



**UTM**  
UNIVERSITI TEKNOLOGI MALAYSIA

**INTERNATIONAL JOURNAL OF  
INNOVATIVE COMPUTING**

ISSN 2180-4370

Journal Homepage : <https://ijic.utm.my/>

# Diabetic Retinopathy Image Classification Using Shift Window Transformer

Rasha Ali Dihin  
Department of Computer Science  
University of Kufa  
Najaf, Iraq  
Email: rasha.aljabry@uokufa.edu.iq

Ebtesam N AlShemmary  
IT Research and Development Center  
University of Kufa  
Najaf, Iraq

Waleed A Mahmoud Al-Jawher  
Department of Electronic & Communication Eng.  
Uruk University  
Baghdad, Iraq

Submitted: 30/11/2022. Revised edition: 31/3/2023. Accepted: 31/3/2023. Published online: 13/9/2023  
DOI: <https://doi.org/10.11113/ijic.v13n1-2.415>

**Abstract**—Diabetic retinopathy is one of the most dangerous complications for diabetic patients, leading to blindness if not diagnosed early. However, early diagnosis can control and prevent the disease from progressing to blindness. Transformers are considered state-of-the-art models in natural language processing that do not use convolutional layers. In transformers, means of multi-head attention mechanisms capture long-range contextual relations between pixels. For grading diabetic retinopathy, CNNs currently dominate deep learning solutions. However, the benefits of transformers, have led us to propose an appropriate transformer-based method to recognize diabetic retinopathy grades. A major objective of this research is to demonstrate that the pure attention mechanism can be used to determine diabetic retinopathy and that transformers can replace standard CNNs in identifying the degrees of diabetic retinopathy. In this study, a Swin Transformer-based technique for diagnosing diabetic retinopathy is presented by dividing fundus images into nonoverlapping batches, flattening them, and maintaining positional information using a linear and positional embedding procedure. Several multi-headed attention layers are fed into the resulting sequence to construct the final representation. In the classification step, the initial token sequence is passed into the SoftMax classification layer, which produces the recognition output. This work introduced the Swin transformer performance on the APTOS 2019 Kaggle for training and testing using fundus images of different resolutions and patches. The test accuracy, test loss, and test top 2 accuracies were 69.44%, 1.13, and 78.33%, respectively for 160\*160 image size, patch size=2, and embedding dimension C=64. While the test accuracy was 68.85%, test loss: 1.12, and test top 2 accuracy: 79.96% when the patch size=4, and

embedding dimension C=96. And when the size image is 224\*224, patch size=2, and embedding dimension C=64, the test accuracy: 72.5%, test loss: 1.07, and test top 2 accuracy: 83.7%. When the patch size =4, embedding dimension C=96, the test accuracy was 74.51%, test loss: 1.02, and the test top 2 accuracy was 85.3%. The results showed that the Swin Transformer can achieve flexible memory savings. The proposed method highlights that an attention mechanism based on the Swin Transformer model is promising for the diabetic retinopathy grade recognition task.

**Keywords**—Diabetic retinopathy, Swin Transformers, Image classification

## I. INTRODUCTION

As one of the leading causes of blindness worldwide, Diabetic Retinopathy (DR) has to be taken seriously. People with diabetes are expected to increase in number in the foreseeable future because of increased life expectancy, decadent lifestyles, and other causes [1]. Diabetic retinopathy is a frequent consequence and a leading cause of blindness in the general population. When blood sugar levels are adequately controlled, and therapy is given on time, many DR problems may be avoided. Medical professionals recommend that diabetic people be checked at least twice a year since the condition progresses, and early indicators of sickness are challenging to detect [2, 3]. Diabetic Mellitus, a condition of sugar metabolism, affects one in eleven people worldwide, and by 2040, it is anticipated to affect one in

10 people [4]. By 2045, this problem is expected to impact more than half of the world's population, reaching 700 million [5, 6].

Diabetic retinopathy may progress through four stages: (i) non-proliferative retinopathy's initial stage, microaneurysms (MA), is linked to mild retinopathy. (ii) Non-proliferative retinopathy of moderate severity, as the condition progresses, the retina's blood vessels may twist and expand, rendering them ineffective in transporting blood. (iii) During severe non-proliferative retinopathy, the retina gets a signal to begin creating new blood vessels when more blood vessels in the retina are blocked. (iv) Excessive glucose levels in the blood may cause diabetic retinopathy (PDR), a condition in which the retina's development characteristics are released and new blood vessels proliferate in the vitreous gel, flooding the eye, (Fig. 1, and Table 1) [7].

