Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Wavelet-Attention Swin for Automatic Diabetic Retinopathy Classification

Rasha Ali Dihin [1], ***Ebtesam N. AlShemmary**** [2], ***Waleed A. M. Al-Jawher*** [3]

[1]Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Kufa, Iraq.
[2]IT Research and Development Center, University of Kufa, Kufa, Iraq.
[3]Uruk University, Baghdad, Iraq.
*Corresponding Author.

Received 11/02/2023, Revised 12/06/2023, Accepted 14/06/2023, Published Online First 20/01/2024

Abstract

Diabetic retinopathy (DR) is a complication of diabetes that affects the eyes by damaging the blood vessels in the retina. High blood sugar levels can cause leakage or blockage of these vessels, leading to vision loss or blindness. Early detection of DR is crucial to prevent blindness, but manually analyzing fundus images can be time-consuming, especially with a large number of images. Swin-Transformers have gained popularity in medical image analysis, reducing calculations and yielding improved results. This paper introduces the WT Attention-Db5 Block, which focuses attention on the high-frequency domain using Discrete Wavelet Transform (DWT). This block extracts detailed information from the high-frequency field while retaining essential low-frequency information. The study discusses findings from the 2019 Blindness Detection challenge (APTOS 2019 BD) held by the Asia Pacific Tele-Ophthalmology Society.The proposed WT-Swin model achieves significant improvements in classification accuracy. For Swin-T, the training and validation accuracies are 99.14% and 98.91%, respectively. For binary classification using Swin-B, the training accuracy is 99.01%, the validation accuracy is 99.18%, and the test accuracy is 98%. In multi-classification, the training and validation accuracies are 93.19% and 86.34%, respectively, while the test accuracy is 86%.In conclusion, early detection of DR is essential for preventing vision loss. The WT Attention-Db5 Block integrated into the WT-Swin model shows promising results in classification accuracy.

Keywords: APTOS Data Set, Diabetic Retinopathy, Swin-B, Swin-T, Wavelet-Attention.

Introduction

The eye is one of the main parts affected by diabetes and thus leads to diabetic retinopathy[1]. It is considered one of the main complications of diabetes, which does not show early symptoms, as the blood vessels in the eye are damaged, due to the high level of sugar in the blood, thus leading to swelling of the vessels[2]. The advanced stage of this disease leads to blindness and thus may be considered the main cause of blindness in adults[3].

DR is associated with two types of diabetes. For the disease type 1, the number of patients is about 75-95% of those who have had diabetes for at least 15 years. With type 2 disease, about 60% of these patients have had diabetes for more than 16 years[4].

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

DR disease is at its beginning without any symptoms in the early stages of it, but when its stages progress, symptoms begin to appear noticeably. Therefore, detection of the disease in its early stages makes treatment more beneficial and effective and thus may prevent the development of the disease[5].

Discrete wavelet transformation is considered one of the methods in which a conversion is made from the time domain to the frequency domain. It is also used in compression clouds, including JPEG 2000. Usually, the waves are functions and these functions are integrated with the zero waves above and below the x-axis[6]. It can reduce spectral noise, but it can preserve spectral details, and it can also be used to reduce local noise, as well as in analyzing information with different bands and thus rebuilding the spectrum to its original form after decomposition[7].

Swin Transformers can be considered the backbone of general-purpose computer vision and has performed well on many tasks including object detection, semantic segmentation, and image classification. The main idea of Swin is to use the hierarchy as well as many prefixes in the encoder to the adapter and the locale and thus all of these help in visual tasks[8].

As it consists of several windows arranged hierarchically, the presence of these windows gives high efficiency by reducing the calculations for self-attention to local windows that are not overlapping in nature, and yet it allows communication through these windows[9].

In this work, a novel Wavelet Attention WTA-Db5 Block is currently being developed by the team, which will selectively capture essential high-frequency information without affecting low-frequency information. Furthermore, a new method called Wavelet-Attention (WTA-Db5-Swin) is being proposed, which will utilize the Wavelet-Db5-Attention Block to extract detailed image data more accurately. This will result in improved efficiency in image classification accuracy, with Swin-T and Swin-B versions of the Swin Transformer being adopted as the default backbone for optimal performance. As the number of patients with diabetes and DR has risen, diagnosing DR has become more challenging in recent years. This has led to an increase in the number of cases that go undiagnosed and untreated. Early detection and treatment of DR are more cost-effective than late or incorrect diagnosis. To address this issue, the study focused on using wavelet attention methods to extract features from fundus images. They then used the Swin Transformer instead of CNN deep learning to speed up the diagnosis process and reduce training and testing time.

The proposed approach contributes to the development of new theories related to the use of wavelet analysis and attention mechanisms in medical image analysis, specifically for the diagnosis of the diabetic retina, and improves the efficiency of existing models that can be further extended and applied to other medical imaging tasks. However, the key contributions of the paper are summarized as follows:

- The proposed approach combines wavelet analysis and attention mechanisms to improve the accuracy of automatic diabetic retinopathy classification.
- The proposed WT Attention-Db5 Block extracts detailed information from the high-frequency domain based on DWT, while preserving basic information in the low-frequency domain, leading to better results and reduced computational burden.
- The proposed approach uses Swin-Transformers, which have the advantage of reducing computation while providing better results, to improve the accuracy of diabetic retinopathy classification.
- The proposed approach is evaluated on the APTOS 2019 dataset and achieves a high accuracy of 98% for binary classification and 86% for multi-classification.

Related Work

Li H, et al. [10], proposed a new technology called DnSwin, which is used to reduce noise, by integrating WSWT, in which the bottom features are extracted from the image first, through the use of CNN. DnSwin is used to extract high and low-frequency information. Through the results, the

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

method has proven that it can reduce noise in the real world. It has a speed when compared with Euclidean-based protocols.

Xiangyu Zhao Wavelet-Attention (WA-CNN) proposal that is used for image classification, where WA-CNN is used in image analysis to extract high and low-frequency features, as well as through which detailed information and noise can be obtained [11]. Through the experimental results when applying the method to CIFAR-10 and CIFAR-100, it has been proven that it achieves good results in obtaining high classification accuracy.

Sabiha G K. et al.[12], proposed a new method that relies on a deep feature generator based on correction, and this method is inspired by ViT. Both ViT (Vision Transformer) used MLP-mixer, which uses a fixed-size square patch to extract features. In this way, rectangular patches were used instead of square patches and copies of these were used. Patches to create deeply hidden patterns, DenseNet201 was used that has 201 layers and trained on the ImageNet dataset for image classification tasks. The method used achieved good results with high accuracy of more than 90% for classification (Normal, NPDR, and PDR) on the APTOS 2019 dataset.

Danny C. et al.[13], transformed UNet (PCAT-UNet), and this unit depends on drawing attention and is in the shape of the letter U, and it is based on a transformer, and to combine the features, a skip connection was used on both sides and the results showed that the proposed method gives good results in segmenting the retinal blood vessels on both datasets (DRIVE, STARE, CHASE_DB1).

Gupta, et al.[14], developed an Optimal Deep Convolutional Neural Network for Retinal Fundus Image Classification (ODCNN-RFIC) model using pre-processing techniques, image segmentation, feature extraction, and a mayfly optimization with kernel extreme learning machine (MFO-KELM) classification model. The proposed method outperformed existing algorithms in terms of accuracy and performance.

Atwany, et al. [15], reviewed and analyzed state-of-the-art deep learning methods in supervised, self-supervised, and Vision Transformer setups, proposing retinal fundus image classification and detection. For instance, referable, non-referable, and proliferative classifications of Diabetic Retinopathy are reviewed and summarized. Moreover, the paper discusses the available retinal fundus datasets for Diabetic Retinopathy that are used for tasks such as detection, classification, and segmentation. The paper also assesses research gaps in the area of DR detection/classification and addresses various challenges that need further study and investigation.

