A SWIN TRANSFORMER IN THE CLASSIFICATION OF IR IMAGES OF PHOTOVOLTAIC MODULE

S.Bharathi Research Scholar, SCSVMV University, Kanchipuram, Tamil Nadu, India,
P.Venkatesan Associate Professor, SCSVMV University, Kanchipuram, Tamil Nadu, India,
bharathieee@kanchiuniv.ac.in; venkatesan.p@kanchiuniv.ac.in

ABSTRACT: The faults occurring in the photo voltaic module as to be detected in order to increase its efficiency. The infrared images, electroluminescent images, and photo luminescent images of the photo voltaic modules have been used to detect and classify the faults. The infrared data set is used for classification and it is highly imbalanced. To make it balanced, generative adversarial networks are used to generate images for each fault category. If the fault classification is done using a convolution neural network, the feature maps are generated by convolution operation with filters on images. The convolution neural network model is to be trained for 1000 epochs and the time required for training is greater than 24 hours. The computational cost of a convolution neural network (CNN) is reduced in transformers through the attention mechanism. The Swin (shifted window) transformers which are used for classification as to be trained for 40 epochs and the maximum training time is less than 6 hours. The computation time, categorical classification accuracy, and top-5-accuracy obtained using Swin classifier are compared with the existing methods. It is found that the computational cost is very much reduced by the Swin transformer and top-5-accuracy of 99.04% is obtained by it while classifying 11 faults of IR images.

Keywords: MLP, IR images, GAN's, MSA, SW-MSA,MHSA

1. INTRODUCTION

In k means algorithm the number of clusters should be specified before the execution of k-means algorithm, In [1] the u-kmeans algorithm by itself finds the value of optimal clusters. image and numerical datasets were classified by u-kmeans algorithm. In [2] PCA is applied to reduce the dimensions of the features obtained by applying images from microstructural datasets to VGG16. later k-means clustering is used to classify it. In [3] variants of VIT and their application is dealt In [4] voltage and current values are measured under faulty conditions on power distribution side. The voltage and current values are given input to DNN for training.The performance of DNN is better than SVM classifier. In [5] through an experimental setup the fault signals such as irradiation,current,voltage, temperature, etc., are measured and they are applied to the stacked auto encoder, the fused features obtained from the auto encoder are applied to the DBM for classification

In [6] the training data's arrangement is indicated in the affinity graph. Depending upon the affinity penetration a cluster of samples having similar nature was found. the group structure is updated with the help of values obtained using the loss function. In [7] 3D convolution neural network is used to classify the 3D images. these 3D images are obtained through the gramian angular field of DC and AC signals in the PV system. In [8] average silhouette, elbow, and Nb-clust R packages are employed to find the optimal number of clusters k. In [9] using the Simulink model the dataset is formed with 8 categories of faults. The fault classification is done by a multi-layer perceptron. In [10] the fault classification of temporary shading faults and faults on DC side of PV system is considered. The residuals which is the difference between measured values and simulated outputs. The abnormal residuals are applied to 1 SVM and to represent the presence of faults clustering techniques are applied and the normal residuals are used for training. In [11] the classification of an open circuit fault, line-to- line fault, and the normal condition is done based on the min-max values of illumination, voltage, and current for each condition. In [12]MRI image restoration is done using Swin transformer and they found that Swin transformer and PIDDGAN yield better results with that Deep ADMM Net, U-Net, DAGAN as well as ground truth MR images (GT) and under-sampled zero-filled MR images (ZF) using Gaussian 1D 30% mask. In [14] surface defects of PV cells are classified by training CNN with R, G, and B components individually. later these features were combined and applied to a fully connected network, finally classification is done by using a soft max classifier. In [15] pre-trained VGG is used to extract

 http://doi.org/10.36893/DRSR.2023.V13I4N13.059-070

features and these features are fed to the SVM classifier. In [16] the PCA is used to classify the dataset consisting of electrical parameters, irradiance, and temperature. In [17] 4 fault classes are classified by using 11 parameters of PV cell. An ensemble model is used for classifying the faults. Chen et [18] proposed cross-attention fusion, in which the class token of one branch interacts with the patch token of another branch to transfer information and to learn about other branch tokens, later the knowledge is shared with its branch to enhance the features of the patches. Fan et[19] proposed transformers working on multiple scales to increase the channel capacity when pooling of resolution is done from the input to the output side. Chen et [20] proposed a region to the local token encoder in which each local token interacts with one of the tokens from the regional self-attention block and then they are fed to the local attention block, later they are fed to the feed-forward network. Zhang et [21] proposed a method in which local token is combined with global token and reduced both memory and computation costs. Xie et [22] used convolutions to obtain positional and spatial information, the cross window approach of multi-head self-attention computation, and finally aggregation of low-level features were done by them. Liu et [23] addressed the training in stability problems by (i) post residual norm (ii)computing attention through scaled cosine approach and (iii) computing bias in log space. Yuan et [24] rectified the issue of (i) redundant attention and (ii) lines and edges in the vicinity are not modeled properly in their work by introducing tokens to the token mechanism. Zhou et [25] rectified the problem of attention collapse by computing self-attention in higher dimensional space and by computing re-attention.

To the best of my knowledge, CNN is used for the classification of faults in IR images but its computational cost is very high. So there exists a need to reduce the computational time. This issue is solved in this paper by using Swin transformer for classification. The novelty in this paper is the use of Swin transformer to classify the faults in infrared images of photo voltaic modules since the computational complexity of it is varying linearly with image size.