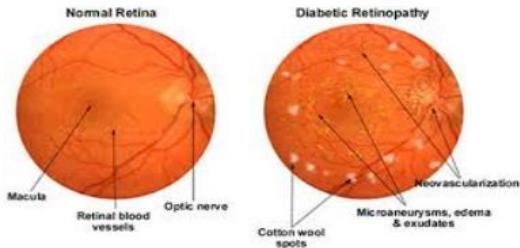


Fig. 1. Depicts a normal retina and the effects of diabetes on the retina

linear complexity with image size  $O(M * N)$  instead of the complexity found in ViT which is quadratic  $O(N^2)$  and thus the swin can be more efficient [9]. Furthermore, they also linked blocks of pay attention to these patch merge blocks, which are used to combine adjacent patches to produce a hierarchical representation to handle differences in the scale of visible entities [12]. The partitioning window in the second layer on the right is shifted by 2 image patches, resulting crossing in the boundary of the previous window [13].

There are two important concepts, hierarchical maps and window attention shifting, that the Swin Transformer introduces to solve problems for ViT. Where the name Swin Transformer comes from "Shifted window". Fig. 2 shows the structure of the Swin Transformer.

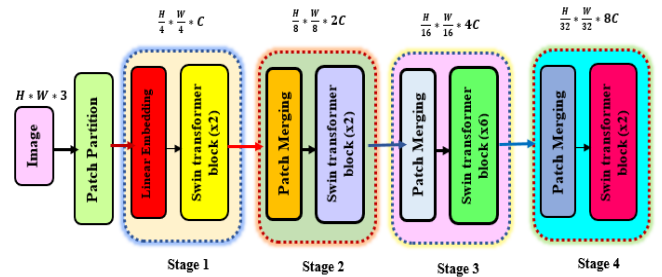


Fig. 2. The Architecture of the Swin Transformer

TABLE I. THE DR STAGES DEPENDING ON LESIONS CLASSIFICATION [7]

DR Severity	Level Lesions
No DR	No lesions.
Mild	DR Micro-aneurysms only.
Moderate	DR has more than MA and less than DR
Severe	more than 20 intraretinal HM in each of 4 quadrants
proliferative DR.	Proliferative DR One or more of the following: neovascularization, pre-retinal HM/ Vitreous

Automatic classification of DR plays a role in decision-making, ViT has gained great importance in the field of computer vision, however, studies include the use of computer vision transducers in the medical field. Vision transformer (ViT) was proposed for the first time to be used with machine translation tasks in natural language processing and had an advanced performance in this field because it learns local and The scalability of training with information is universal across different layers, which is the opposite of CNN, which has very limited sensory fields [8, 9]. Recently a pure transformer has been proposed called Shifted windows (Swin) Transformers. This model consists of several blocks of Swin transformers, providing a hierarchical representation of the input image which is then used in various computer vision tasks. The authors [10] achieved advanced performance in image classification, object detection, and semantic segmentation [11]. The possibility of applying transformers in the field of view of the calculator such as image recognition demonstrates the success of ViT and Swin Transformer [8-12], the architecture of the swin that uses local computing via non-overlapping windows and thus achieves

Architecture of Swin transformer consisting of 4 Stages:

### A. Patch Partitioning

The primary role of the Patch Partition is to convert the input image into correction blocks, where each of these blocks consists of four adjacent pixels. Where the image entered into Patch Partition, each  $(4 \times 4)$  adjacent pixel is divided into a patch, where each  $(4 \times 4)$  and three channels of the color image are converted into patches  $(1 \times 1)$  that are flat and have (48) channel as shown in Fig. 3. Since the size of each patch is  $4 \times 4$  so the number of pixels per patch will be  $4 \times 4 = 16$  in a flat shape, so that each of these pixels contains three values of R, G, and B, so it's  $16 \times 3 = 48$  as show in Fig. 4.