AlShemmary and Omran[16] proposed a method for detecting pupils in eye images using a combination of morphological operations and Hough Transform. The method converts the local iris area into a rectangular block to calculate inconsistencies in the image. There is a potential relation between the method for detecting pupils in eye images using a combination of morphological operations and Hough Transform and diabetic retinopathy classification, as diabetic retinopathy can cause changes in the retinal blood vessels and may result in changes in the shape and size of the pupil. The method for detecting pupils in eye images can be used to identify abnormalities in the pupils of diabetic retinopathy patients.

Jaskari, et al.[17], presented novel results for 9 BNNs by investigating a clinical dataset and a 5-class classification scheme, as well as benchmark datasets and a binary classification scheme. A novel uncertainty measure is also proposed, which improves performance on some datasets. The findings suggest that BNNs can be utilized for uncertainty estimation in classifying diabetic retinopathy on clinical data, but proper uncertainty measures are needed to optimize performance, and methods developed for benchmark datasets might not generalize to clinical datasets.

Zia, et al.[18], proposed a computerized learning model utilizing deep neural networks that have the potential to accurately detect key precursors of Diabetic Retinopathy (DR) from retinal images. By combining the strengths of selected models (VGG and Inception V3) and using an entropy concept to select the most discriminant features, the model can classify features such as enlarged veins, liquid dribble, exudates, hemorrhages, and miniaturized scale aneurysms into different classes and determine the severity level of DR in diabetic retinopathy images. This model can be a useful tool for the accurate diagnosis and treatment of patients with DR.

Ashour [19], highlighted the effectiveness of Artificial Neural Networks (ANN) in time-series applications, particularly Back Propagation and Recurrent neural networks, in solving linear, semi-linear, and non-linear time series. The study employed forecast skill

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

(SS), mean square error, and absolute mean square error to measure the efficiency and accuracy of the estimation methods used. The study found that RBF neural networks were less efficient and accurate in solving nonlinear time series, but showed good efficiency in the case of linear or semi-linear time series. Overall, the study provides insights into improving modern methods for time series forecasting.

Materials and Methods

In this section, the global context-modeling framework is first introduced, followed by a detailed discussion of the design shown in Fig. 1.

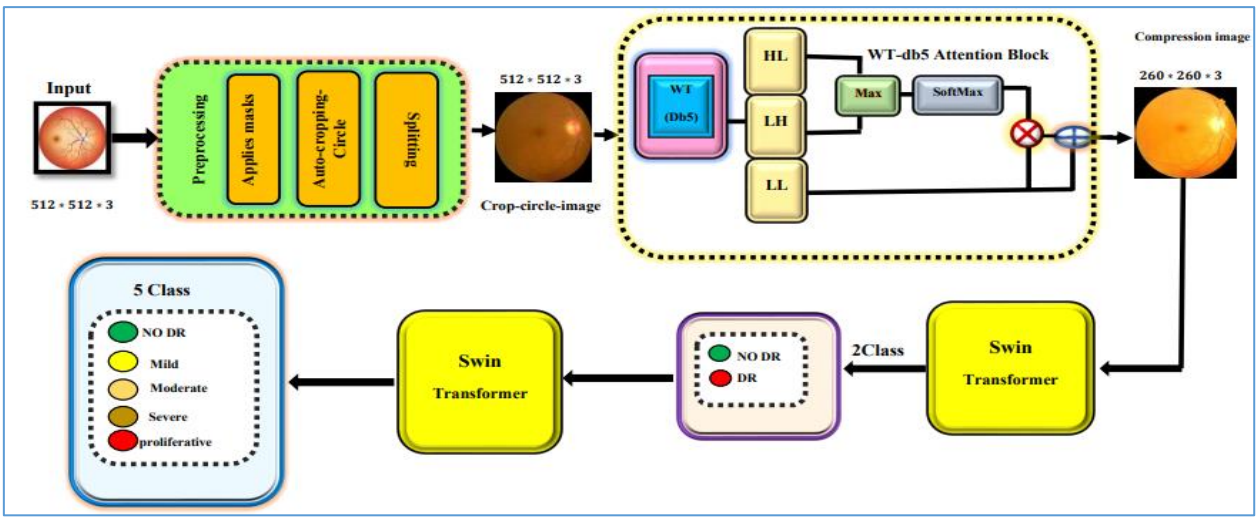

Figure 1. Proposed framework.

The Swin Transformer is a state-of-the-art deep learning architecture that has been shown to achieve high accuracy on a wide range of computer vision tasks, including image classification. In the context of DR classification for the APSTOS dataset, the Swin Transformer has been used to address several gaps in the literature, including:

- Limited data availability: The APSTOS dataset is relatively small compared to other medical image datasets. This can make it difficult to train accurate deep-learning models. The Swin Transformer's ability to efficiently model long-range dependencies and its scalability to larger datasets has helped to address this limitation.
- Variability in image quality: The APSTOS dataset contains images captured using different cameras and imaging conditions, resulting in variability in image quality. The Swin Transformer's self-attention mechanism can selectively focus on relevant image regions and suppress noise and artifacts in the input image, improving the model's robustness to variability in image quality.
- Limited interpretability: Deep learning models can be difficult to interpret, making it challenging to understand how the model arrives at its predictions. The Swin Transformer's architecture includes a hierarchical attention mechanism that can highlight the most informative image regions for the classification task, improving the interpretability of the model.

The Swin Transformer's ability to model long-range dependencies, handle variability in image quality, and improve interpretability has made it a promising approach for DR classification on the APSTOS dataset. A new approach to feature extraction using a technique called Wavelet attention (WTA) is introduced. This technique is inspired by the wavelet transform, which is a mathematical tool commonly used in signal processing and image analysis. Wavelet attention (WTA) allows the Swin Transformer to capture

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

multi-scale features from an image, which is particularly important for DR classification where important features may be present at different scales in the retina. Traditional convolutional neural networks (CNNs) typically rely on a fixed-size receptive field to capture features, which can limit their ability to capture information at multiple scales. WTA works by decomposing the input image into a set of wavelet coefficients at multiple scales. These coefficients are then used as inputs to the attention mechanism, which selectively weighs each coefficient based on its relevance to the classification task. This allows the model to focus on important features at multiple scales, improving its ability to capture complex and subtle patterns in the image. The use of WTA in the Swin Transformer has been shown to improve the performance of DR classification on the APSTOS dataset, compared to traditional CNNs and other state-of-the-art deep learning models. This demonstrates the effectiveness of the approach in addressing gaps in the literature related to feature extraction for DR classification.

Pre-processing

Retina images are susceptible to various issues, such as inconsistent image sizes caused by the use of cameras with varying aspect ratios and heights. This inconsistency affects the image quality and may result in the appearance of black areas around the eye, which do not provide any useful information for diagnosis. To improve the model's performance, it is necessary to standardize the retinal images. This can be achieved by cropping a circular area containing only the eyeball and removing all the pixels around the eye that are not relevant.

One approach to cropping the black areas is to convert the image to grayscale to identify the black areas accurately, based on their pixel density. A mask can then be created by defining the rows and columns, which can help remove the vertical and horizontal black rectangles that may appear in the upper right areas of the image. Afterward, the image can be resized to the desired width and height, denoted as R.

Another critical issue is the shape of the eye, which can vary from circular to oval. To standardize the shape of the eye, a circular crop can be made around the center of the image. These steps help ensure that all input images are similar, which is necessary for our Swin framework that requires input images of (RxR x3).

