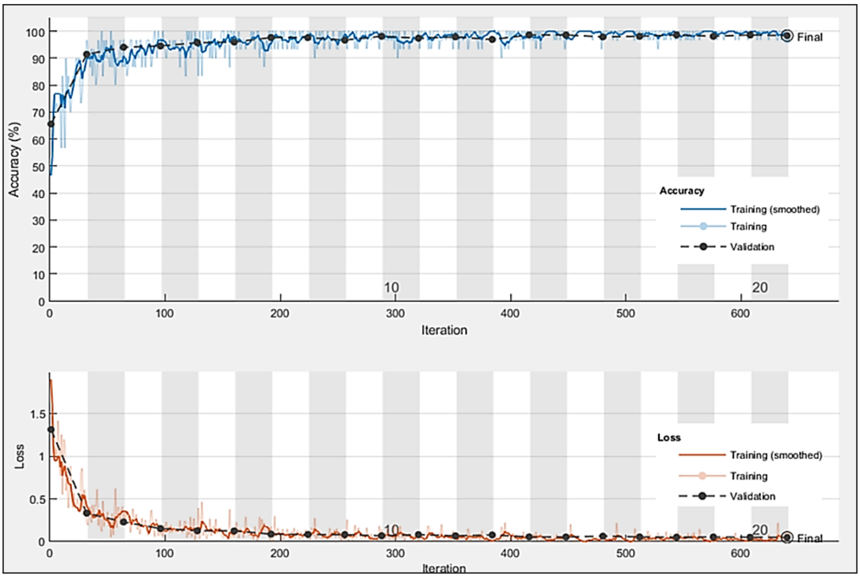

Figure 14. The training progress of GoogLeNet

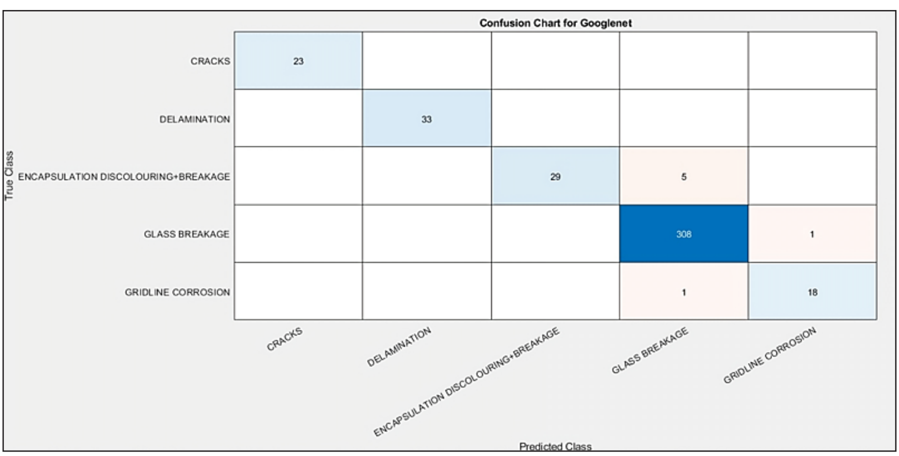

Figure 15. Confusion Matrix of GoogLeNet

Comparison of Performance Metrics

Performance metrics, including precision, F1-score, specificity, and sensitivity, are used to evaluate the effectiveness of the network. In this study, the performance of augmented and non-augmented datasets was compared to determine which approach produced superior results. The performance metrics for both analyses are presented in Figures 16 and 17. The augmented dataset was found to consistently outperform the non-augmented approach, significantly enhancing these metrics. However, strong results were achieved in terms of specificity for both approaches, with only a minimal difference in scores.

Without Data Augmentation

Figure 17 presents a comparison of performance metrics for different networks when the dataset is augmented. The F1-score analysis indicates that GoogLeNet performs well, achieving 74.75%, despite having a longer processing time than other networks, as reported by Rattanawin et al. (2023). In contrast, AlexNet demonstrates the lowest F1-score at 68.77%, a result consistent with findings by Tyassari et al. (2022), who noted its lower F1-score but emphasised its advantage of faster training time. VGG-19 achieves the highest sensitivity, whereas Inception-V3 records the lowest sensitivity score. In terms of specificity, AlexNet excels in accurately predicting the negative class, with most networks demonstrating strong performance in this metric. Regarding precision, ResNet-18 outperforms other networks with a score of 88.55%, while VGG-19 exhibits the lowest precision. These results highlight the varying strengths of series and DAG networks in transfer learning, where DAG networks such as ResNet-18 and GoogLeNet demonstrate superior feature extraction capabilities, while series networks like AlexNet and VGG-19 show trade-offs between speed and accuracy.