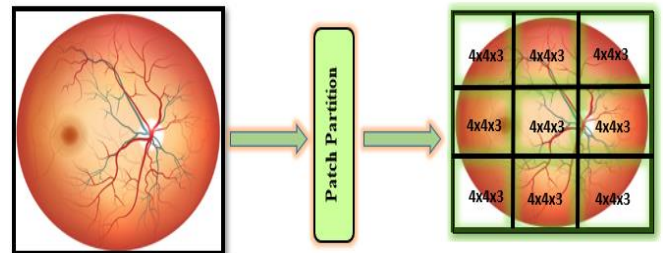


Fig. 3. Patch partition

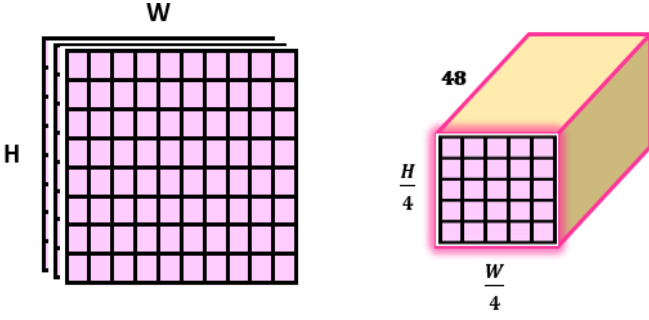


Fig. 4. Image from Patch partition

The input image passes through patch partition) which is divided into (4x4) sized falls. This helps in creating debug symbols, where after patch partition the shape of the image is changed to be whose shape is (W/4, H/4, × channel) = (W/4, H/4, × 3).

**B. Linear Embedding**

Linear Embedding performed after the patch partition step, which implement a linear transformation on the image channel. This step is executed to convert the input tokens and the output tokens to vectors as well as. The feature in this layer looks like a convolution layer, the number of channels in which the convolution mapping is done is from 48 dimensions to 96 dimensions. On the application side, patch embed is used to combine steps patch, partition and linear mbedding where convolution, kernel is used.

**C. Swin Transformer Block**

Swin Transformer is designed by replacing the multi-head MSA to a switched window-based module with a switched window-based module, this layer has been swapped but the rest of the layers remain without any replacement as shown in Fig. 5.

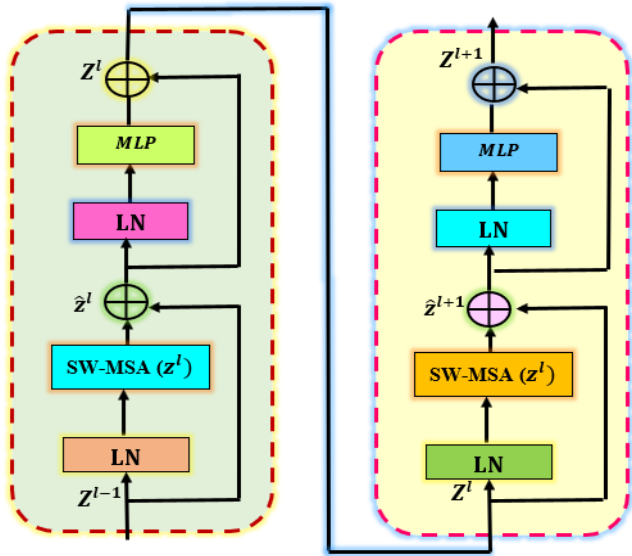


Fig. 5. Swin transformer blocks

The Swin transformer block contains two sub-modules, where the first module uses W1-MSA and in the second module “SW-1MSA” is used. Each of these sub-layers contains a normalization layer [14-16]. Each W-MSA layer is followed by a 2-layer MLP with GELU nonlinearity between them where LN is applied before both MSA and MLP. Patch tokens pass through the linear embedding with size image (W/4, H/4, C) and are referred as “Stage 1”. The number of 2 \* 2 patch tokens is reduced and the shape of the tokens is (W/8, H/8, 2C) is denoted as “Stage 2”, (W/16, H/16, 4C) as “Stage 3”, and (W/32, H/32, 8C) as “Stage 4” respectively. Equations (1-4) state the mathematical expression of W-MSA and SW-MSA as follows [17]:

$$z^{\wedge l} = W - MSA(LN(z^{l-1})) + z^{l-1} \tag{1}$$

$$z^l = MLP(LN(z^{l-1})) + z^{\wedge l} \tag{2}$$

$$z^{\wedge l+1} = SW - MSA(LN(z^l)) + z^l \tag{3}$$

$$z^{l+1} = MLP(LN(z^{\wedge l+1})) + z^{\wedge l+1} \tag{4}$$

Where (z ^ l) is the item in the current block, (z l – 1) is the item in the previous block, (LN) is layer-norm, (MLP) is multi-layer perceptron, (W1-1MSA) is window self-attention, and (SW1-1MSA) is shift window self-attention [15].