The original images in the dataset are high-resolution color images with different sizes (R1 x R2 x3) captured by various cameras. To obtain the desired input size for our Swin framework, the pixels are cropped from the right and left sides of each original image to achieve a square shape and remove any non-relevant parts, as illustrated in Fig. 2. Cropping and resizing the images was carried out in the following steps:

Step 1: Find the height of the image (R1) and width of the image (R2).
Step 2: Crop a part from left (cpr_left) and right (cpr_righ) for each of the images, as in Eq. 1, and Eq. 2

$$cpr_{\_left} = \frac{R1 - R2}{2} \quad \ldots\ldots 1$$

$$cpr_{\_right} = R1 - \left(\frac{R1 - R2}{2} + R1\right) \quad \ldots\ldots 2$$

Step 3: Resize the resulting image from step 2 to (512 x 512).
Step 4: Repeat the steps on all dataset images.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

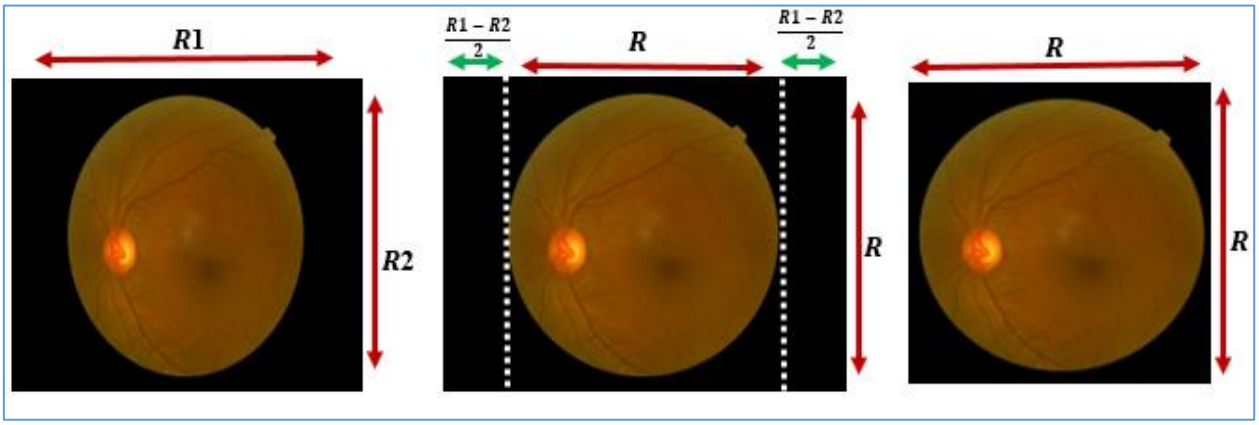

Figure 2. Circular crop.

Analysis of Method

In this section, the principle of Discrete Wavelet Transform (DWT) is first introduced, followed by a detailed discussion of the Wavelet-Db5-Attention Block design.

Discrete Wavelet Transform

Wavelet transform (WT) is a mathematical method used in the signal analysis in the field of signal processing that uses a set of orthogonal waves[11]. It has been widely used recently in various fields, including noise reduction, pressure, and analysis[20]. The basic idea of a wave transform is that any function can be represented as a superposition of a group of waves, which thus forms the basic function of the transform, and it uses wavelets as functions localized in both time and frequency. Different times can be utilized to capture local features in the frequency domain, which enables the extraction of information in the time and frequency domains simultaneously[21]. DWT is used to analyze the data into several different components and at different frequency intervals as well, and this helps in image processing because it adapts the separate data[11]. There are many basic functions in this conversion, including Haar wave Daubechies (db) and Symlets (sym), and other waveform transformers[22]. The 2D-DWT can be used in digital image processing, as it converts the input image into a set of low-frequency information and high-frequency information that may be vertical, horizontal, or diagonal, as shown in Fig. 3 [20]. In this study, the wavelets that provide the best classification performance are obtained as "db5" for the Daubechies wavelet family.

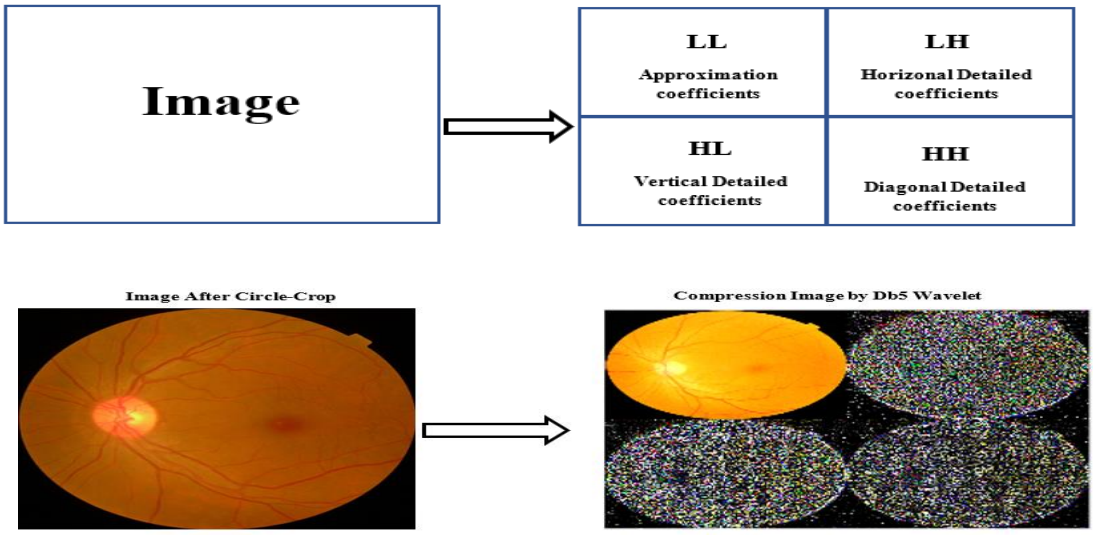

Figure 3. Retina image transformation by wavelet transforms family (db5).

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Wavelet-Db5-Attention Block

Swin Transformer with WTA discovered a novel approach to feature extraction that significantly improves the accuracy of DR classification on the APSTOS dataset. By leveraging the multi-scale information captured by the wavelet transform and the attention mechanism of the Swin Transformer, the authors achieved state-of-the-art performance on the dataset, surpassing previous deep learning models and human experts. This discovery highlights the potential of the Swin Transformer with WTA for accurate and efficient DR classification, which has important implications for the early detection and treatment of diabetic retinopathy, a leading cause of blindness worldwide.

This study is significant in several ways. Firstly, it demonstrates the potential of using the Swin Transformer with WTA to improve the accuracy of DR classification on the APSTOS dataset. Secondly, the use of WTA for feature extraction in the Swin Transformer is a novel approach that has not been extensively explored in the literature on DR classification. Thirdly, the authors' approach outperformed previous deep learning models and human experts, suggesting its potential to improve the efficiency and accuracy of DR diagnosis. Finally, the study highlights the potential for deep learning approaches to address gaps in the literature related to DR classification, paving the way for future research. The structure of the proposed WT-db5 block is shown in Fig. 4.

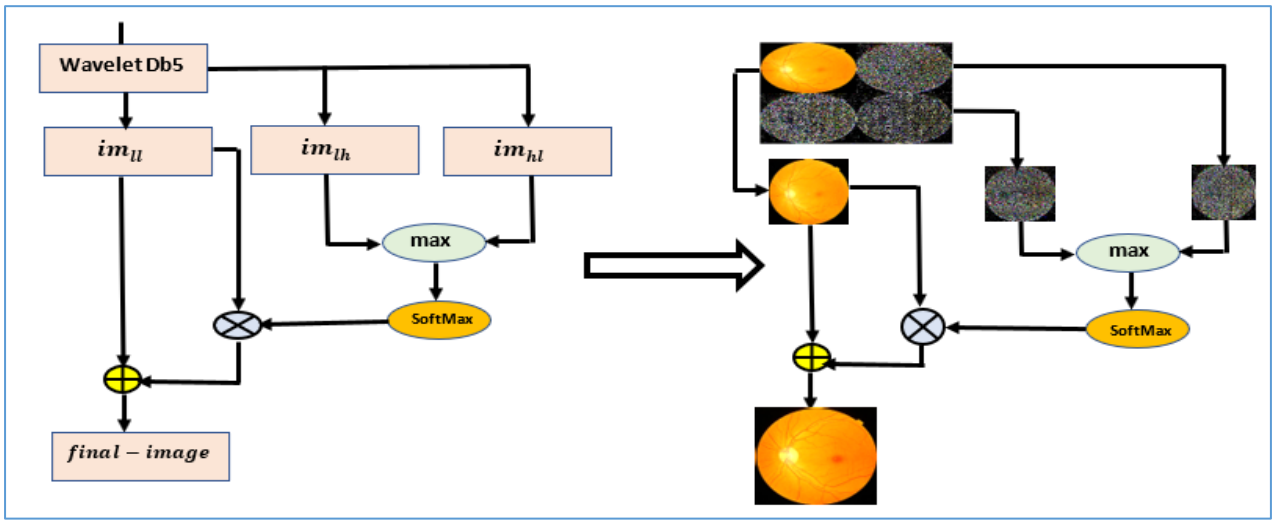

Figure 4. WT Attention-Db5 Block.