Comparison of Performance Analysis Between Data Augmentation and Without Data Augmentation

A comparison of performance metrics between augmented and non-augmented datasets has been conducted, revealing that the augmented dataset consistently outperforms the non-augmented dataset. As shown in Figures 17 and 18, the F1-score for the non-augmented dataset ranges from 68.77% to 74.75%, sensitivity from 68.32% to 73.14%, specificity from 94.24% to 95.26%, and precision from 55.42% to 88.55%. In contrast, the augmented dataset achieves significantly improved results, with F1-scores ranging from 96.93% to 97.89%, sensitivity from 95.72% to 98.10%, specificity from 98.85% to 99.53%, and precision from 96.77% to 98.65%. Notably, specificity exhibits strong performance in both cases compared to other metrics such as F1-score, sensitivity, and precision. This variation in performance can be attributed to the structural differences between series and DAG networks, as each network's unique layer arrangement influences its ability to optimise certain parameters (Shorten & Khoshgoftaar, 2019; Zhou et al., 2020).

Comparison of Performance Analysis Between CNN Architectures and EfficientNetV2-S

The comparison between seven CNN architectures and EfficientNetV2-S was conducted based on performance metrics. EfficientNetV2 was selected over EfficientNetV1 for the analysis because it offers faster training speed, improved accuracy, and reduced computational complexity, making it more suitable for efficient PV defect image classification (Kamal et al., 2023). The results indicate that ResNet-50, VGG-19, and VGG-16 outperformed EfficientNetV2-S in terms of accuracy, F1-score, sensitivity, and precision.

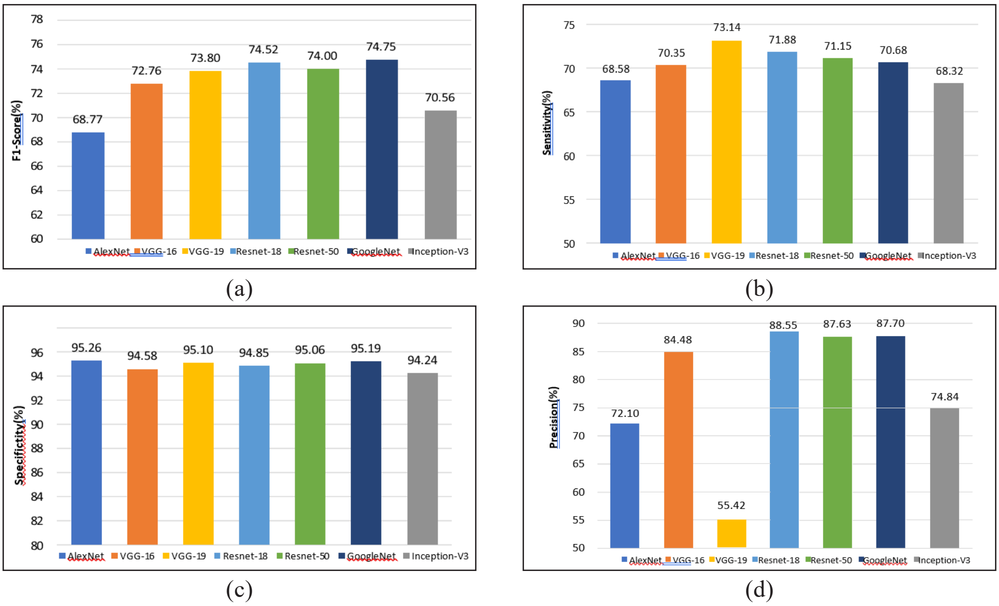

Figure 16. Comparison of Performance Metrics: (a) F1-Score; (b) Sensitivity; (c) Specificity; and (d) Precision between AlexNet, VGG-16, VGG-19, Resnet-18, Resnet-50, GoogLeNet and Inception-V3 Architectures (Without Data Augmentation)

