Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Automated Binary Classification of Diabetic Retinopathy by SWIN Transformer

Rasha Ali Dihin[a], *Ebtesam N. AlShemmary*[b]*, *Waleed A. Mahmoud Al-Jawher*[c]

[a]*Department of Computer Science, University of Kufa, Iraq. Email: rashaa.aljabry@uokufa.edu.iq*

[b]*IT Research and Development Center, University of Kufa, Iraq. Email: dr.alshemmary@uokufa.edu.iq*

[c]*Uruk University, Iraq. Email: profwaleed54@gmail.com*

ARTICLEINFO

Article history:

Received: 25/01/2023
Revised form: 11/03/2023
Accepted: 16/03/2023
Available online: 31/03/2023

Keywords:

Diabetic retinopathy

Swin transformer

Gaussian

APTOS

CLAHE

ABSTRACT

Diabetic retinopathy is a medical condition that affects the eyes and is caused by damage to the blood vessels in the retina (the light-sensitive part of the eye) due to high blood sugar levels in individuals with diabetes. This damage can lead to vision loss or even blindness. It is a common complication of diabetes and a leading cause of blindness in working-age adults. In this paper, to automatically classify images of the retina as having either diabetic retinopathy or not. The goal of this classification is to assist medical professionals in diagnosing diabetic retinopathy more accurately and efficiently, potentially improving patient outcomes. In this process, the Swin transformer model is trained on the APTOS dataset of retinal images and then used to automatically classify new images as either positive or negative for diabetic retinopathy. Used CLAHE and Gaussian, to improve the input image, and the model achieved a Test Accuracy of 96%, Sensitivity of 96%, F1 Score of 96% for Swin-T and Test Accuracy of 98% for Swin-B, Sensitivity of 98%, and F1 Score of 98%.

https://doi.org/*10.29304/jqcm.2023.15.1.1166*

1. Main text

The eye is a vital part of the human body, serving as the primary source of external information. According to the International Diabetes Federation (IFD), it is predicted that by 2040, there will be 642 million individuals with diabetes worldwide. It is crucial to accurately assess the severity of diabetic retinopathy to recommend the appropriate treatment. However, manual diagnosis can be time-consuming and dependent on the physician's expertise. Therefore, automating this process is desirable. With the development of artificial intelligence, better results have been achieved in image classification and identifying crucial image features, allowing for automatic screening of the retina [1].

The classification of DR is an important aspect in determining the stages of severity and prognosis in its development. Over time, several classifications have been established, with the Airlie House classification being the most widely

∗Corresponding author *Ebtesam N. AlShemmary*

Email addresses: *dr.alshemmary@uokufa.edu.iq*

Communicated by ‘sub etitor’

used as the basis for current classifications. This classification separates DR into two groups: non-proliferating and proliferating [2].

Swin transformer is a type of neural network architecture in the field of machine learning. It is a variant of the Transformer architecture and has become popular in natural language processing tasks [3]. The extends the Transformer architecture to handle computer vision tasks, specifically image classification tasks. Swin Transformer incorporates the ideas of local and global attention mechanisms, allowing it to handle both fine-grained and global contextual information in images. It has been used in various computer vision tasks such as object recognition and semantic segmentation and has shown promising results [4].

This paper's sections were arranged as follows: in Section 2, Related Works, in Section 3, we elaborate on the proposed framework, including CLAHE and Gaussian with Swin transformer, we present and discuss the experimental results in Section 4 and finally summarize the conclusions in Section 5.

2. Related work

This section reviews the relevant literature.

S. Sanjana et al.[5] proposed a binary classification of DR using five Transfer Learning models Xception, InceptionResNetV2, MobileNetV2, DenseNet121, and NASNet Mobile. The highest validation accuracy was achieved by InceptionResNetV2 with 96.25%. In the pre-processing phase, the images were transformed using techniques such as rescaling, shearing, zooming, and flipping horizontally.

Y. Li. et al. [6] proposed a Semi-supervised Auto-encoder Graph Network (SAGN) for challenging DR diagnosis. The SAGN model has three components: auto-encoder feature learning, neighbor correlation mining, and graph representation. The performance of SAGN on the APTOS 2019 dataset was evaluated with an accuracy of 94.4%. In the pre-processing phase, to reduce the impact of irrelevant regions in the fundus images, the black regions were removed through cropping and the images were resized to 512x512 pixels before being fed into the network. Additionally, each image was augmented by random horizontal and vertical rotation.