**D. Patch Merging**

The main function of patch merging is to reduce both the height and width of the image by sampling downwards to reduce the accuracy. This is before the start of each phase and this function is equivalent to the down sampling process that is in CNN. In the patch merging stage, each (2x2) of the adjacent pixels is divided into a patch, and the pixels of the same colour are grouped to get 4 feature maps. These four feature maps are linked in the depth direction, the output from here being passed through a layer (LN) and layer (FC) is used to linearly change the depth direction of the feature map, whereby the feature map depth is changed from C to C/2 as shown in Fig. 6.

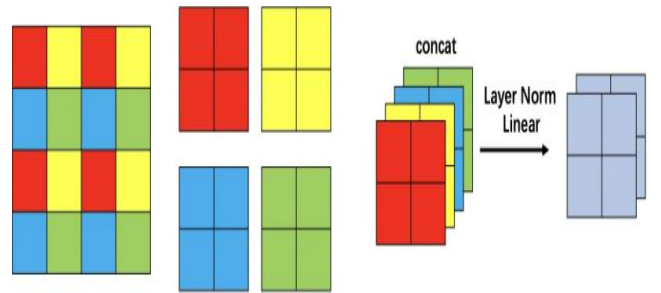


Fig. 6. Patch Merging

**II. RELATED WORK**

Cao et al. proposed a Swin-UNet, where the Swin Transformer block used as a basic unit in building the model for segmentation of medical images. A benchmarking comparison

was presented for the development of transformers in the field of medical image, where the (Swin-Unet) has a patch expansion layer to map the decoder's feature maps and shows superior performance in restoring fine details compared to binary down sampling [8].

Huang *et al.* suggested a SwinMR which is a new Swin-based method used for rapid MRI reconstruction. The built model was consisting of an input unit (IM), a feature extraction unit (FEM), and an output unit (OM). When applying SwinMR, it achieved excellent results for high reconstruction. Quality compared to other measurement methods under noise interruption and on different data sets [17].

Jiang *et al.* contemplated a new approach called SwinBTS for 3D medical image segmentation. This approach combines a Transformer, convolutional neural network, and decoder architecture to determine the 3D brain tumor semantic segmentation. Experimental results showed that the method gives better performance in segmenting MRI images of a brain tumor when compared with some state-of-the-art methods (for example "Residual U-Net", "Attention U-Net", and "TransBTS") [18-21].

Hao *et al.* Which is a Swin converter that is two-stream and used in the classification of remote sensing images, and TSTNet consists of two vertical parts which are the original stream and the second part is the edge stream, and through them the features are combined, and it has achieved good performance and also designed a driver unit called DESOM based on Sobel To extract the features in the edge and thus give us a better rating [22].

### III. EXPERIMENTS SETTINGS

#### A. Datasets

The APTOS 2019 (Asia Pacific Teleophthalmology Society) Kaggle benchmark dataset This rule contains images of the retina that were taken using fundus imaging, and the conditions that were used in the imaging were very diverse, and this rule was used in the challenge of detecting blindness This data has been manually classified by specialists into 5 classes (0 to 4) where "0" means no DR; "1" means Mild1; "2" means 1Moderate; "3" means Severe1; and "4" means Proliferative1

DR2) to indicate different severity levels of DR [23]. Table II shows the number of retinal images in the dataset to indicate the level1 of meverity, [24-30].

TABLE II. DATASET SUMMARY OF APTOS DATASET

Severity level	Number of images
Class 0 (Normal )	1805
Class 1(Mild )	370
Class 2 (Moderate )	999
Class 3 ( Severe)	193
Class 4 (Proliferative )	295
Total	3662

#### B. Implementation details

Used Swin transformer to automatically recognize Diabetic Retinopathy progression level because the Swin transformer can achieve flexible memory savings. Where used GPU with 5.71 GB memory/12.68 GB Disk, we set the input image size as  $160 \times 160$  and  $224 \times 224$ . Here in the Table III below is the combination of hyperparameters we are chosen after carefully tuning each of them across a wide range of values.