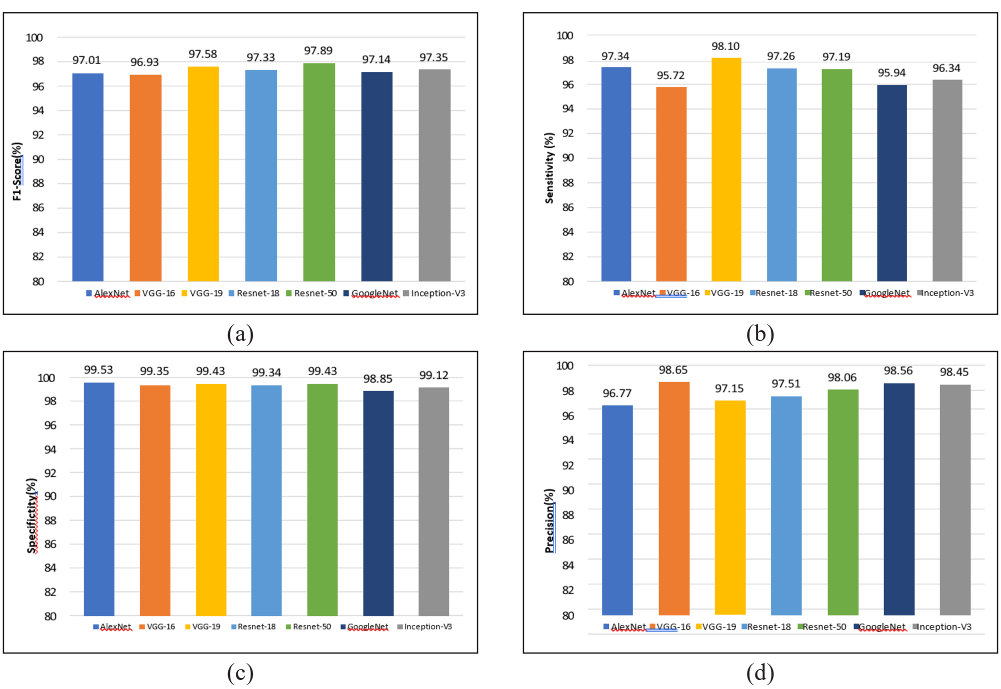

Figure 17. Comparison of Performance Metrics; (a) F1-Score; (b) Sensitivity; (c) Specificity; and (d) Precision between AlexNet, VGG-16, VGG-19, Resnet-18, Resnet-50, GoogLeNet and Inception-V3 Architecture (With data augmentation)

Figure 18 presents the comparison of the highest performance metric scores achieved by the CNN models against EfficientNetV2-S. As shown in Figure 18(a), ResNet-50 achieved a higher accuracy of 98.96% compared to 93.36% for EfficientNetV2-S. Similarly, Figure 18(b) demonstrates that ResNet-50 outperformed in F1-score with 97.89% against 93.69%. In terms of sensitivity, Figure 18(c) shows VGG-19 achieving 98.10% compared to 93.36% for EfficientNetV2-S. Finally, Figure 18(d) illustrates that VGG-16 achieved superior precision with 98.65%, outperforming EfficientNetV2-S at 93.75%.