2. RELATED WORK

2.1. Generator adversarial Network

Data augmentation aims to generate more images that resemble the original images. Figure 1 indicates the main block of the GAN network. The GAN network consists of a generator and discriminator. The generator tries to generate data similar to the original data by getting a noise vector as the input. The input noise vector is referred to as the latent vector. The discriminator's input may be from the original dataset or the generator generated fake data. The discriminator tries to identify whether the input data given to it is real or fake. The generator and the discriminator work in such a way to minimize their loss functions respectively. Initially, the generator will be generating an irrelevant output, over time it tries to generate data resembling the original data. In the same way initially, the discriminator predicts the fake data as the original data and vice versa, later it learns well and predicts correctly.

The structure of the generator and discriminator network used in the production of gray images of each category of faults is shown in figure 2. The latent vector size is 100. The generated image is of size 40X 24.

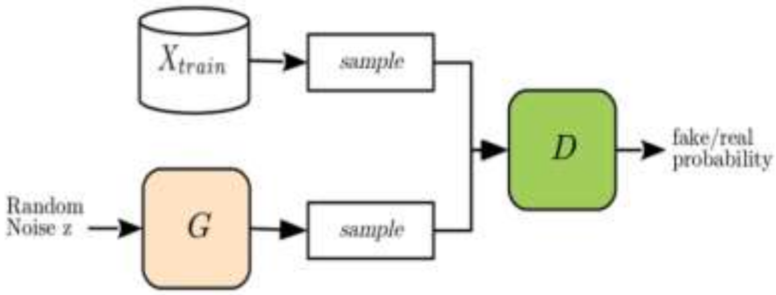


Figure 1: Representation of generative adversarial network.

2.1.1 Activation layer

An activation function gives a small output for small inputs, and a larger value if its input is above the threshold. Activation functions are used to introduce non-linear ties into neural networks, due to that the neural networks can learn complex computations.

 http://doi.org/10.36893/DRSR.2023.V13I4N13.059-070

2.1.2 ReLu layer

ReLu is a linear function that gives output that is equal to that of input if the input applied to this layer is positive. This is used to avoid the diminishing gradient.

2.1.3 Soft Max layer

An activation function chosen in the last layer of the discriminator is a soft max function which is used to normalize the outputs of the last fully connected layer to probabilities of the labeled class. The soft max function acts on a vector k consisting of real numbers and it outputs the probability to the vector of k that sum to 1. The expression used to find the probability values is shown in equation (1)

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \qquad (1)$$

2.2 Swin transformer

Swin transformer forms the hierarchical feature representations and the computational complexity of it varies linearly with image size. In hierarchical representation, self-attention is computed within the patch of the image. the number of patches in a window is set as constant so the computational cost is varying linearly to image size. In layer 1 a normal window scheme is followed and in the subsequent layer (l+1) window segregation is done in such a way that novel windows formed in this layer cross the boundaries of the preceding layer and provide connections among them. The architecture of the swin transformer and sub-blocks of the swin transformer is given in Figure 3.Non-overlapping patches are formed in an RGB image. each patch is considered as a "token", The feature of a token is obtained by concatenating the raw pixel RGB values. If the patch size is taken as 4 X 4, then the feature map dimension is 4 X 4 X 3 for that patch. The input to the transformer blocks with modifying self-attention is the vector that converts the raw pixel value into a vector of fixed dimension. The linear layer output is fed to the swin transformer block. The output of the first swin transformer block will be of size (H/4 X W/4.). The tokens in the subsequent layers of the hierarchical structure is decreased by merging patches.

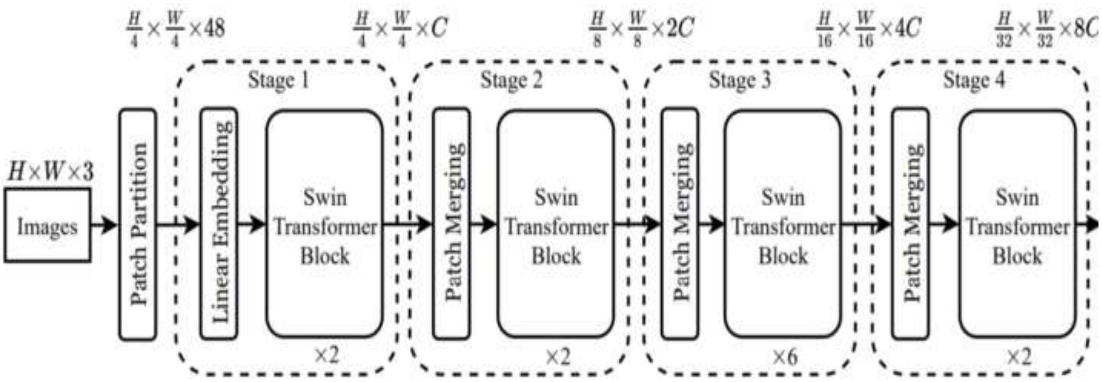


Figure 3 Obtained from [13] (a) The architecture of a Swin Transformer (Swin-T)

The features of each group of 2 X 2 neighboring patches are concatenated and a linear layer is applied to these concatenated features .this decreases the number of tokens by 2 X2 . the output dimension of the second transformer block is H/8 X w/8 X 2C. In Swin transformer block a layer normalization block, a shifted windows-based multi-shift attention block with a sip connection, then a layer normalization block and an MLP block with a skip connection is present. In the second Swin transformer block all blocks are the same as that of the first Swin transformer but instead of W_MSA, SW_MSA is present.

3. Proposed method

The block diagram of the IR fault classification system is shown in figure 4. The IR dataset images and the DCGAN-generated images are applied to the preprocessing block to reshape the images. the reshaped images are applied to the SWIN classifier to perform classification. Around 11 faults of IR images are classified based on self-head attention computation.