The WT-db5 block performs DWT on the image (im) to obtain a low-frequency component $\boldsymbol{im_{ll}}$ and $\boldsymbol{im_{lh}, im_{hl}, and\ im_{hh}}$ have three high-frequency components. Take the $\boldsymbol{im_{lh}, im_{hl}}$. Where when comparing low frequencies with high frequencies, where low frequencies contain the basic information of the image and can preserve it from damage, while high frequencies only contain a lot of noise, but they retain the detailed information from the image. Since the WT-db5 block can be defined as in Eq. 3:

$$R = F(\boldsymbol{im_{ll}}, \boldsymbol{\delta r}\left(\boldsymbol{im_{ll}}, \boldsymbol{Sm}\left(\boldsymbol{max}(\boldsymbol{im_{lh}}, \boldsymbol{im_{hl}})\right)\right) \ldots\ldots\ldots \mathbf{3}$$

The function F (·, ·) collects the final features of the image, while Sm (.) refers to the SoftMax and δr (,) is used to create an attention map. The WT-db5 block is utilized to extract the horizontal feature from image $\boldsymbol{im_{lh}}$ as well as the vertical feature $\boldsymbol{im_{hl}}$. The max operation is then applied to each feature to obtain a comprehensive detail feature. Next, Sm is applied to each $\boldsymbol{im_{lh}}$ and $\boldsymbol{im_{hl}}$ to normalize the global detail feature of the images. Multiplication of the SoftMax results with the low-frequency component is performed, and finally, the component $\boldsymbol{im_{ll}}$ is added to the map of interest.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Results and Discussion

The results of the study on DR classification using the Swin Transformer with WTA have several potential benefits, for the research community, clinicians, and healthcare systems. This study improves the accuracy of DR diagnosis; deep learning models like the Swin Transformer with WTA have the potential to facilitate earlier detection and treatment, ultimately reducing the burden of disease. Deep learning models can process large amounts of data quickly and accurately, potentially reducing the workload of clinicians and enabling more efficient use of resources. Additionally, the interpretability of the Swin Transformer with WTA may make it easier for clinicians to understand and trust the model's predictions, increasing its utility in clinical practice.

The proposed architecture was developed using a software package (Python), and the implementation was specific to central processing units (CPUs). All experiments were conducted on Google Colaboratory (Colab) using a 15G Graphics Processing Unit (GPU). The use of a GPU allowed for faster computations and improved performance compared to using only a CPU. It also enables to train of larger models and process larger datasets, which was essential for achieving the research goals. Table 1 concludes the parameters set for the proposed model.

Table 1. The setting of hyper parameter.

Hyper parameter	Setting
Input Size	224*224
Training Splitting Ratio	Train:80%, Val:10%, Test:10%
Batch Size	128
Epoch size	100
Learning Rate	0.001
Dropout	0.3
Optimizer	Adam

Datasets

The APTOS 2019 dataset, created by the Asia Pacific TeleOphthalmology Society (APTOS) and used in the challenge of detecting blindness, consists of fundus imaging of the retina with diverse imaging conditions[23]. The dataset has been manually classified into five severity levels of Diabetic Retinopathy (DR) by specialists, ranging from 0 (no DR) to 4 (Proliferative DR), with varying levels of severity in between; where “1” means Mild1; “2” means Moderate; and “3” means Severe[24, 25].

Research on the Different Classifications

Binary Classification

The training model for DR binary classification with wavelet attention WTA using Swin Transformer (Swin-T) and (Swin-B) was performed for 100 epochs. During each epoch, the model was trained on the training dataset using (Adam) optimizer with a learning rate of 0.001, momentum of 0.9, and weight decay of 0.0005. The learning rate was reduced by a factor of 0.1 after every 30 epochs.

In each epoch, the training and validation loss and accuracy were calculated. The training and validation datasets were processed using the WTA module for feature extraction. The images were classified into two categories, normal or with DR, using Swin-T. The model underwent 46 minutes and 21 seconds of training, achieving a best validation accuracy of 0.9973 at epoch 100. The final training accuracy was 0.9920, and the final validation loss was 0.2205. On the other hand, Swin-B was trained for 76 minutes and 57 seconds, with a best validation accuracy of 0.9973 at epoch 100. The final training accuracy was 0.9911, and the final validation loss was 0.2170. Fig. 5 displays part of the training process for DR binary classification using Swin-T and Swin-B.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

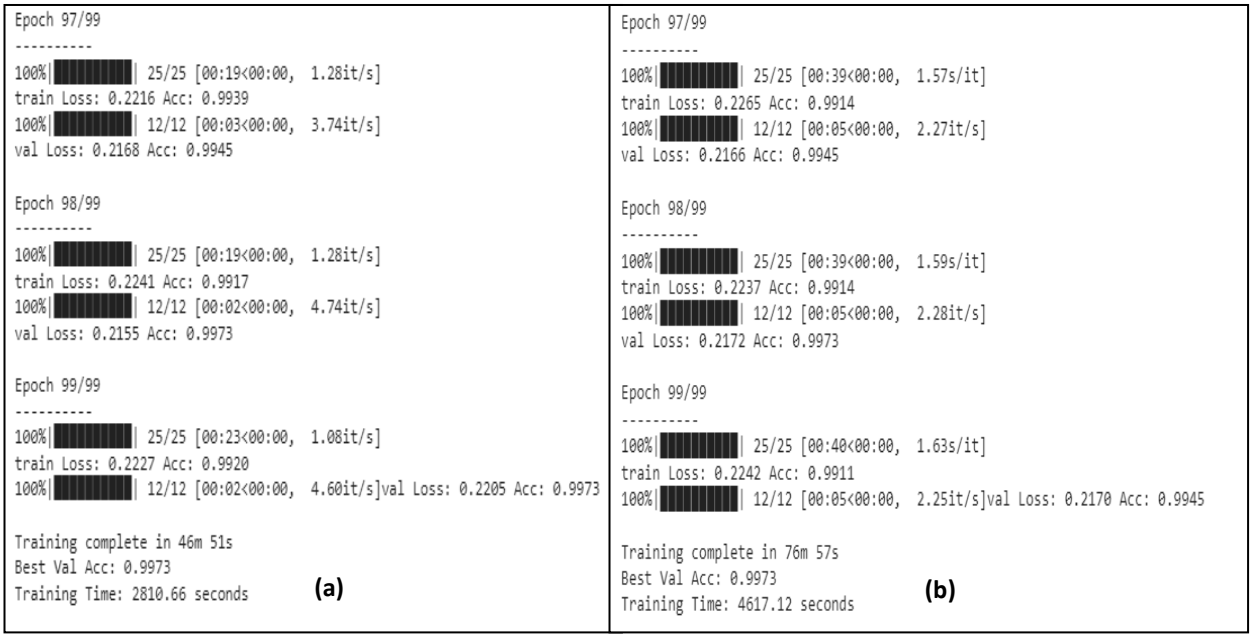

Figure 5. Part of the training process implementation for DR binary classification using (a) Swin-T, and (b) Swin-B.

To improve the performance of the swin_tiny (Swin-T) and swin_base (Swin-B) models, the WT Attention-Db5 Block was applied. This block utilized wavelet (db5) on the pre-processed image resulting from the circular crop, which had a size of 260, 260. The results of binary classification using this block are presented in Table 2 and Table 3, and the visual representation of the training process is shown in Fig. 6 and Fig. 7.

For binary classification of DR, the Swin-T model achieved impressive results as shown in Table 3, with an accuracy of 98% on the test dataset, including 97% accuracy for images without DR. The test loss was also low at 0.0102. Similarly, the Swin-B model performed exceptionally well, achieving a test accuracy of 98% with a test loss of 0.0079. This indicates that the model can generalize well and avoid overfitting.