S. Gungor K. et al. [7] proposed a new method that leverages a deep feature generator based on correction and is inspired by the Vision Transformer (ViT). The ViT uses an MLP mixer, which extracts features using a fixed-size square patch. In this method, rectangular patches were utilized instead of square patches to create deeply hidden patterns, and a DenseNet201 was employed. This approach achieved good results, with a classification accuracy of over 90% for the categories 'Normal', 'NPDR', and 'PDR' on the APTOS 2019 dataset".

D. Chen. et al. [8] introduced "A new unit called the Transformer UNet (PCAT-UNet) was proposed in the paper. This unit combines attention mechanisms and the shape of a letter U and is based on a transformer architecture. To combine features, skip connections are utilized on both sides. The results indicated that the proposed method produces excellent results in segmenting retinal blood vessels in both DRIVE, STARE, and CHASE_DB1 datasets."

H. Siyuan et al. presented a two-stream network named TSTNet for the classification of remote sensing images. The Swin Transformer was employed as the foundation of each stream, producing impressive results. The experiments on three demanding public datasets indicated that TSTNet has a better classification performance compared to other leading models [9].

3. Method

In this section, we first introduce the global context-modeling framework and then discuss in detail the design shown in Figure 1.

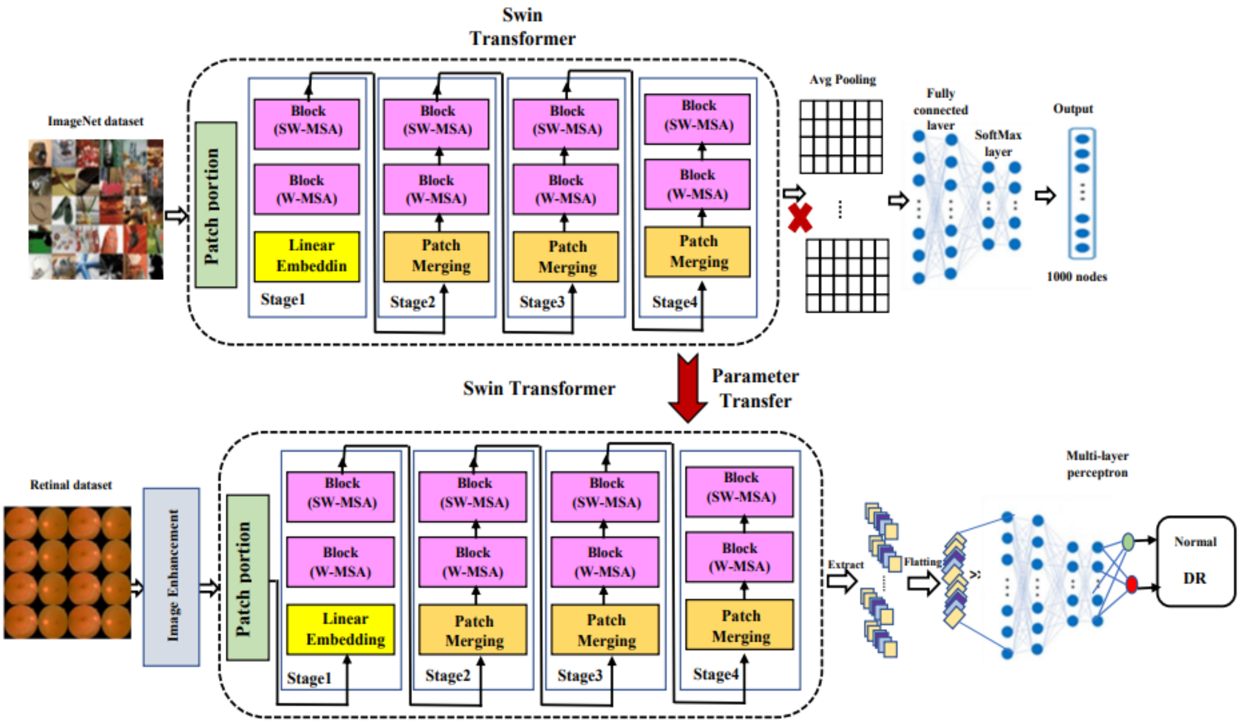


Fig. 1 – The proposed framework.