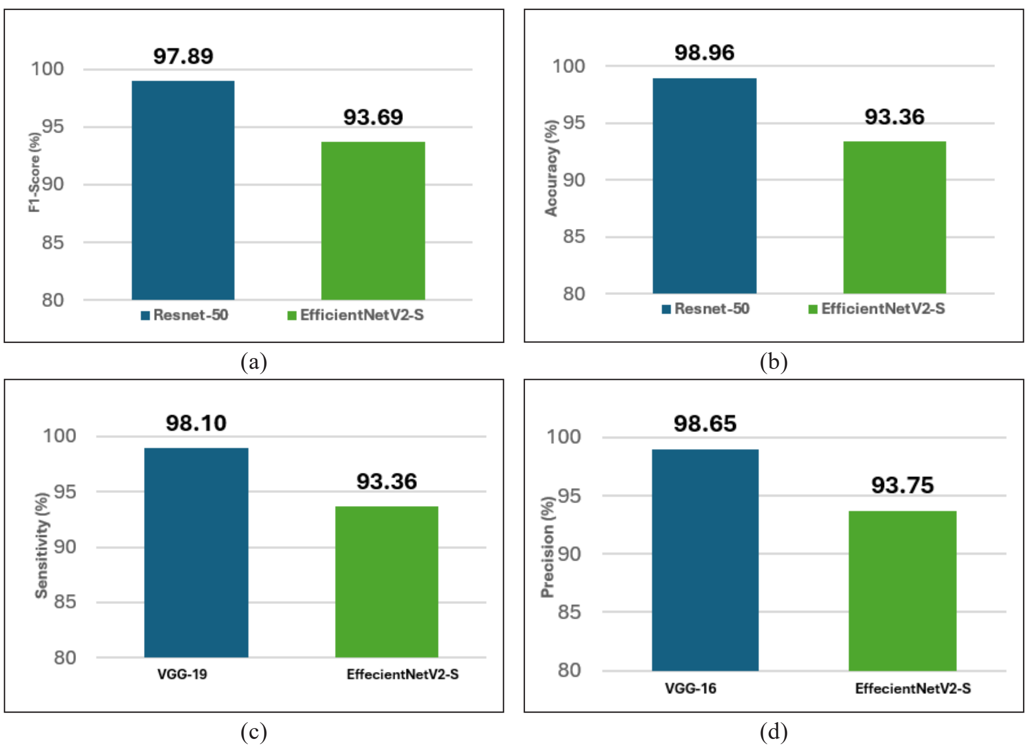

Figure 18. Comparison of Performance Metrics: (a) F1-Score; (b) Accuracy; (c) Sensitivity; and (d) Precision between CNN architecture and EfficientNetV2-S (With data augmentation)

CONCLUSIONS

The results indicated that the augmented dataset consistently outperformed the non-augmented dataset. The analysis of seven CNN architectures, namely AlexNet, VGG-16, VGG-19, ResNet-18, ResNet-50, GoogLeNet, and Inception-V3 demonstrate varying performance across different evaluation metrics. ResNet-50, a DAG network, achieved the highest classification accuracy and the best F1-score, highlighting its strength in deep feature extraction and residual learning. AlexNet, a series network, excelled in specificity, while VGG-19 demonstrated superior sensitivity, and VGG-16 achieved the highest precision. These findings align with the structural differences between series and DAG networks, where series networks perform well in straightforward feature extraction, while DAG networks leverage deeper architectures for enhanced learning. In addition to CNN-based architectures, this study compared EfficientNetV2-S to demonstrate the potential of alternative deep learning methods. Future work could further explore advanced approaches such as Vision Transformers or self-supervised learning, as well as multi-modal datasets such as RGB and infrared, to address challenges like subtle defect detection under varying conditions and enhance PV monitoring.

Although ResNet-50 performed the best in terms of accuracy and F1-score, its computational complexity and memory constraints may make it difficult to use in real-time PV inspection systems. Inference speed and consumption of energy are important considerations in resource-constrained situations, such as embedded devices or edge computing systems frequently employed in solar farm monitoring. Real-time applications may benefit more from simpler architectures (like AlexNet or VGG-16), which may provide faster inference at the cost of slightly lower accuracy. Therefore, a trade-off exists between maximising classification accuracy and ensuring operational feasibility in large-scale PV monitoring systems. Future work should investigate lightweight models or model compression techniques to balance accuracy with deployment efficiency.

Finally, while it is true that ResNet-50's superior performance can be anticipated due to its residual learning capabilities, the novelty of this work lies in providing a broader comparative study across seven CNN architectures for PV defect classification. Unlike prior studies that typically evaluate only a few models, this research systematically compares Series and DAG networks using multiple performance metrics, highlighting not only overall accuracy but also class-wise strengths such as sensitivity, precision, and specificity. This provides deeper insights into which architectures are more reliable for defect types, offering practical guidance for selecting models in future PV defect classification studies.

ACKNOWLEDGEMENT

Authors acknowledge the Universiti Teknologi MARA for funding under the Geran Penyelidikan SRC (600-RMC/SRC/5/3(032/2020).