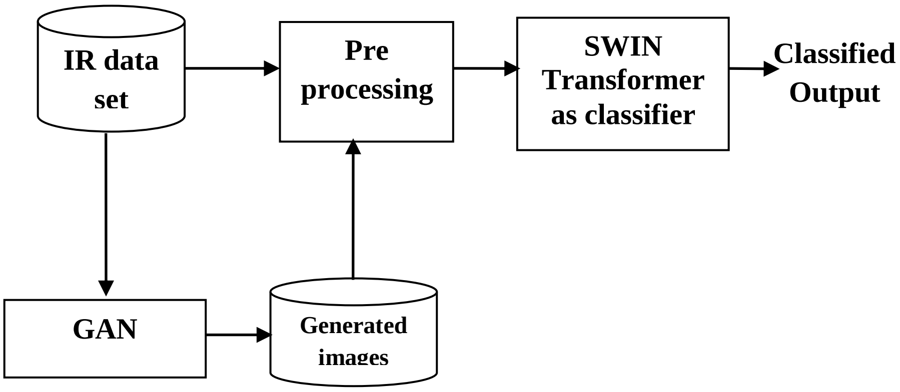


Figure 4. Block diagram of IR Fault classification system

3.1 Data set:

IR images are taken from the link https://github.com/RaptorMaps/InfraredSolarModules. 12 categories of infrared images are there in the dataset. It consists of 10000 normal images and 10,000 fault images. The fault images are grouped in one of the 11 fault categories. They are cell, cell-multi, hot-spot, hot-spot multi, diode, diode-multi, soiling, shadowing, vegetation, and off-line module. Thenumber of images in each fault category is not equal, the number of images in the diode-multi category is 175 and the number of images in the cell fault category is 1877. To achieve a uniform number of images in each fault category DCGANs' are used. The DCGANs generate images for each fault category to make the dataset balanced. Table .1 indicates the spread of images in every GAN-generated images and the bar diagram of image distribution in every fault category is shown in figure 5.

Table 1. The count of images in the IR dataset (in row1) and the count of images generated by GANs (row 2) for each fault category

Image source	Normal images	Number of images in each fault category										
		Cell	**Cell-Multi**	**Cracking**	**Hot spot**	**Hot spot multi**	**Diode**	**Diode-multi**	**Offline Module**	**Soiling**	**Shadowing**	**Vegetation**
IR dataset	10000	1877	1288	940	249	246	1499	175	827	204	1056	1639
GAN generated images	-	3123	3712	4060	4751	4754	3501	4825	4173	4796	3944	3361

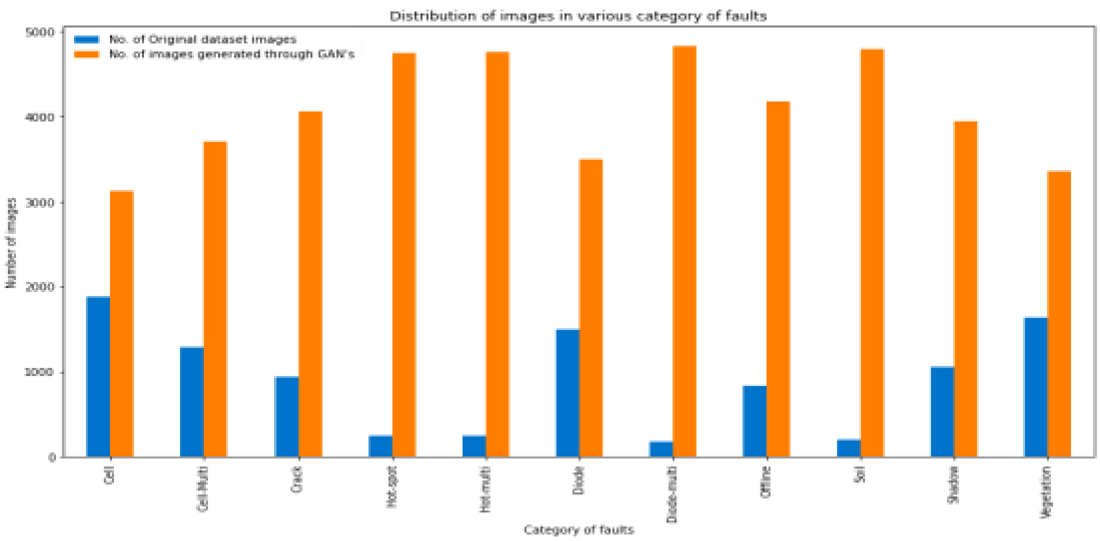


Figure 5. Distribution of images in original dataset and GAN generated dataset in every category of faults

 http://doi.org/10.36893/DRSR.2023.V13I4N13.059-070

3.2 Preprocessing

The IR dataset consists of 20,000 images of which 10,000 images belong to the normal category and the remaining 10,000 images are images of 11 different fault categories. The number of images in each of the fault categories is not the same, so to get the same number of images in each category GANs are used. The images of size 40 x 28 x 3 are reshaped to 32 x 32 x 3.