Table IV, Fig. 7 and Fig. 8 shows the loss and accuracy across the training and validation process with input image size  $160 \times 160$  when Patch size =2, C=64, Patch size =4, C=96 and Table V, Fig. 9 and Fig. 10 shows the loss and accuracy across the training and validation process with input image size  $224 \times 224$  when Patch size =2, C=64, Patch size =4, C=96.

TABLE III. HYPER-PARAMETERS IN SWIN TRANSFORMER TRAINING

Hyper-parameter	Value	
Batch Size	16	
Learning Rate	0.05	
Size of shifting window	1	
Size of attention window	2	
Epoch	80	
Weight decay	1e-3	
optimizer	Adam	
Patch size	2	4
Embedded dimension	64	96
# Param	987,381	567,445

TABLE IV. CLASSIFICATION ACCURACY AND LOSS (IMAGE SIZE (160\*160))

Epoch	Patch size =2, C=64						Patch size =4, C=96					
	Train-Acc	Train-loss	top-2-acc	Val-Acc	Val-loss	Val-top-2-acc	Train-Acc	Train-loss	top-2-acc	Val-Acc	Val-loss	Val-top-2-acc
1	0.4564	11.7449	0.6646	0.4125	8.2687	0.4669	0.4537	23.350	0.6786	0.6786	3.7794	0.7857
10	0.6612	1.1450	0.7935	0.5992	1.3052	0.7860	0.6914	1.0644	0.8272	0.6865	1.0510	0.8095
20	0.7059	1.0276	0.8312	0.6498	1.1441	0.7821	0.7381	0.9463	0.8690	0.7421	0.9478	0.8611
30	0.7323	0.9323	0.8716	0.6459	1.1282	0.8132	0.8961	0.8961	0.8832	0.6905	1.0901	0.7778
40	0.7488	0.9079	0.8781	0.6498	1.0653	0.7665	0.8100	0.8016	0.9114	0.6984	1.0232	0.8135
50	0.7714	0.8586	0.8924	0.6654	1.1918	0.7977	0.8395	0.7601	0.9290	0.7024	1.0530	0.8373
60	0.8208	0.7717	0.9150	0.6654	1.2200	0.7704	0.8395	0.7513	0.9317	0.7063	1.0252	0.8294
70	0.8499	0.7265	0.9380	0.6732	1.2898	0.7588	0.8642	0.7142	0.9484	0.6905	1.1238	0.8492
80	0.8803	0.6785	0.9458	0.6576	1.2626	0.7743	0.8849	0.6690	0.9563	0.6984	1.0851	0.7976



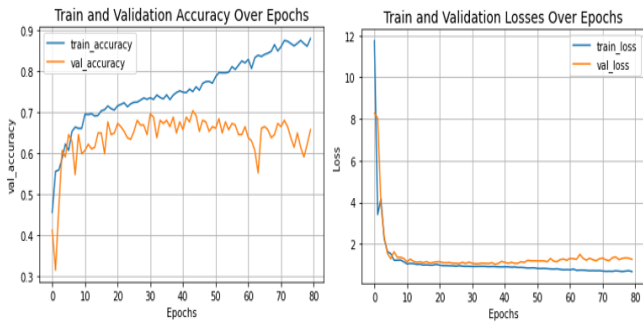


Fig. 7. Training and validation over epoch for APTOS 2019 dataset, (a) accuracy, (b) loss, (epochs=80), image size (160×160), patch size =2, C=64

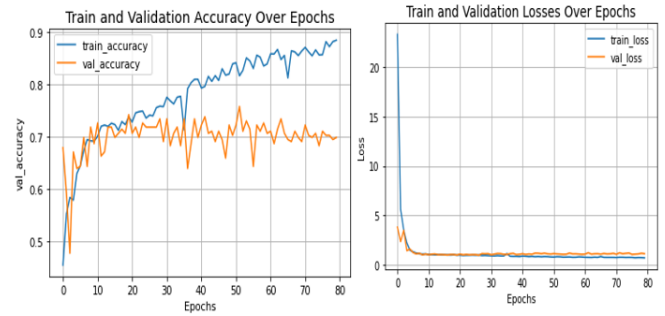


Fig. 8. Training and validation over epoch for APTOS 2019 dataset, (a) accuracy, (b) loss, (epochs=80), image size (160×160), patch size =4, C=96

TABLE V. CLASSIFICATION ACCURACY AND LOSS (IMAGE SIZE (224\*224))