LIST OF ABBREVIATIONS

AI	:	Artificial Intelligence
C-CE	:	Commercial Counter Electrode
CE	:	Counter Electrode
CNN	:	Convolutional Neural Network
DAG	:	Directed Acyclic Graph
DSSC	:	Dye-Sensitised Solar Cells

EL : Electroluminescence
GFDI : Ground Fault Detection Interrupters
H-CE : Homemade Counter Electrode
OCPD : Overcurrent Protection Devices
PV : Photovoltaic
RESNET: Residual Neural Network
RGB : Red Green Blue
VGG : Visual Geometry Group
ViT : Vision Transformer

REFERENCES

Ahmad, Abu Salim, S., Zakaria, N., Mokhtar, N., Mostafa, S. A., & Abdulkadir, S. J. (2022). Data augmentation on intra-oral images using image manipulation techniques. In the *2022 International Conference on Digital Transformation and Intelligence (ICDI)* (pp. 117-120). IEEE. https://doi.org/10.1109/ICDI57181.2022.10007158

Akşehir, Z. D., & Kiliç, E. (2022). How to handle data imbalance and feature selection problems in CNN-based stock price forecasting. *IEEE Access, 10*, 31297-31305. https://doi.org/10.1109/ACCESS.2022.3160797

Alsudi, I. M., & Al-Aubidy, K. M. (2023). Deep learning-based fault classification and location for photovoltaic systems. In *2023 16th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 760-764). IEEE. https://doi.org/10.1109/DeSE60156.2023.10393961

Altinkaya, C., Atli, A., Atilgan, A., Salimi, K., & Yildiz, A. (2020). Facile fabrication of low-cost, low-temperature carbon-based counter electrode with an outstanding fill factor of 73% for dye-sensitised solar cells. *International Journal of Energy Research, 44*(12), 9570-9581. https://doi.org/10.1002/er.5174

Alves, R. H. F., de Deus Junior, G. A., Marra, E. G., & Lemos, R. P. (2021). Automatic fault classification in photovoltaic modules using convolutional neural networks. *Renewable Energy, 179*, 502-516. https://doi.org/10.1016/j.renene.2021.07.070

Anilkumar, K. K., Manoj, V. J., & Sagi, T. M. (2021). Automated detection of leukaemia by pretrained deep neural networks and transfer learning: A comparison. *Medical Engineering & Physics, 98*, 8-19. https://doi.org/10.1016/j.medengphy.2021.10.006

Appiah, A. Y., Zhang, X., Ayawli, B. B. K., & Kyeremeh, F. (2019). Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques. *International Journal of Photoenergy, 2019*, Article 6953530. https://doi.org/10.1155/2019/6953530

Arnob, A. S., Kausik, A. K., Islam, Z., Khan, R., & Rashid, A. B. (2025). Comparative result analysis of cauliflower disease classification based on deep learning approach VGG16, Inception V3, ResNet, and a custom CNN model. *Hybrid Advances, 7*, Article 100440. https://doi.org/10.1016/j.hybadv.2025.100440

Arvinth, S. (2024). *VGG16 vs. VGG19 vs. EfficientNetB5: Which model is best for your project?* Medium. https://medium.com/

Azis, S. F. S., Shahbudin, S., Kassim, M., Mohamad, R., & Rahman, F. Y. A. (2022). Photovoltaic module defects classification analysis using DenseNet architecture. In *2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA)* (pp. 1-5). IEEE. https://doi.org/10.1109/ISIEA54517.2022.9873681

Bagaskara, A., & Suryanegara, M. (2021). Evaluation of VGG-16 and VGG-19 deep learning architectures for classifying people with dementia. In *2021 4th International Conference on Computer and Informatics Engineering (IC2IE)* (pp. 1-4). IEEE. https://doi.org/10.1109/IC2IE53219.2021.9649132

Bakhshayesh, P. R., Ejtehadi, M., Taheri, A., & Behzadipour, S. (2022). The effects of data augmentation methods on the performance of human activity recognition. In *2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICSPIS56952.2022.10043959

Balani, H., Darji, P., Sanghvi, R., Kalavadiya, A., & Parekh, R. (2021). Performance analysis of a single junction crystalline solar cell using 1D drift diffusion modelling. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 339-346). IEEE. https://doi.org/10.1109/ICESC51422.2021.9533003