3.3 Self-attention computation in non overlapped windows

Local windows are alone used to find self-attention. If 2X 2 patch is formed in a window comprising of M X M patches and the image size is h X w

$$\Omega(MSA) = 4hwc^2 + 2(hw)^2c \qquad (2)$$

$$\Omega(W-MSA) = 4hwc^2 + 2(M)^2hwc \qquad (3)$$

Where $\Omega(MSA)$ of equation (2) is the computational cost incurred by MSA and $\Omega(W-MSA)$of equation (3) is the MSA obtained by the windowing approach. In The shifted window method connections are introduced between the non-overlapping windows lying in its neighborhood. the batch is shifted toward the top left direction cyclically. One of the drawbacks of the Swin method is the increase in the number of windows when shifting of windows is done, some windows size will not be equal to M XM size in that condition padding will be done to maintain the window size. Attention is calculated using equation (4). The padded values will be neglected during the calculation of attention.

$$attention(Q,K,V) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (4)$$

Where Q is the Query matrix, K is the Key matrix, V is the value matrix, and M^2 is the number of patches in a window. The features obtained from successive swin blocks with regular and shifted windows are found using (5).

$$\begin{aligned} \hat{z}l &= W-MSA\left(LN(z^{l-1})\right) + z^{l-1} \\ z^l &= MLP\left(LN(\hat{z}l)\right) + \hat{z}l \\ \hat{z}l+1 &= SW-MSA\left(LN(z^l)\right) + z^l \\ z^{l+1} &= MLP\left(LN(\hat{z}l+1)\right) + \hat{z}l+1 \end{aligned} \qquad (5)$$

Where $W-MSA$ and $SW-MSA$ indicates self attention of consecutive swin blocks.

3.4. Training phase

The dataset of IR images and gan generated images are represented as {(x1, y1), (x2, y2) …, (xn, yn)}, where x is the image from the IR dataset or GAN-generated image and y is the corresponding fault label, n is the number of IR images. a categorical cross-entropy loss as given in equation (7) is used. the CNN is trained to give a probability over the number of classes. for each image. one hot encoding of labels is done, so only the terms containing positive class Cp are present in the loss function. one element in the label vector will be non-zero. So the terms containing label elements whose values are zero in the summation are neglected.

$$CE = -log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right) \qquad (7)$$

s_p is the score of the CNN for positive fault class

4. Result and discussion

The input to the GAN network is the gray image of the IR dataset, The size of the input image is 40 X 28, The images under one fault category are applied to the GAN network and it generates images of that particular category. The number of images that has to be generated by the GAN is indicated in Table 1, The GAN model is used repeatedly 11 times to generate images for 11 fault category by considering one fault category at a time. The noise vector of dimension 100 is used as input to the GAN. The GAN generates gray images of size 40 X 28 at its output. These images are available in GAN_dataset. Figure 6 shows the images of 11 categories of faults generated using generative adversarial networks.The images available in the IR data set and GAN data set are reshaped to 32 X32 X3 and the images are normalized by dividing by 255. These images are grouped to train and test images in the ratio of 8:2. The image labels are the fault name in the dataset. The integer label values are obtained using the label encoder of preprocessing library. One hot encoding

 http://doi.org/10.36893/DRSR.2023.V13I4N13.059-070

is performed on the integer labels. The pixels of images are normalized and these normalized images are divided into train and test images using a train-test split.The structure of the classifier model with shifted window transformer is shown in figure 4. The random crop layer crops images from a randomly chosen location and the shape of the output of this layer are the same as the target size. The random flip layer performs a flipping operation on the input in the horizontal direction and produces a mirror image. The patch extract layer obtains patches from the input image by partitioning them. The patch embedding layer is the fully connected layer and it gives output based on the embedding dimension parameter (embed_dim). Since the embed_dim value is 64, the shape of the output of the patch embedding layer is 256 X 64. The Swin transformer block 1 computes self-attention of patches in the window, but the second Swin block computes self-attention on the shifted windows. The number of layers of the multi-layer perceptron num_mlp in the Swin blocks is set as 256. The fully connected layer is present in the patch merging layer whose output shape is 256 X 64. The patch merging layer will increase the number of channels by a factor of 2 and decrease the patch size. The parameter details of the Swin classifier are shown in Table 2. Only the patch size is varied and other hyper parameter values are kept constant.

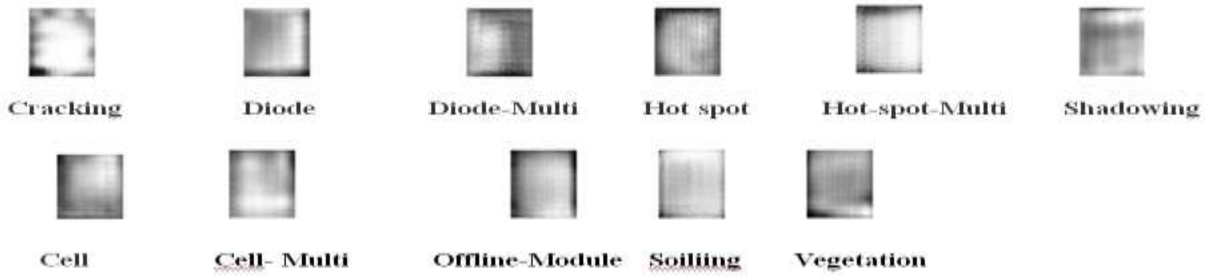


Figure 6. Representation of images of 11 fault categories generated through GAN's

Table 2. Training parameters of the Swin classifier model

Description	Value
learning_rate	1e-3
batch_size	128
validation_split	0.1
weight_decay	0.0001
label_smoothing	0.1
loss	Categorical cross-entropy
optimizer	AdamW optimizer