Table 2. Classification binary class accuracy and loss for Swin-T & Swin-B.

	Swin-T				Swin-B			
Epochs	Train-loss	Train-acc	Val-loss	Val-acc	Train-loss	Train-acc	Val-loss	Val-acc
10	0.3593	0.9147	0.2847	0.9699	0.3942	0.8813	0.2650	0.9836
20	0.2556	0.9736	0.2320	0.9891	0.2446	0.9793	0.2427	0.9809
30	0.2444	0.9815	0.2298	0.9891	0.2370	0.9850	0.2319	0.9836
40	0.2393	0.9825	0.2305	0.9891	0.2342	0.9850	0.2283	0.9863
50	0.2360	0.9847	0.2323	0.9891	0.2297	0.9892	0.2265	0.9863
60	0.2346	0.9873	0.2304	0.9863	0.2274	0.9901	0.2229	0.9945
70	0.2292	0.9879	0.2214	0.9918	0.2280	0.9911	0.2219	0.9918
80	0.2279	0.9901	0.2203	0.9891	0.2248	0.9908	0.2207	0.9891
90	0.2282	0.9895	0.2184	0.9918	0.2242	0.9901	0.2238	0.9891
100	0.2245	0.9914	0.2177	0.9891	0.2233	0.9917	0.2179	0.9918

Table 3. Test Accuracy and test loss of binary class for Swin-T & Swin-B.

Type	Swin-T		Swin-B	
	Test Accuracy	Test loss	Test Accuracy	Test loss
No DR	97%		97%	
DR	98%		98%	
Average	98	0.0102	98	0.0079

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

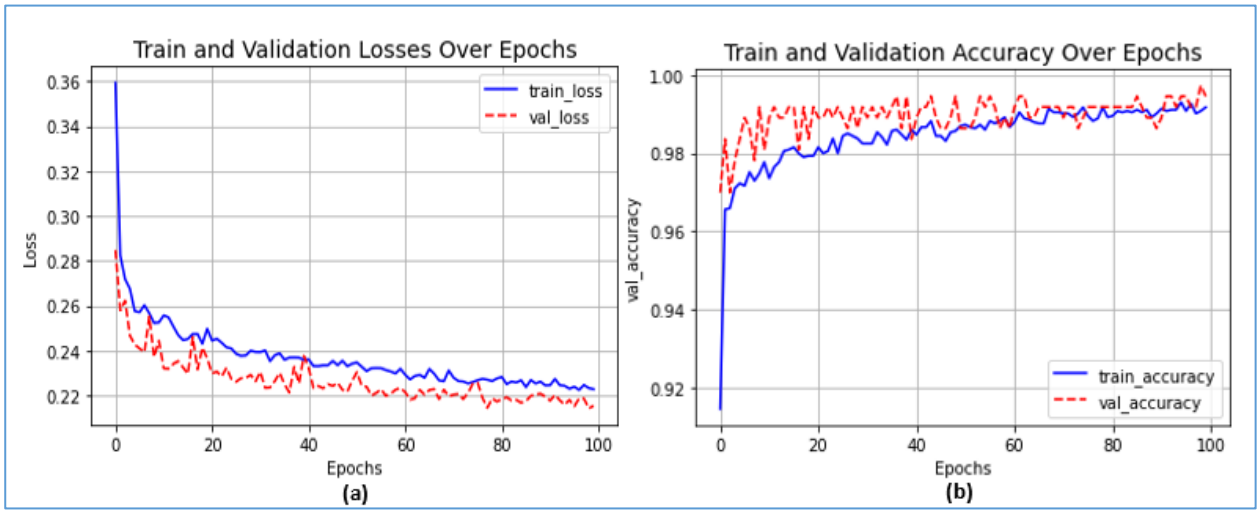

Figure 6. Training and validation over epochs for APTOS 2019 dataset, (a) loss, (b) accuracy, (epochs=100), WT Attention-Db5 Block-Swin-T to binary class.

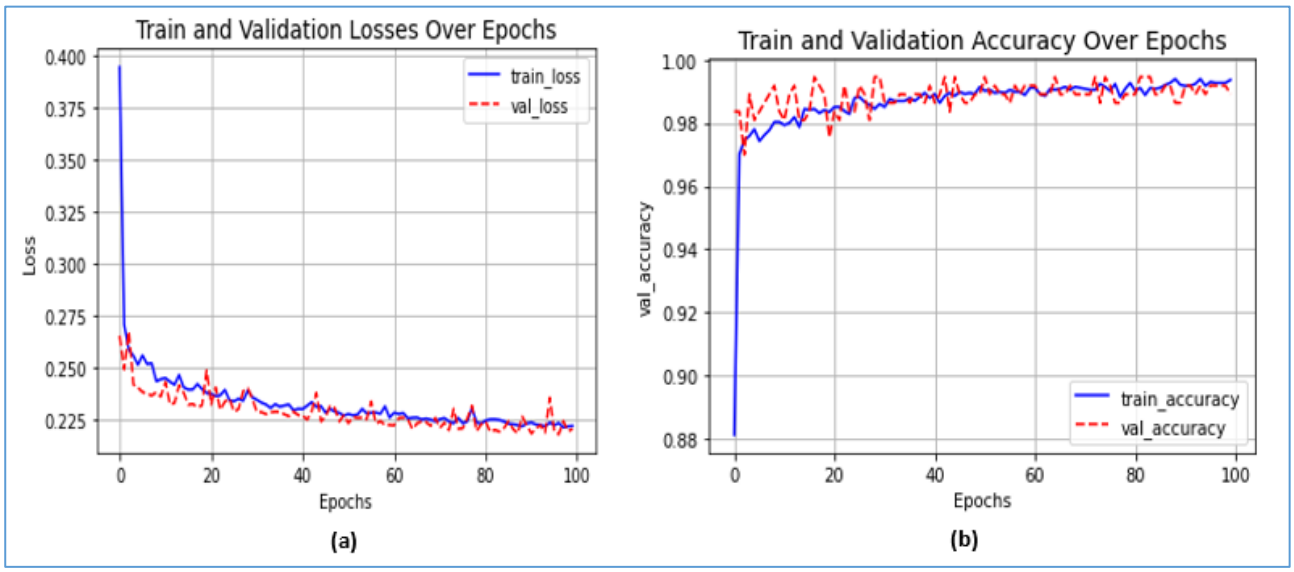

Figure 7. Training and validation over epochs for APTOS 2019 dataset, (a) loss, (b) accuracy, (epochs=100), WT Attention-Db5 Block-Swin-B to binary class.

Multiple Classifications

This is a description of a model designed for classifying diabetic retinopathy into multiple categories using Swin Transformers (Swin-T) and (Swin-B) with wavelet attention (WTA) for feature extraction. The training process consisted of training the model on the training dataset for 100 epochs with the Adam optimizer and a learning rate of 0.001. Throughout the training process, both Swin-T and Swin-B models were evaluated for their performance in binary classification of DR. The models were trained for a total of 46 minutes and 25 seconds (Swin-T) and 78 minutes and 59 seconds (Swin-B) over the course of 100 epochs. Swin-B demonstrated better validation accuracy with a score of 0.8743 at epoch 100, compared to Swin-T which scored 0.8552. The final training accuracy for Swin-B was 0.9208, while the final validation loss was 0.5759. For Swin-T, the final training accuracy was 0.9042, while the final validation loss was 0.6126. Fig. 8 presents a visual representation of the DR multiple classification training process using both Swin-T and Swin-B.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