Basnet, B., Chun, H., & Bang, J. (2020). An intelligent fault detection model for fault detection in photovoltaic systems. *Journal of Sensors, 2020*, Article 6960328. https://doi.org/10.1155/2020/6960328

Chindarkkar, A., Priyadarshi, S., Shiradkar, N. S., Kottantharayil, A., & Velmurugan, R. (2020). Deep learning-based detection of cracks in electroluminescence images of fielded PV modules. In *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)* (pp. 1612-1616). IEEE. https://doi.org/10.1109/PVSC45281.2020.9300615

Cipriani, G., D'Amico, A., Guarino, S., Manno, D., Traverso, M., & Di Dio, V. (2020). Convolutional neural network for dust and hotspot classification in PV modules. *Energies, 13*(23), Article 6357. https://doi.org/10.3390/en13236357

Drir, N., Mellit, A., & Bettayeb, M. (2025). A novel ensemble CNN framework with weighted feature fusion for fault diagnosis of photovoltaic modules using thermography images. *IEEE Journal of Photovoltaics, 15*(1), 146-154. https://doi.org/10.1109/JPHOTOV.2024.3492283

Faheem, M. B., Khan, B., Feng, C., Subhani, W. S., Mabrouk, S., Sayyad, M. H., Yildiz, A., Zhang, W.-H., & Qiao, Q. (2022). Van der Waals epitaxial growth for high-performance organic-free perovskite solar cell: Experimental and theoretical insights. *Advanced Materials Interfaces, 9*(20), Article 2200421. https://doi.org/10.1002/admi.202200421

Feng, C., Zhang, H., Wang, S., Li, Y., Wang, H., & Yan, F. (2019). Structural damage detection using a deep convolutional neural network and transfer learning. *KSCE Journal of Civil Engineering, 23*(10), 4493-4502. https://doi.org/10.1007/s12205-019-0437-z

Hussein, L. A., Ayadi, O., & Fathi, M. (2021). Performance comparison for sun-tracking mechanism photovoltaic (PV) and concentrated photovoltaic (CPV) solar panels with fixed system PV panels in Jordan. In *2021 12th International Renewable Energy Conference (IREC)* (pp. 1-8). IEEE. https://doi.org/10.1109/IREC51415.2021.9427808

Kamal, M. R. M., Shahbudin, S., & Rahman, F. Y. A. (2023). Photovoltaic (PV) module defect image classification analysis using EfficientNetV2 architectures. In *2023 IEEE 14th Control and System Graduate Research Colloquium (ICSGRC)* (pp. 236-241). IEEE. https://doi.org/10.1109/ICSGRC57744.2023.10215491

Kellil, N., Aissat, A., & Mellit, A. (2023). Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under Algerian climatic conditions. *Energy, 263*, Article 125902. https://doi.org/10.1016/j.energy.2022.125902

Kocak, Y., Atli, A., Atilgan, A., & Yildiz, A. (2019). Extraction method-dependent performance of bio-based dye-sensitised solar cells (DSSCs). *Materials Research Express, 6*(9), Article 095512. https://doi.org/10.1088/2053-1591/ab2ef7

Li, B., & He, Y. (2018). An improved ResNet based on the adjustable shortcut connections. *IEEE Access, 6*, 18967-18974. https://doi.org/10.1109/ACCESS.2018.2814605

Li, J., Huang, H., Hu, F., & Ou, Y. (2022). Classification of ceramics based on an improved AlexNet convolutional neural network. In *2022 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, and International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* (pp. 1-8). IEEE. https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9927857

Li, X., Li, W., Yang, Q., Yan, W., & Zomaya, A. Y. (2020). An unmanned inspection system for multiple defect detection in photovoltaic plants. *IEEE Journal of Photovoltaics, 10*(2), 568-576. https://doi.org/10.1109/JPHOTOV.2019.2955183

Luu, T. T., & Huynh, D. A. (2025). A ResNet-based deep reinforcement learning framework using soft actor-critic for remaining useful life prediction of rolling bearings. *Results in Engineering, 25*, Article 106739. https://doi.org/10.1016/j.rineng.2025.106739