Table 3. Hyperparameters and their values

Symbol	**Description**	**Values**
patch_size	The size of the patch	2X2, 4x4, 8x8,12x12,16x16
dropout_rate	Neglecting nodes randomly during each updation	0.03
num_heads	The number of attention heads	8
embed_dim	The embedding dimension	64
num_mlp	The number of multi-layer perceptron layers	256
window_size	The size of the window	2
shift_size	The stride size	1

 http://doi.org/10.36893/DRSR.2023.V13I4N13.059-070

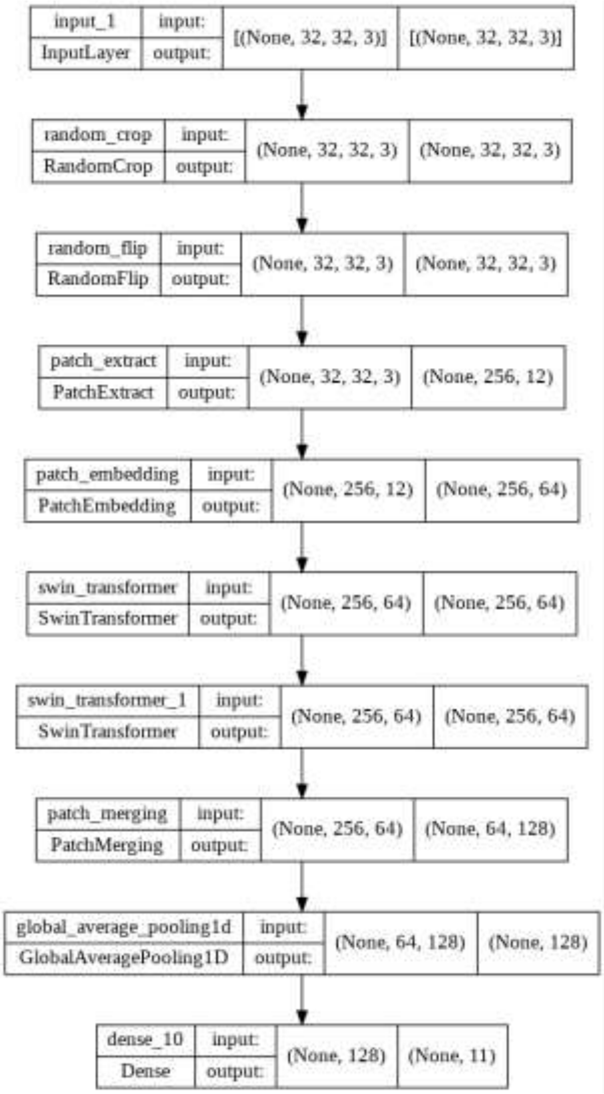


Figure 7. Structure of the classifier model with Swin blocks

Evaluation of the performance of the Swin classifier fault detection solution by comparing with existing models. The parameter values of the Swin classifier model during the training phase are specified in Table 2. the hyperparameters of the Swin classifier model are indicated in Table 3. The number of IR images of the training set, validation, and test dataset is shown in Table 4.

Table 4. The number of IR images in each dataset

Dataset (11 types of faults)	Dataset size (number of IR images)
Training set	43956 × 32 x 32 x3
Validation set	44 × 32 × 32 x 3
Test set	1000 × 32 x 32 x 3

The Swin classifier model code is written in python using the Tensor flow platform. The experiments are carried out in google colab and the specifications of the running environment of google colab are shown in Table 5.

Table 5. Details about the Google colab run time environment

Model name	Intel(R) Xeon(R) CPU @ 2.20GHz
Socket(s)	1
Core(s) per socket	1
Thread(s) per core	2
L3 cache	56320K

CPU MHz	2200.184
Hard disk space used	13GB

4.1. **Training phase assessment**

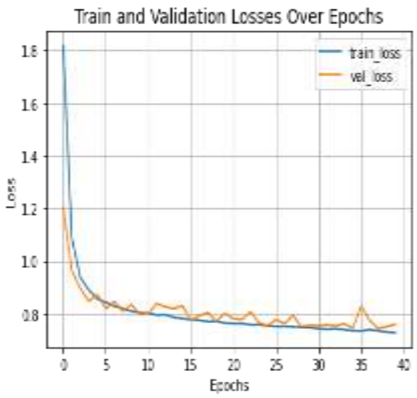
Fig. 8 and Fig.9 show the loss and accuracy curves of the training set and validation data set in detail versus the number of epochs while training the classifier for various values of patch size. Fig. 8 shows the significant decrease in losses for the training and testing datasets as the number of epochs increases and Fig 9 shows that accuracy increases at a faster rate for the training and testing datasets as the number of epochs increases, and the process of training is terminated at 40 epochs. Once the model is trained, the evaluation of the classification model is done with test images. From the figure, it is found that the difference between the accuracy of the training and validation datasets becomes small as the number of epochs increases, and only at a certain epoch, the difference is larger.

The loss, accuracy, and top-5-accuracy are measured during testing. The test accuracy, test top_5_accuracy, and testloss obtained for each value of patch size are indicated in Table 6. and their line plots are shown in figure 10. It can be observed from that table the time taken for training the Swin transformer is varying patch sizes. It is found that for the patch size of 8X8 highest accuracy is obtained by minimizing the test loss.

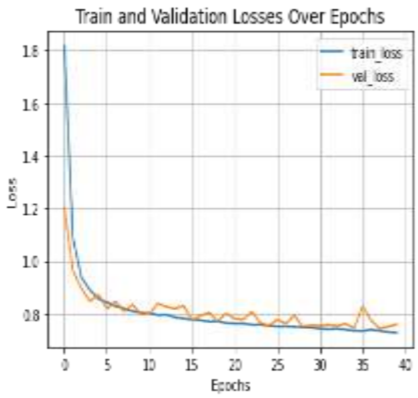
Table 6. Performance measures of Swin classifier for different values of patch size

Size of the patch	Test loss	Test accuracy	Test top_5_ accuracy	Training time /computationncost
16 x 16	0.75	91.09%	98.73%	10.5 min
12 x 12	0.75	90.34%	98.4%	11.43 min
8 x 8	**0.72**	**92.2%**	**99.04 %**	**15.2 min**
4 x 4	0.73	1.85%	98.7 %	85 min
2 x 2	0.78	90.49%	98.14%	332min