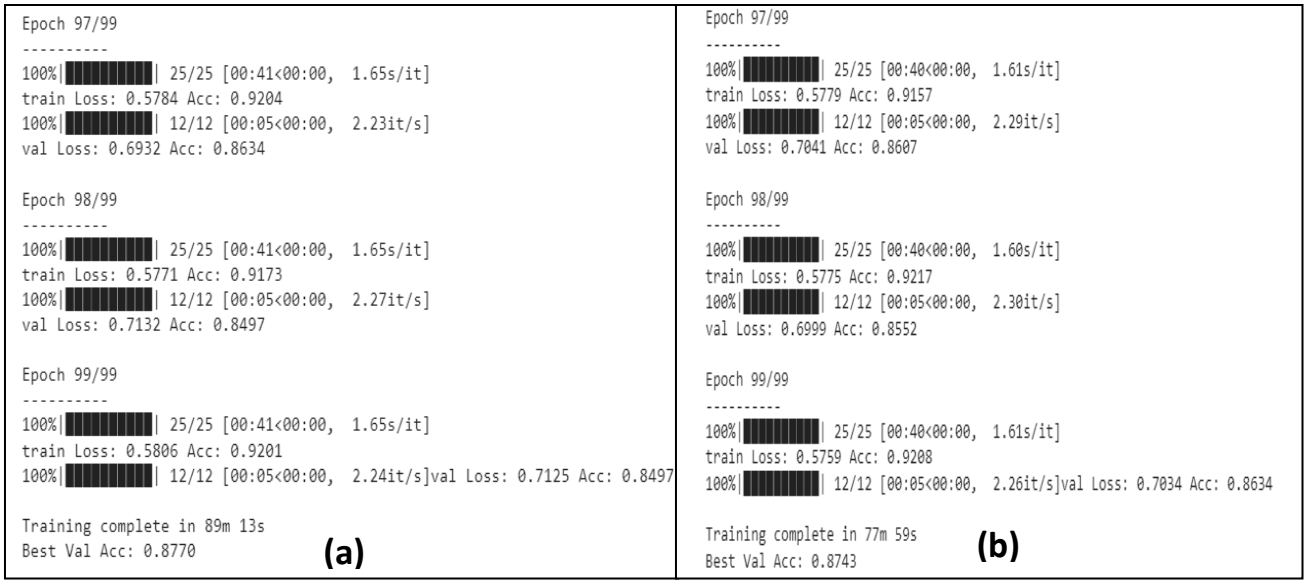

Figure 8. Part of the training process implementation for DR multiple classification using (a) Swin-T, and (b) Swin-B.

Table 4, Table 5, Fig. 9, and Fig. 10 summarize the performance of the WT Attention-Db5 Block with swin_tiny (Swin-T) and swin_base (Swin-B) for multiple classifications.

Regarding multiple classification of DR, the Swin-T model's performance varied depending on the severity of the condition, as shown in Table 5. While the model achieved high accuracy (98%) for identifying images without DR, its accuracy dropped significantly as the severity of DR increased, with an average accuracy of 84%. This suggests that the Swin-T model may have limitations in identifying more severe cases of DR.

On the other hand, the Swin-B model for multiple classification of DR performed reasonably well, with an average accuracy of 86% and a test loss of 0.0327. The model achieved high accuracy 97% for identifying images without DR, indicating its ability to accurately classify healthy retinal images.

Table 4. Classification of multiple class accuracy and loss for Swin-T & Swin-B.

	Swin-T				Swin-B			
Epochs	**Train-loss**	**Train-acc**	**Val-loss**	**Val-acc**	**Train-loss**	**Train-acc**	**Val-loss**	**Val-acc**
10	1.0179	0.6948	0.8840	0.7596	1.0818	0.6617	0.8917	0.7350
20	0.7364	0.8310	0.7614	0.8115	0.7201	0.8377	0.7590	0.8142
30	0.6904	0.8533	0.7175	0.8333	0.6963	0.8577	0.7448	0.8279
40	0.6686	0.8759	0.7131	0.8333	0.6659	0.8730	0.7185	0.8306
50	0.6336	0.8902	0.7082	0.8443	0.6502	0.8794	0.7195	0.8388
60	0.6226	0.8899	0.6884	0.8388	0.6363	0.8873	0.7248	0.8470
70	0.6009	0.9093	0.6984	0.8634	0.6135	0.8972	0.6975	0.8552
80	0.5905	0.9141	0.6752	00.8497	0.6008	0.9058	0.6959	0.8661
90	0.5731	0.9239	0.6792	0.8525	0.5987	0.9052	0.7112	0.8470
100	0.5759	0.9042	0.7398	0.8497	0.5759	0.9208	0.7034	0.8634

Table 5. Test Accuracy and Test Loss of Multiple Classes for Swin-T & Swin-B.

Type	**Swin-T**		**Swin-B**	
	Test Accuracy	**Test loss**	**Test Accuracy**	**Test loss**
No DR	98%		97%	
Mild DR	48%		60%	
Moderate DR	91%		93%	
Severe DR	25%		37%	
Proliferative DR	41%		48%	
Average	84%	0.0243	86%	0.0327

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

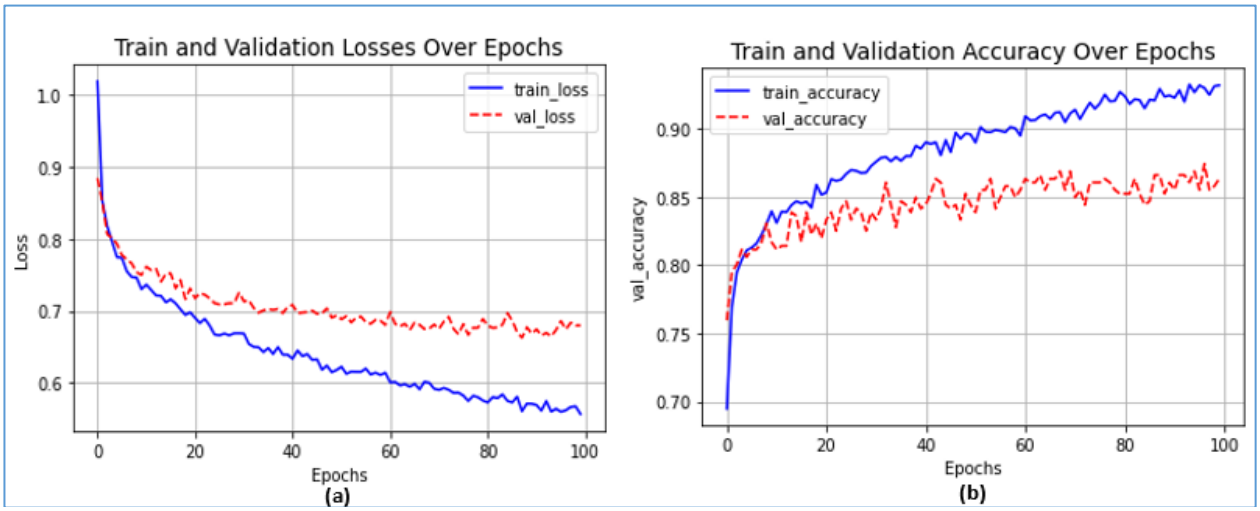

Figure 9. Training and validation over epochs for APTOS 2019 dataset, (a) loss, (b) accuracy, (epochs=100), WT Attention-Db5 Block-Swin-T to multi-class.