Epoch	Patch size =2, C=64						Patch size =4, C=96					
	Train-Acc	Train-loss	top-2-acc	Val-Acc	Val-loss	Val-top-2-acc	Train-Acc	Train-loss	top-2-acc	Val-Acc	Val-loss	Val-top-2-acc
1	0.5518	5.7511	0.6460	0.6187	4.4559	0.4280	0.4684	26.3868	0.6667	0.4757	7.6186	0.7431
10	0.6993	1.0575	0.8325	0.5759	1.2574	0.7549	0.6987	1.0220	0.8387	0.6389	1.0630	0.8507
20	0.7128	0.9727	0.8521	0.7043	1.0055	0.8210	0.7184	0.9704	0.8642	0.7604	0.9229	0.8681
30	0.7401	0.9212	0.8725	0.7198	0.9782	0.8171	0.7357	0.9529	0.8611	0.7604	0.9818	0.8229
40	0.7761	0.8509	0.9007	0.7082	1.0046	0.8132	0.7828	0.8561	0.8939	0.7222	1.0437	0.8438
50	0.7879	0.8458	0.8998	0.6887	1.0596	0.8560	0.8233	0.7796	0.9240	0.7674	0.9862	0.8576
60	0.8204	0.7826	0.9150	0.7043	1.0609	0.8171	0.8465	0.7415	0.8465	0.7326	0.9955	0.8438
70	0.8811	0.6802	0.9423	0.6965	1.1076	0.8288	0.8623	0.7167	0.9444	0.7361	1.1023	0.8299
80	0.8810	0.6801	0.9393	0.7043	1.0609	0.8289	0.8781	0.6865	0.9533	0.7500	1.0727	0.8750

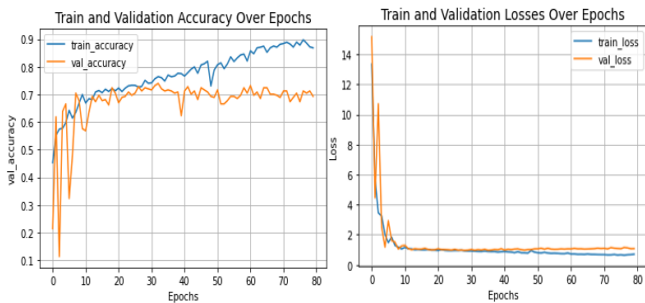


Fig. 9. Training and validation over epoch for APTOS 2019 dataset, (a) accuracy, (b) loss, (epochs=80), image size (224×224), patch size =2, C=64

#### IV. DISCUSSION

We tested model on APTOS-2019 Blindness Detection dataset whose data distribution between different classes is unbalanced, pre-training the model affects the performance of the transformer-dependent model. In this work, we use two different input size of image (160\*160) and (224\*224), patch size is 2 and 4, and embedding dimension is 64 and 96. In the case of selecting the patch size 4 and embedding dimension 96, the model give performance better than the patch size 2 and embedding dimension 64 for each input image sizes where the number of parameters in this case was 567,445 but in the second case was 987,381. Other parameters can be changed, as well as the change in the Swin transformer block to see its effect on the results.

#### V. CONCLUSION

Transformer has made great technical advances in deep learning and has achieved wide spread in the field of NLP and CV, because medical imaging is very similar to “CV”. Since the Swin transformer has great flexibility in modelling and its computational complexity is linear complexity proportional to the size of the image. This work used the Swin transform on APTOS 2019 Kaggle. It was also used to classify the DR to 5 class, The results demonstrated that the Swin transform

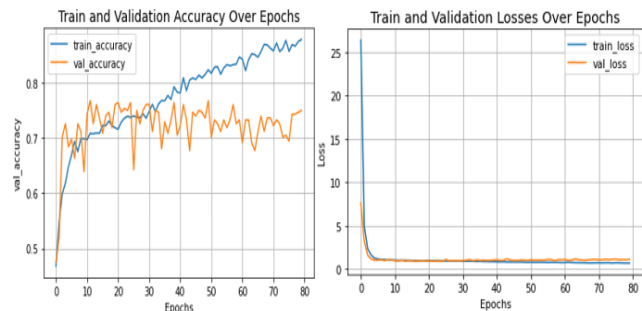


Fig. 10. Training and validation over epoch for APTOS 2019 dataset, (a) accuracy, (b) loss, (epochs=80), image size (224×224), patch size =4, C=96.