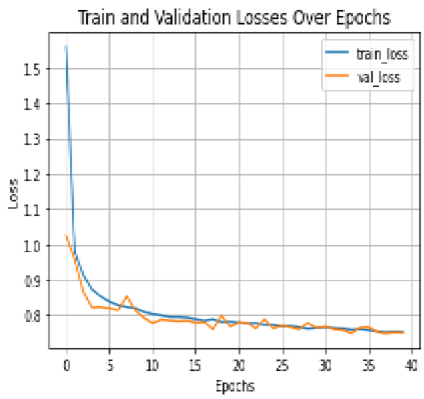
In [26] electroluminescent images (EL) are classified using CNN to detect the faults. the test accuracy of 0.83 is obtained by them while classifying 4 faults. In [27] CNN is used to classify 11 faults in infrared images. the test accuracy of 0.6373 is obtained by them.



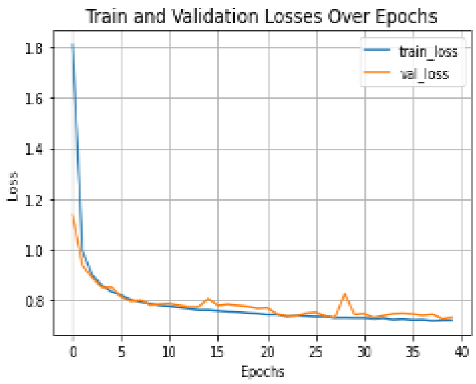
(a) 16X16 patch size



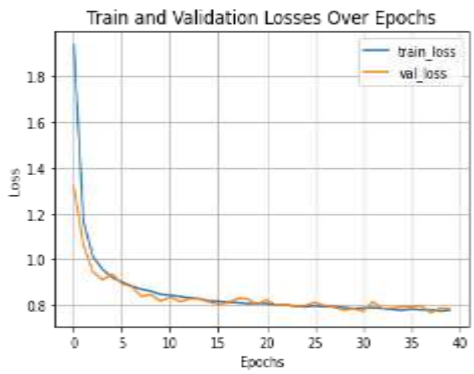
(b) 12 X 12 patch size



(c) 8X8 patch size

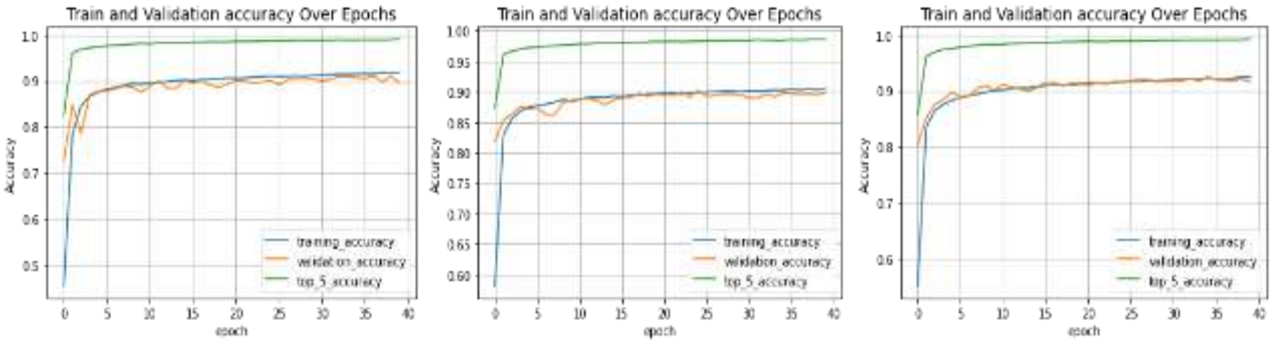


(d) 4 X 4 patch size

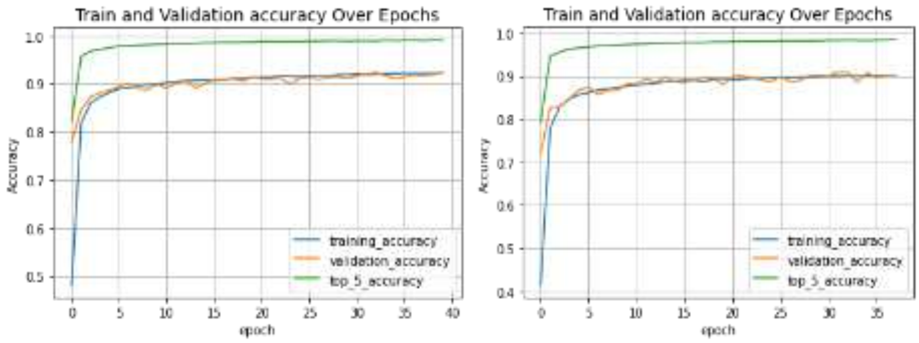


(e) 2 X 2 patch size

Figure 8. Train and Validation losses plots for different values of patch size



(a) 16 X 16 patch size **(b) 12X12 patch size** **(c) 8 X 8 patch size**



(d) 8 X 8 patch size **(e) 2 X2 patch size**

Figure 9. Train, Validation accuracy, Top-5-accuracy plots for different values of patch size