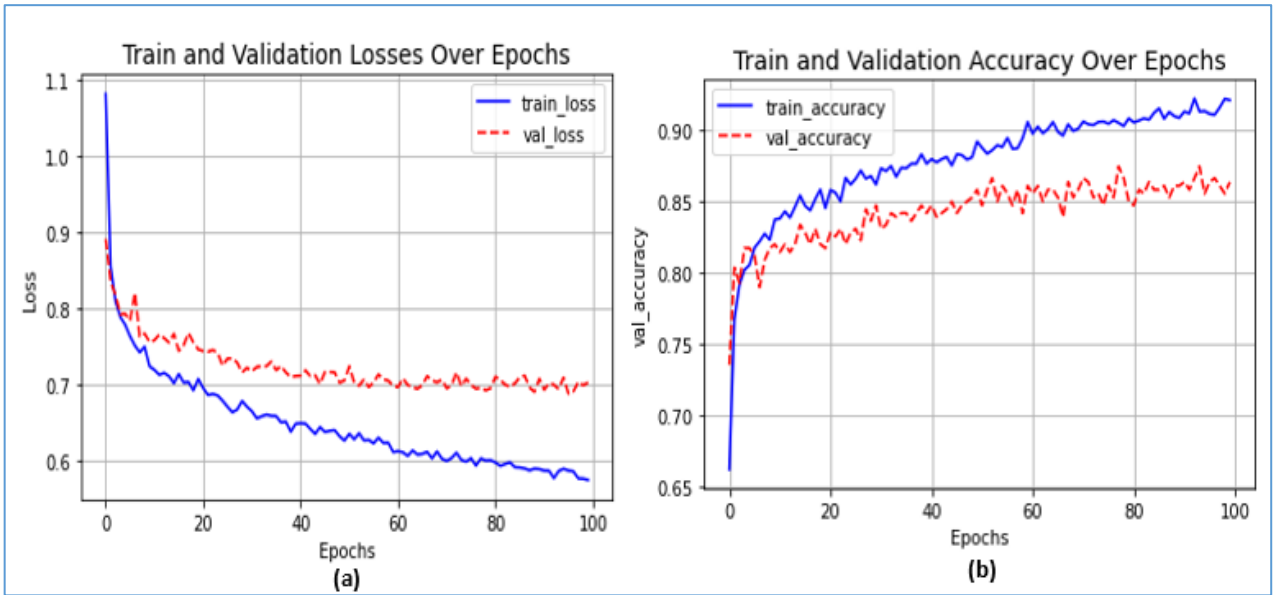

Figure 10. Training and validation over epochs for APTOS 2019 dataset, (a)loss, (b) accuracy, (epochs=100), WT Attention-Db5 Block-Swin-B to multi-class.

The proposed model Swin Transformer with Wavelet Attention (WTA) achieves superior accuracy and performance compared to conventional deep learning approaches by being more effective, quicker, and more precise. Using Swin Transformer allows accessing non-local information and, combined with other techniques such as wavelet attention, helps extract remote features, leading to increased diagnosis efficiency and better accuracy while reducing training and testing time. Table 6 presents a comparison between the proposed model and the existing state-of-the-art research in the field of image analysis and deep learning in medical imaging. This comparison provides insights into the effectiveness of the proposed model in terms of accuracy and performance, as compared to the existing approaches.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 6. Comparison of the proposed model with the state-of-the-art research.

Author (Ref.)	Method	Database	Efficiency	Research Advantages
Li H, et al.[10]	DnSwin	Real-world images	Proven to reduce noise, Speed compared to Euclidean-based protocols	Extracts high and low-frequency information
Xiangyu Zhao[11]	WA-CNN	CIFAR-10, CIFAR-100	Achieves good results in obtaining high classification accuracy	Extracts high and low-frequency features, obtains detailed information
Sabiha G. K. et al.[12]	Deep feature generator based on correction	APTOS 2019	Achieves good results with high accuracy of more than 90% for classification (Normal, NPDR, and PDR)	Uses rectangular patches to create deeply hidden patterns
Danny C. et al.[13]	PCAT-UNet	DRIVE, STARE, CHASE_DB1	Gives good results in segmenting the retinal blood vessels	Based on a transformer, uses a skip connection to combine features
Gupta, et al.[14]	ODCNN-RFIC	Retinal fundus images	Outperforms existing algorithms in terms of accuracy and performance	Uses pre-processing techniques, image segmentation, feature extraction, and a mayfly optimization with kernel extreme learning machine (MFO-KELM) classification model
Atwany, et al.[15]	Review and analysis of state-of-the-art deep learning methods	Various retinal fundus datasets	Discusses available datasets for detection, classification, and segmentation	Reviews and summarizes current research in the field, addresses research gaps and challenges
AlShemmary and Omran[16]	Combination of morphological operations and Hough Transform	Eye images	Potential Relation to diabetic retinopathy classification	Method for detecting pupils in eye images can identify abnormalities in diabetic retinopathy patients
Jaskari, et al.[17]	9 BNNs	Clinical dataset and benchmark datasets	Utilizes BNNs for uncertainty estimation in classifying diabetic retinopathy	Proposes a novel uncertainty measure, provides insights into using BNNs for clinical data
Zia, et al.[18]	Computerized learning model utilizing deep neural networks	Retinal images	Accurately detects key precursors of diabetic retinopathy	Combines strengths of VGG and Inception V3, uses entropy concept to select most discriminant features
Ashour[19]	Artificial neural networks (ANN)	Time series	RBF neural networks are less efficient and accurate in solving nonlinear time series	Highlights effectiveness of ANN in time-series applications
Proposed model	Swin Transformer with Wavelet Attention (WTA)	APTOS 2019	Compared to conventional deep learning approaches, it exhibits superior accuracy and performance by being more effective, quicker, and more precise.	Using Swin Transformer allows accessing non-local information and, combined with other techniques such as wavelet attention, helps extract remote features, leading to increased diagnosis efficiency and better accuracy while reducing training and testing time.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Conclusion

In this paper, a WT-Swin model based on the new WT Attention-Db5 Block was presented for early-stage detection of two and five-severity grades for Diabetic Retinopathy. The network was trained on APTOS 2019 dataset. The test accuracy of 98% and loss of 0.0102 for Swin-T, while the test accuracy of 98% and test loss of 0.0079 was reported with the binary-label classification for Swin-B when they apply Swin-T for multiple class the test accuracy of 84% and the loss of 0.0243, while the test accuracy of 86% and test loss of 0.0327 for multi-label classification when used Swin-B. The proposed method achieved better performance than computer-assisted diabetic retinopathy detection systems in terms of speed and accuracy, and it could be good for use in clinical applications to detect DR. The study on DR classification using the Swin Transformer with WTA has the potential to help researchers uncover critical areas related to the pathophysiology of diabetic retinopathy and the development of new diagnostic and treatment approaches. The study may help researchers reveal new insights into the complex and subtle features of the retina that are associated with diabetic retinopathy. The use of WTA for feature extraction in the Swin Transformer allows for the capture of multi-scale features, which may detect previously unrecognized patterns in the images that are associated with the disease. By analyzing large datasets of retinal images, deep learning models like the Swin Transformer with WTA can identify patterns and features that are associated with different stages and subtypes of the disease. This could ultimately lead to the development of personalized diagnostic and treatment approaches that are tailored to the individual patient.

As for future work, the proposed model can be utilized for detecting specific lesions in diabetic retinopathy, such as (AM) and (HE), it can be applied to other medical image analysis tasks, such as the detection and classification of lesions in other diseases or medical conditions.

Acknowledgment

The research behind this paper would not have been possible without the exceptional support of the Advisor 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.

Author's Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Kufa.

Author's Contribution Statement

R.A.D. Developed the Wavelet-Attention Swin (WAS) model for automatic diabetic retinopathy classification, integrating wavelet transform and attention mechanism into the Swin Transformer architecture. E.N.A collected a large dataset of retinal images, annotated them for diabetic retinopathy signs, and performed preprocessing. Ensured the availability of a reliable dataset for training and evaluation and contributed to manuscript writing and revision. W.A.M.A. Conducted experiments to evaluate the Wavelet-Attention Swin model, compared it with existing approaches, and performed statistical analyses.