performs well in classifying DR with linear computational complexity as compared to the ViT transformer's quadratic computational complexity. Test accuracy for size image 160 \*160, patch size =2, embedding dimension C=64 was 69.44%, the Test loss: 1.13, and Test top 2 accuracy: 78.33%, while when the patch size =4, embedding dimension C=96, the Test loss was 1.12, Test accuracy: 68.85% and Test top 2 accuracy: 79.96%. For the size image 224\*224, patch size =2, embedding dimension C=64 the Test loss: 1.07, Test accuracy: 72.5% and Test top 2 accuracy: 83.7%, while when the patch size =4, embedding dimension C=96 the Test loss: 1.02, Test accuracy: 74.51% and Test top 2 accuracy: 85.3%. From the results we can conclude that it is possible to change other parameters, changing the Swin transformer block or combining it with deep learning algorithms that can give better results for automatically Diabetic Retinopathy recognition. Since the architecture of transformers is still quite new together in the field of CV and It is possible in the future to discover other variants of the attention layers, and also plan to use the hypercomplex to detect the mixing of symbols.

#### ACKNOWLEDGMENTS

The research behind this paper would not have been possible without the exceptional support of Prof. Ebtesam AlShemmary and Prof. Waleed AlJawher. Throughout researching image processing journals and writing this paper, they have inspired me with their enthusiasm, knowledge, and attention to detail.

#### REFERENCES

- [1] Li, X., Pang, T., Xiong, B., Liu, W., Liang, P., & Wang, T. (2017, October). Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. *2017 10th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI)* (pp. 1-11). IEEE.
- [2] Waheed, S. R., Suaib, N. M., Rahim, M. S. M., Adnan, M. M., & Salim, A. A. (2021, April). Deep learning algorithms-based object detection and localization revisited. *Journal of Physics: Conference Series* (Vol. 1892, No. 1, p. 012001). IOP Publishing.
- [3] Salim, A. A., Ghoshal, S. K., Suan, L. P., Bidin, N., Hamzah, K., Duralim, M., & Bakhtiar, H. (2018). Liquid media regulated growth of cinnamon nanoparticles: Absorption and emission traits. *Malaysian Journal of Fundamental and Applied Sciences*, *14*(3-1), 447-449.
- [4] Adnan, M. M., Rahim, M. S. M., Al-Jawaheri, K., Ali, M. H., Waheed, S. R., & Radie, A. H. (2020, September). A survey and analysis on image annotation. *2020 3rd International Conference on Engineering Technology and its Applications (IICETA)* (pp. 203-208). IEEE.
- [5] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., ... & Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, *135*, 104599.
- [6] Salim, A. A., Ghoshal, S. K., & Bakhtiar, H. (2022). Prominent absorption and luminescence characteristics of novel silver-cinnamon core-shell nanoparticles prepared in ethanol using PLAL method. *Radiation Physics and Chemistry*, *190*, 109794.
- [7] Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., & Qian, W. (2019, July). An interpretable ensemble deep learning model for diabetic retinopathy disease classification. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2045-2048). IEEE.
- [8] Abbas, S. I., Hathot, S. F., Abbas, A. S., & Salim, A. A. (2021). Influence of Cu doping on structure, morphology and optical characteristics of SnO<sub>2</sub> thin films prepared by chemical bath deposition technique. *Optical Materials*, *117*, 111212.
- [9] Tang, Y., Yang, D., Li, W., Roth, H. R., Landman, B., Xu, D., ... & Hatamizadeh, A. (2022). Self-supervised pre-training of swin transformers for 3d medical image analysis. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20730-20740).
- [10] Waheed, S. R., Sakran, A. A., Rahim, M. S. M., Suaib, N. M., Najjar, F. H., Kadhim, K. A., Salim A. A. & Adnan, M. M. (2023). Design a crime detection system based fog computing and IoT. *Malaysian Journal of Fundamental and Applied Sciences*, *19*(3), 345-354.
- [11] Nguyen, C., Asad, Z., Deng, R., & Huo, Y. (2022, April). Evaluating transformer-based semantic segmentation networks for pathological image segmentation. *Medical Imaging 2022: Image Processing* (Vol. 12032, pp. 942-947). SPIE.
- [12] He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., ... & Shen, D. (2022). Transformers in medical image analysis: A review. *Intelligent Medicine*.
- [13] Salim, A. A., Bakhtiar, H., Bidin, N., & Ghoshal, S. K. (2018). Unique attributes of spherical cinnamon nanoparticles produced via PLAL technique: Synergy between methanol media and ablating laser wavelength. *Optical Materials*, *85*, 100-105.
- [14] Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., & Xu, D. (2022, July). Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021, Virtual Event, September 27, 2021, Revised Selected Papers, Part I* (pp. 272-284). Cham: Springer International Publishing.
- [15] Aljewaw, O. B., Karim, M. K. A., Kamari, H. M., Zaid, M. H. M., Salim, A. A., & Mhareb, M. H. A. (2022). Physical and spectroscopic characteristics of lithium-aluminium-borate glass: Effects of varying Nd<sub>2</sub>O<sub>3</sub> doping contents. *Journal of Non-Crystalline Solids*, *575*, 121214.
- [16] Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., & Zhang, D. (2022). Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-15.
- [17] Huang, J., Fang, Y., Wu, Y., Wu, H., Gao, Z., Li, Y., ... & Yang, G. (2022). Swin transformer for fast MRI. *Neurocomputing*, *493*, 281-304.
- [18] Jiang, Y., Zhang, Y., Lin, X., Dong, J., Cheng, T., & Liang, J. (2022). SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer. *Brain Sciences*, *12*(6), 797.
- [19] Jiang, Y., Zhang, Y., Lin, X., Dong, J., Cheng, T., & Liang, J. (2022). SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer. *Brain Sciences*, *12*(6), 797.
- [20] Aldhuhaibat, M. J., Amana, M. S., Aboud, H., & Salim, A. A. (2022). Radiation attenuation capacity improvement of various oxides via high density polyethylene composite reinforcement. *Ceramics International*, *48*(17), 25011-25019.
- [21] Al-Jawher, W. A. M., & Awad, S. H. (2022). A proposed brain tumor detection algorithm using Multi wavelet Transform (MWT). *Materials Today: Proceedings*, *65*, 2731-2737.