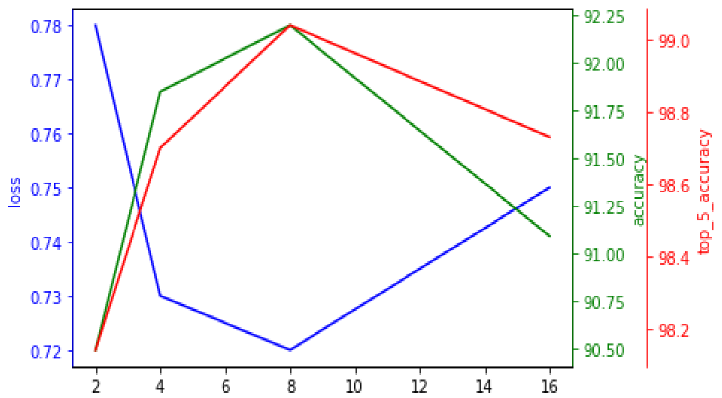


Figure 10. Loss, accuracy, top_5_accuracy against patch size plot

In this work by using the Swinclassifier 11 categories of faults are classified with a testing accuracy of 0.922. The comparison of accuracy obtained with the Swin classifier and that of existing methodsis given in table7.

Table 7. Comparison of accuracy of Swin classifier with existing methods

Image type	Number of faults	Method	Test Accuracy
EL image	4	CNN [26]	83 %
IR image	11	CNN [27]	63.73%
IR image	**11**	**Swin classifier**	**92.2%**

5. Conclusion

Swin Transformer is used for the classification of faults in IR images of the photovoltaic system since the computational complexity of it is varying linearly with image size. The generative adversarial network is used to generate images to make the dataset balanced.11 faults occurring in the PV module are classified by using the Swin transformer. One of the hyper parameters of the Swin transformer is the patch size. Training of Swin transformer is carried out for different values of patch size. The test accuracy obtained through this method is compared with that of the

 http://doi.org/10.36893/DRSR.2023.V13I4N13.059-070

classification results of CNN on EL images and CNN on IR images and it is found that the Swin classifier produces more promising results than other methods. For the patch size of 8X8 Swin transformer gives an accuracy of 92.2 %, and a top_10_accuracy of 99.04%. The computational time required to train the Swin transformer is 15.2 minutes for the patch size of 8X8, this time is very less when compared with CNN training time. By using new augmentation techniques and transformers a search for further improvement of accuracy can be done.

References

[1] Sinaga, Kristina P., and Miin-Shen Yang. "Unsupervised K-means clustering algorithm." *IEEE access* 8 (2020): 80716-80727.

[2] Cohn, Ryan, and Elizabeth Holm. "Unsupervised machine learning via transfer learning and k-means clustering to classify materials image data." *Integrating Materials and Manufacturing Innovation* 10.2 (2021): 231-244.

[3] K. Han *et al*., "A Survey on Vision Transformer," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2022.3152247.

[4] Hong, Cui, et al. "Deep-Belief-Networks Based Fault Classification in Power Distribution Networks." *IEEJ Transactions on Electrical and Electronic Engineering* 15.10 (2020): 1428-1435.

[5] Chen, Zhuyun, and Weihua Li. "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network." *IEEE Transactions on Instrumentation and Measurement* 66.7 (2017): 1693-1702.

[6] Huang, Jiabo, et al. "Unsupervised deep learning via affinity diffusion." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 07. 2020.

[7] Hong, Ying-Yi, and Rolando A. Pula. "Detection and classification of faults in photovoltaic arrays using a 3D convolutional neural network." *Energy* 246 (2022): 123391.

[8] Et-taleby, Abdelilah, Mohammed Boussetta, and Mohamed Benslimane. "Faults detection for photovoltaic field based on k-means, elbow, and average silhouette techniques through the segmentation of a thermal image." *International Journal of Photoenergy* 2020 (2020).

[9] Rao, S., Spanias, A., &Tepedelenlioglu, C. (2019). Solar array fault detection using neural networks. In *Proceedings - 2019 IEEE International Conference on Industrial Cyber Physical Systems, ICPS 2019* (pp. 196-200). [8780208] (Proceedings - 2019 IEEE International Conference on Industrial Cyber Physical Systems, ICPS 2019). Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/ICPHYS.2019.8780208.

[10] Harrou, Fouzi, et al. "An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine." *Solar Energy* 179 (2019): 48-58.

[11] Barun Basnet, Hyunjun Chun, Junho Bang, "An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems", *Journal of Sensors*, vol. 2020, Article ID 6960328, 11 pages, 2020. PV system, Solar Energy,Volume 179,2019,pp. 1-10.

[12] Jiahao Huang, Yingying Fang, Yinzhe Wu, Huanjun Wu, Zhifan Gao, Yang Li, Javier Del Ser, Jun Xia, GuangYang,Swin transformer for fast MRI,Neurocomputing,Volume 493,2022,Pages 281-304,ISSN 0925-2312,

[13] Liu, Ze and Lin, Yutong and Cao, Yue and Hu, Han and Wei, Yixuan and Zhang, Zheng and Lin, Stephen and Guo, Baining "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows",ar Xiv preprint arXiv:2103.14030,2021

[14] Chen, Haiyong, et al. "Solar cell surface defect inspection based on multispectral convolutional neural network." *Journal of Intelligent Manufacturing* 31.2 (2020): 453-468.

[15] Li, Xiaoxia, et al. "Intelligent fault pattern recognition of aerial photovoltaic module images based on deep learning technique." *J. Syst. Cybern.Inf* 16 (2018): 67-71.

[16] S. Fadhel, C. Delpha, D. Diallo, I. Bahri, A. Migan, M. Trabelsi, M.F. Mimouni, PV shading fault detection and classification based on I-V curve using principal component analysis: Application to isolated

[17] CeyhunKapucu, Mete Cubukcu,"A supervised ensemble learning method for fault diagnosis in photovoltaic strings",Energy,Volume227,2021.