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

References

1. Farooq MS, Arooj A, Alroobaea R, Baqasah AM, Jabarulla MY, Singh D, et al. Untangling computer-aided diagnostic system for screening diabetic retinopathy based on deep learning techniques. Sensors MDPI. 2022 24; 22(5): 1803. https://doi.org/10.3390/s22051803
2. Alyoubi WL, Abulkhair MF, Shalash WM. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. Sensors MDPI. 2021; 21(11): 3704. https://doi.org/10.3390/s21113704
3. Hameed EK, Al-Ameri LT, Hasan HS, Abdulqahar Z. The Cut-off Values of Triglycerides-Glucose Index for Metabolic Syndrome Associated with Type 2 Diabetes Mellitus. Baghdad Sci J. 2021; 19(2): 340-346. http://dx.doi.org/10.21123/bsj.2022.19.2.0340
4. Qureshi I, Ma J, Abbas Q. Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. Multimed Tools Appl. 2021; 80: 11691-11721. https://doi.org/10.1007/s11042-020-10238-4
5. Hasan DA, Zeebaree SR, Sadeeq MA, Shukur HM, Zebari RR, Alkhayyat AH. Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems-A Survey. 1st Babylon Int Conf Inf Technol Sci. 2021 28 (pp. 16-21). IEEE. https://doi.org/10.1109/BICITS51482.2021.9509920
6. Dutta S, Saini K. Securing data: A study on different transform domain techniques. WSEAS Trans Syst Control. 2021; 16: 110-120. https://doi.org/10.37394/23203.2021.16.8
7. Liu J, Ding J, Ge X, Wang J. Evaluation of total nitrogen in water via airborne hyperspectral data: potential of fractional order discretization algorithm and discrete wavelet transform analysis. Remote Sens. MDPI. 2021; 13(22): 4643. https://doi.org/10.3390/rs13224643
8. Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, et al. Swin transformer v2: Scaling up capacity and resolution. Proc IEEE/CVF Conf CVPR 2022: 12009-12019. https://doi.org/10.48550/arXiv.2111.09883
9. Xie Z, Lin Y, Yao Z, Zhang Z, Dai Q, Cao Y, et al. Self-supervised learning with swin transformers. arXiv preprint arXiv. 2021 10; 2105: 1-8. https://doi.org/10.48550/arXiv.2105.04553
10. Hao Li, Zhijing Yang, Xiaobin Hong, Ziying Zhao, J unyang Chen, Yukai Shi, et al. DnSwin: Toward real-world denoising via a continuous Wavelet Sliding Transformer. Knowl Based Syst . 2022, 14; 255: 109815. https://doi.org/10.1016/j.knosys.2022.109815
11. Zhao X, Huang P, Shu X. Wavelet-Attention CNN for image classification. Multimedia Systems. 2022 ; 28(3): 915-24. https://doi.org/10.1007/s00530-022-00889-8
12. Kobat SG, Baygin N, Yusufoglu E, Baygin M, Barua PD, Dogan S, et al. Automated Diabetic Retinopathy Detection Using Horizontal and Vertical Patch Division-Based Pre-Trained DenseNET with Digital Fundus Images. Diagnostics. 2022 15; 12(8): 1975. https://doi.org/10.3390/diagnostics12081975
13. Chen D, Yang W, Wang L, Tan S, Lin J, Bu W. PCAT-UNet: UNet-like network fused convolution and transformer for retinal vessel segmentation. Raja G, editor. PLoS One. 2022 , 24; 17(1): e0262689. https://dx.plos.org/10.1371/journal.pone.0262689
14. Gupta IK, Choubey A, Choubey S. Mayfly optimization with deep learning enabled retinal fundus image classification model. Comput Electr Eng. 2022 , 1; 102: 108176. https://doi.org/10.1016/j.compeleceng.2022.108176
15. Atwany MZ, Sahyoun AH, Yaqub M. Deep learning techniques for diabetic retinopathy classification: A survey. IEEE Access. 2022 8: 28642 - 28655. https://doi.org/10.1109/ACCESS.2022.3157632
16. Omran M, AlShemmary EN. Towards accurate pupil detection based on morphology and Hough transform. Baghdad Sci J. 2020 ,1; 17(2): 583-590. http://dx.doi.org/10.21123/bsj.2020.17.2.0583
17. Jaskari J, Sahlsten J, Damoulas T, Knoblauch J, Särkkä S, Kärkkäinen L, et al. Uncertainty-aware deep learning methods for robust diabetic retinopathy classification. IEEE Access. 2022; 10: 76669-76681. https://doi.org/10.1109/ACCESS.2022.3192024
18. Zia F, Irum I, Qadri NN, Nam Y, Khurshid K, Ali M, et al. A multilevel deep feature selection framework for diabetic retinopathy image classification. CMC 2022; 70(2): 2261-2276. https://doi.org/10.32604/cmc.2022.017820
19. Ashour M A H. Optimized Artificial Neural network models to time series. Baghdad Sci J. 2022 ;19(4): 0899-0904. https://doi.org/10.21123/bsj.2022.19.4.0899
20. Liu Y, Guan L, Hou C, Han H, Liu Z, Sun Y, et al. Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform. Appl Sci. 2019 , 15; 9(6): 1108. https://doi.org/10.3390/app9061108
21. Nobre J, Neves RF. Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. Expert Systems with Applications. 2019, 1; 125: 181-94. https://doi.org/10.1016/j.eswa.2019.01.083

Published Online First: January, 2024
https://doi.org/10.21123/bsj.2024.8565
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

22. Freire PK de MM, Santos CAG, Silva GBL da. Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting. Appl Soft Comput ASC. 2019; 80: 494–505. https://doi.org/10.1016/j.asoc.2019.04.024
23. Tymchenko B, Marchenko P, Spodarets D. Deep learning approach to diabetic retinopathy detection. arXiv preprint arXiv: 2020 , 3; 2003: 02261.
24. Bodapati JD, Naralasetti V, Shareef SN, Hakak S, Bilal M, Maddikunta PKR, et al. Blended Multi-Modal Deep ConvNet Features for Diabetic Retinopathy Severity Prediction. Electronics. 2020 , 30; 9(6): 914. https://doi.org/10.3390/electronics9060914
25. Khalaf M, Dhannoon BN. MSRD-Unet: Multiscale Residual Dilated U-Net for Medical Image Segmentation. Baghdad Sci J. 2022 , 5; 19(6(Suppl.)): 1603-1611. https://doi.org/10.21123/bsj.2022.7559

الانتباه الموجي لتحويل النوافذ المتنقل Swin لتصنيف اعتلال الشبكية السكري التلقائي

رشا علي دهن [1]، ابتسام نجم الشمري [2] ، وليد محمود الجوهر [3]

[1]قسم علوم الحاسوب، كلية علوم الحاسوب والرياضيات، جامعة الكوفة، الكوفة، العراق.
[2]مركز البحث والتأهيل المعلوماتي، جامعة الكوفة، الكوفة، العراق.
[3]جامعة اوروك، بغداد، العراق.

الخلاصة

اعتلال الشبكية السكري (DR) هو أحد مضاعفات مرض السكري الذي يؤثر على العين عن طريق إتلاف الأوعية الدموية في شبكية العين. يمكن أن يؤدي ارتفاع مستويات السكر في الدم إلى تسرب أو انسداد هذه الأوعية ، مما يؤدي إلى فقدان البصر أو العمى. يعد الاكتشاف المبكر لـ DR أمرًا ضروريًا لمنع العمى ، ولكن التحليل اليدوي لصور قاع العين يمكن أن يستغرق وقتًا طويلاً ، خاصة مع عدد كبير من الصور. اكتسبت Swin-Transformers شعبية في تحليل الصور الطبية ، مما أدى إلى تقليل الحسابات وتحقيق نتائج أفضل. تقدم هذه الورقة WT Attention-Db5 Block ، والتي تركز الانتباه على مجال التردد العالي باستخدام تحويل المويجات المنفصل .(DWT) تستخرج هذه الكتلة معلومات مفصلة من مجال التردد العالي مع الاحتفاظ بالمعلومات الأساسية منخفضة التردد. تناقش الدراسة نتائج تحدي كشف العمى لعام 2019 (APTOS 2019 BD) الذي عقدته جمعية آسيا والمحيط الهادئ لطب العيون عن بُعد. يحقق نموذج WT-Swin المقترح تحسينات كبيرة في دقة التصنيف. بالنسبة إلى Swin-T ، تبلغ دقة التدريب والتحقق من الصحة 99.14٪ و 98.91٪ على التوالي. بالنسبة للتصنيف الثنائي باستخدام Swin-B ، تبلغ دقة التدريب 99.01٪ ، ودقة التحقق 99.18٪ ، ودقة الاختبار 98٪. في التصنيف المتعدد ، تبلغ دقة التدريب والتحقق 93.19٪ و 86.34٪ على التوالي ، بينما تبلغ دقة الاختبار 86٪. في الختام ، يعد الاكتشاف المبكر لـ DR ضروريًا لمنع فقدان البصر. تُظهر كتلة WT Attention-Db5 المدمجة في نموذج WT-Swin نتائج واعدة في دقة التصنيف.

الكلمات المفتاحية: قاعدة بياناتAPTOS ، اعتلال الشبكية السكري، الانتباه الموجي، محولSwin-B، محول Swin-T .