- [22] Hao, S., Wu, B., Zhao, K., Ye, Y., & Wang, W. (2022). Two-stream swin transformer with differentiable sobel operator for remote sensing image classification. *Remote Sensing*, 14(6), 1507.
- [23] Waheed, S. R., Rahim, M. S. M., Suaib, N. M., & Salim, A. A. (2023). CNN deep learning-based image to vector depiction. *Multimedia Tools and Applications*, 1-20.
- [24] Bodapati, J. D., Naralasetti, V., Shareef, S. N., Hakak, S., Bilal, M., Maddikunta, P. K. R., & Jo, O. (2020). Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. *Electronics*, 9(6), 914.
- [25] Abbas, A. M., Abid, M. A., Abbas, K. N., Aziz, W. J., & Salim, A. A. (2021, April). Photocatalytic activity of Ag-ZnO nanocomposites integrated essential ginger oil fabricated by green synthesis method. *Journal of Physics: Conference Series* (Vol. 1892, No. 1, p. 012005). IOP Publishing.
- [26] Hathot, S. F., Jubier, N. J., Hassani, R. H., & Salim, A. A. (2021). Physical and elastic properties of TeO<sub>2</sub>-Gd<sub>2</sub>O<sub>3</sub> glasses: Role of zinc oxide contents variation. *Optik*, 247, 167941.
- [27] Salim, A. A., Bakhtiar, H., Shamsudin, M. S., Aziz, M. S., Johari, A. R., & Ghoshal, S. K. (2022). Performance evaluation of rose bengal dye-decorated plasmonic gold nanoparticles-coated fiber-optic humidity sensor: A mechanism for improved sensing. *Sensors and Actuators A: Physical*, 347, 113943.
- [28] Kadhim, K. A., Najjar, F. H., Waad, A. A., Al-Kharsan, I. H., Khudhair, Z. N., & Salim, A. A. (2023). Leukemia classification using a convolutional neural network of AML Images. *Malaysian Journal of Fundamental and Applied Sciences*, 19(3), 306-312.
- [29] Waheed, S. R., Saadi, S. M., Rahim, M. S. M., Suaib, N. M., Najjar, F. H., Adnan, M. M., & Salim, A. A. (2023). Melanoma skin cancer classification based on CNN Deep learning algorithms. *Malaysian Journal of Fundamental and Applied Sciences*, 19(3), 299-305.
- [30] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022).