[18] Chen, Chun-Fu Richard, Quanfu Fan, and Rameswar Panda. "Crossvit: Cross-attention multi-scale vision transformer for image classification." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
[19] Fan, Haoqi, et al. "Multi scale vision transformers." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
[20] Chen, Chun-Fu, Rameswar Panda, and Quanfu Fan. "Regionvit: Regional-to-local attention for vision transformers." *arXiv preprint arXiv:2106.02689* (2021).
[21] Zhang, Pengchuan, et al. "Multi-scale vision longformer: A new vision transformer for high-resolution image encoding." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
[22] X. Xie, H. Liu, Z. Na, X. Luo, D. Wang and B. Leng, "DPiT: Detecting Defects of Photovoltaic Solar Cells With Image Transformers," in *IEEE Access*, vol. 9, pp. 154292-154303, 2021, doi: 10.1109/ACCESS.2021.3119631.
[23] Liu, Ze, et al. "Swin transformer v2: Scaling up capacity and resolution." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
[24] Yuan, Li, et al. "Tokens-to-token vit: Training vision transformers from scratch on imagenet." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.
[25] Zhou, Daquan, et al. "Deepvit: Towards deeper vision transformer." *arXiv preprint arXiv:2103.11886* (2021).
[26] Tang W., Yang Q., Xiong K., *et al* ,"Deep learning based automatic defect identification of photovoltaic module using electroluminescence images" Sol. Energy, 201 (2020), pp. 453-460
[27] Ricardo Henrique Fonseca Alves, Getúlio Antero de Deus Júnior, Enes Gonçalves Marra, Rodrigo Pinto Lemos,"Automatic fault classification in photovoltaic modules using Convolutional Neural Networks",Renewable Energy, Volume 179, 2021, pp. 502-516,

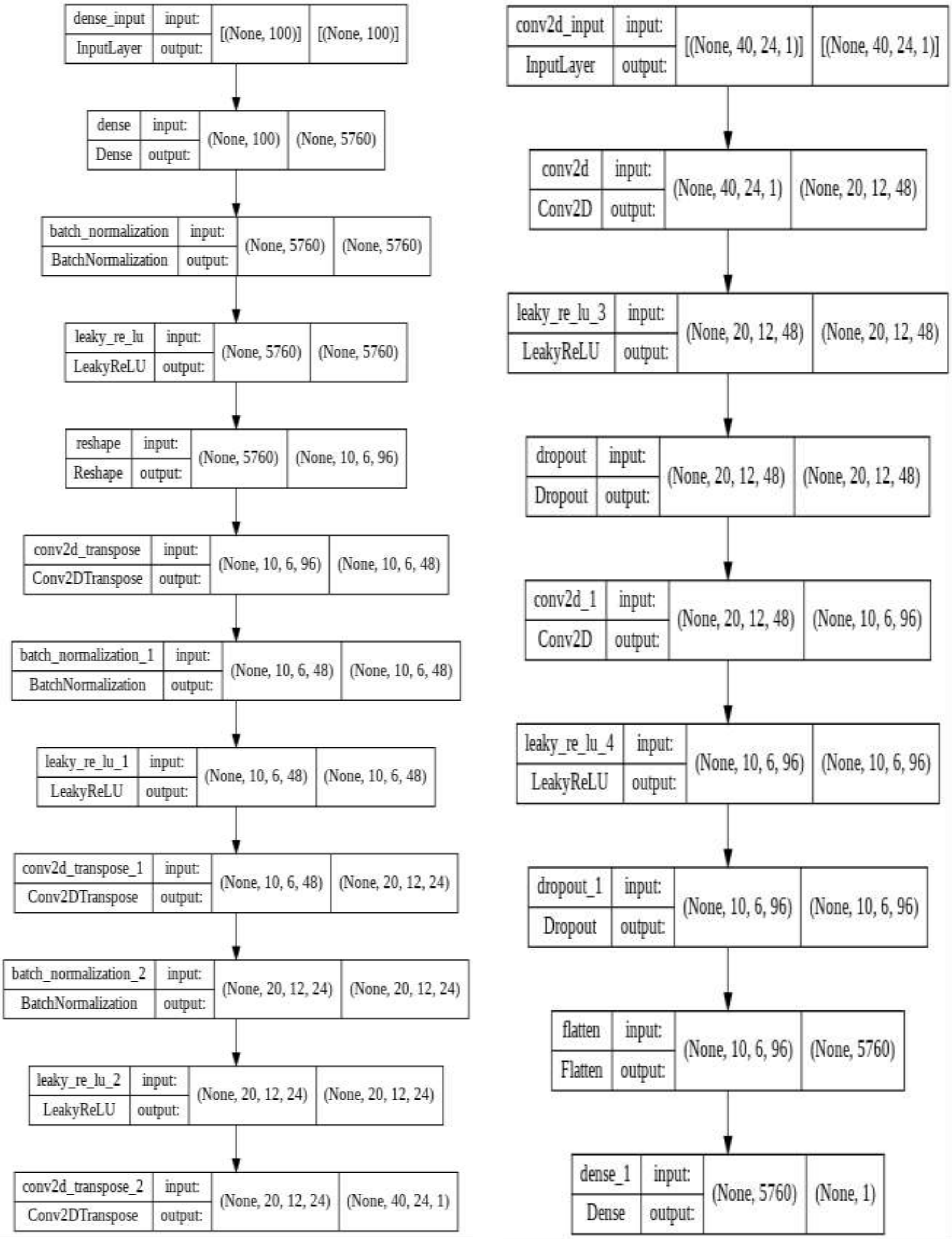


Figure 2. Structure of (a) generator network (b) discriminator network of GAN used to produce gray images.