Meng, L., Wang, Y., Wang, H., Tong, X., Sun, Z., & Cai, Z. (2023). Task offloading optimisation mechanism based on a deep neural network in an edge-cloud environment. *Journal of Cloud Computing, 12*(1), Article 76. https://doi.org/10.1186/s13677-023-00450-6

Negara, B. S., Satria, E., Sanjaya, S., & Santoso, D. R. D. (2021). ResNet-50 for classifying Indonesian batik with data augmentation. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICOTEN52080.2021.9493488

Nylund, S., & Barbari, Z. (2019). *Study of defects in PV modules: UV fluorescence and thermographic photography for photovoltaics (PV) field application* (master's thesis). Mälardalen University, Sweden.

Özel, K., & Yıldız, A. (2022). Comprehensive understanding of the role of emitter layer thickness for metal oxide–semiconductor–based solar cells. *IEEE Journal of Photovoltaics, 12*(1), 251-258. https://doi.org/10.1109/JPHOTOV.2021.3119612

Pangilinan, J. R., Legaspi, J., & Linsangan, N. (2022). InceptionV3, ResNet50, and VGG19 performance comparison on tomato ripeness classification. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 619-624). IEEE. https://doi.org/10.1109/ISRITI56927.2022.10052920

Pavlitskaya, S., Oswald, J., & Zöllner, J. M. (2022). Measuring overfitting in convolutional neural networks using adversarial perturbations and label noise. In *2022 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1551-1559). IEEE. https://doi.org/10.1109/SSCI51031.2022.10022094

Pierdicca, R., Malinverni, E. S., Piccinini, F., Paolanti, M., Felicetti, A., & Zingaretti, P. (2018). Deep convolutional neural network for automatic detection of damaged photovoltaic cells. *The International*

Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2, 893–900. https://doi.org/10.5194/isprs-archives-XLII-2-893-2018

Rahman, M. R. U., Chen, H., & Xi, W. (2019). U-Net-based defects inspection in photovoltaic electroluminescence images. In *2019, IEEE International Conference on Big Knowledge (ICBK)* (pp. 215-220). IEEE. https://doi.org/10.1109/ICBK.2019.00036

Rahman, M. R., Tabassum, S., Haque, E., Nishat, M. M., Faisal, F., & Hossain, E. (2021). CNN-based deep learning approach for micro-crack detection of solar panels. In *2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE. https://doi.org/10.1109/STI53101.2021.9732592

Rao, S., Muniraju, G., Tepedelenlioglu, C., Srinivasan, D., Tamizhmani, G., & Spanias, A. (2021). Dropout and pruned neural networks for fault classification in photovoltaic arrays. *IEEE Access, 9*, 120034-120042. https://doi.org/10.1109/ACCESS.2021.3108684

Rattanawin, P., Pakinsee, T., & Songmuang, P. (2023). A GoogLeNet performance approach for COVID-19 detection using chest X-ray images. In *2023 15th International Conference on Knowledge and Smart Technology (KST)* (pp. 1-5). IEEE. https://doi.org/10.1109/KST57286.2023.10086817

Salih, N., Ksantini, M., Hussein, N., Ben Halima, D., Abdul Razzaq, A., & Ahmed, S. (2023). Prediction of ROP zones using deep learning algorithms and the voting classifier technique. *International Journal of Computational Intelligence Systems, 16*(1), Article 86. https://doi.org/10.1007/s44196-023-00268-9

Salimi, K., Atilgan, A., Aydin, M. Y., Yildirim, H., Celebi, N., & Yildiz, A. (2019). Plasmonic mesoporous core-shell Ag-Au@TiO_2 photoanodes for efficient light harvesting in dye-sensitised solar cells. *Solar Energy, 193*, 820-827. https://doi.org/10.1016/j.solener.2019.10.039

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), Article 60. https://doi.org/10.1186/s40537-019-0197-0

Siuly, S., Khare, S. K., Kabir, E., Sadiq, M. T., & Wang, H. (2024). An efficient Parkinson's disease detection framework: Leveraging time-frequency representation and AlexNet convolutional neural network. *Computers in Biology and Medicine, 174*, Article 108462. https://doi.org/10.1016/j.compbiomed.2024.108462

Swarup, C., Singh, K. U., Kumar, A., Pandey, S. K., & Singh, T. (2023). Brain tumour detection using CNN, AlexNet, and GoogLeNet ensembling learning approaches. *Electronic Research Archive, 31*(5), 2864-2885. https://doi.org/10.3934/era.2023146

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (pp. 6105-6114). PMLR.

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. In *Proceedings of the 38th International Conference on Machine Learning* (pp. 10096-10106). PMLR.

Tao, Y. (2022). Image style transfer based on the VGG neural network model. In *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 1475-1482). IEEE. https://doi.org/10.1109/AEECA55500.2022.9918891

Tella, H., Hussein, A., Rehman, S., Liu, B., Balghonaim, A., & Mohandes, M. (2025). Solar photovoltaic panel cells' defects classification using deep learning ensemble methods. *Case Studies in Thermal Engineering, 66*, Article 105749. https://doi.org/10.1016/j.csite.2024.105749

Triyadi, A. B., Bustamam, A., & Anki, P. (2022). Deep learning in image classification using VGG-19 and residual networks for cataract detection. In *2022 2nd International Conference on Information Technology and Education (ICITE)* (pp. 293-297). IEEE. https://doi.org/10.1109/ICITE54466.2022.9759886

Tyassari, W., Jusman, Y., Riyadi, S., & Sulaiman, S. N. (2022). Classification of cervical precancerous cells of ThinPrep images based on the deep learning model AlexNet and InceptionV3. In *2022, IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 276-281). IEEE. https://doi.org/10.1109/CSNT54456.2022.9787658

Verma, H., Siruvuri, S. D. V. S. S. V., & Budarapu, P. R. (2024). A machine learning-based image classification of silicon solar cells. *International Journal of Hydromechatronics, 7*(1), 49-66. https://doi.org/10.1504/IJHM.2024.137958

Wang, S., & Feng, Z. (2024). Multi-sensor fusion rolling bearing intelligent fault diagnosis based on VMD and ultra-lightweight GoogLeNet in industrial environments. *Digital Signal Processing, 145*, Article 104306. https://doi.org/10.1016/j.dsp.2023.104306

Xia, H., Zhang, Y., Yang, M., & Zhao, Y. (2020). Visual tracking via deep feature fusion and correlation filters. *Sensors, 20*(12), Article 3370. https://doi.org/10.3390/s20123370

Xu, W., Zeng, T. H., & Shalaginov, M. (2022). Class activation mapping-enhanced AlexNet convolutional neural networks for early diagnosis of Alzheimer's disease. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2550-2555). IEEE. https://doi.org/10.1109/BIBM55620.2022.9994868

Yu, F., Lu, T., & Xue, C. (2023). Deep learning-based intelligent apple variety classification system and model interpretability analysis. *Foods, 12*(4), Article 885. https://doi.org/10.3390/foods12040885

Zhang, Y., Zhang, Y., Wang, L., Dong, X., Li, Y., Sun, H., & Yang, X. (2022). Crack detection of electrical equipment based on an improved GoogLeNet. In the *2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICARCE55724.2022.10046575

Zhu, W., & Mousavi, S. M. (2020). Seismic signal augmentation to improve generalisation of deep neural networks. In G. C. B. (Ed.), *Advances in Geophysics* (Vol. 61, pp. 181-222). Elsevier. https://doi.org/10.1016/bs.agph.2020.08.004

Zou, S., Chen, W., & Chen, H. (2020). Image classification model based on deep learning in the Internet of Things. *Wireless Communications and Mobile Computing, 2020*, Article 6677907. https://doi.org/10.1155/2020/6677907

Zulkarnain, I., Nurmalasari, R. R., & Azizah, F. N. (2022). Table information extraction using data augmentation on deep learning and image processing. In *2022 16th International Conference on Telecommunication Systems, Services, and Applications (TSSA)* (pp. 1-6). IEEE. https://doi.org/10.1109/TSSA56819.2022.10063909

Zyout, I., & Oatawneh, A. (2020). Detection of PV solar panel surface defects using transfer learning of the deep convolutional neural networks. In *2020 Advances in Science and Engineering Technology International Conferences (ASET)* (pp. 3-6). IEEE. https://doi.org/10.1109/ASET48392.2020.9118382