3.1. Image enhancement:

The benefits of image enhancement include:

Improved visual quality: Enhanced images have a higher visual quality and appear clearer, sharper, and more detailed. Better feature extraction: Image enhancement can improve the visibility of features, making them easier to detect and extract, Increased accuracy of image analysis: Enhanced images are more suitable for computer vision and image analysis tasks, leading to increased accuracy and improved results and Enhanced interpretation: Image enhancement is a valuable technique that aids humans in interpreting and comprehending images more easily. One way to achieve this is by utilizing a Gaussian filter and CLAHE to remove noise and improve contrast, as illustrated in Figure 2. The goal of noise removal is to eliminate any image disturbance caused by noise. Gaussian filter is the simplest low-pass filter that effectively removes high-frequency noises. Contrast Limited Adaptive Histogram Equalization (CLAHE) has a controllable parameter to limit the contrast and the technique is successfully enhancing the low-contrast images[10]

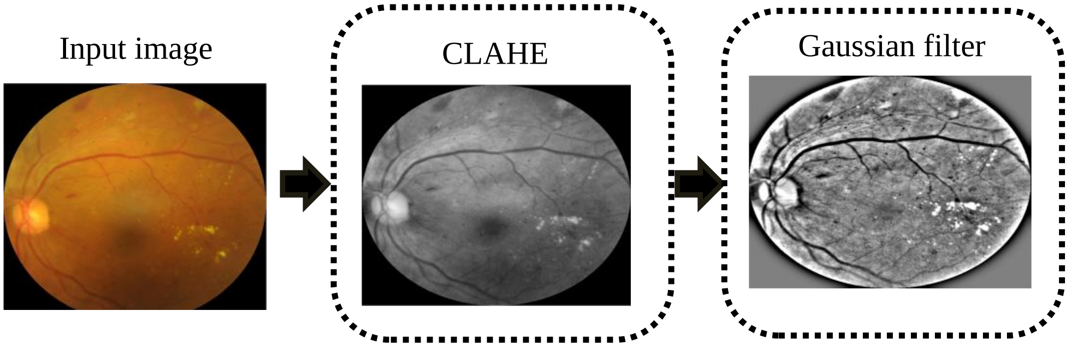


Fig. 2 - CLAHE and Gaussian method

When using CLAHE on an image, the contrast limiting threshold is set to 2. When applying a Gaussian filter, determine the values of alpha and beta as 4 and -4 respectively. Add 128 to the image after adjusting the gamma value, and if a value exceeds the maximum expected value, set it to 0. If a value is greater than 1, set it to 1.

Algorithm CLAHE and Gaussian Filter

```
Input: A Retinal image
Output: enhancement image with CLAHE and Gaussian
Step1: Read the image and store it in a variable (e.g., img)
Step2: Convert the image to grayscale and save it in a new variable (e.g., im1)
Step3: Apply the CLAHE function with a clip limit of 2 and a matrix size of 8x8 onim1
Step4: Convert the result of CLAHE back to RGB and save it in a new variable (e.g. im2)
Step5: Determine the values of $\alpha$ and $\beta$ for the weighted addition (e.g., $\alpha = 4$ and $\beta = -4$)
Step6: Determine the value of $\gamma$ to add to the image (e.g., $\gamma = 128$)
Step7: Apply a Gaussian filter with a kernel size of (0,0) and a standard deviation of 1 on im2
Step8: Add the result of the Gaussian filter to the original image using the weighted addition (cv2.addWeighted(im2,4, cv2.GaussianBlur(img, (0,0),1), -4,128)) and save the result in a new variable (e.g., im3)
Step9: Display the final image (im3)
```

4. Result:

The proposed architecture was developed using a software package (Python). The implementation was central processing unit (CPU) specific. All experiments were performed on Colab with GPU 15G.

4.1. Datasets

The APTOS 2019 Kaggle benchmark dataset is a collection of fundus images of the retina taken under various conditions. These images have been manually categorized by specialists into 5 classes, ranging from "0" meaning no diabetic retinopathy to "4" indicating severe proliferative diabetic retinopathy. Figure 3 displays the retinal images in the dataset that corresponds to each level of severity. This dataset is used in the challenge of detecting blindness and provides valuable information for the research of diabetic retinopathy [11].

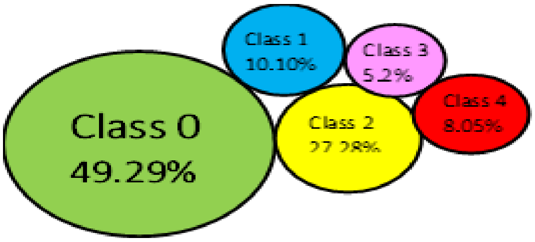


Fig. 3 - Datasets class percentage distribution

4.2. Evaluation Metrics

The evaluation metrics for model performance in DR detection included two-class, The performance is measured with various performance measures including accuracy, Specificity, Sensitivity, and F1-Score, [12], [13].

Where

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

$$SE = \frac{TP}{TP + FN} \qquad (2)$$

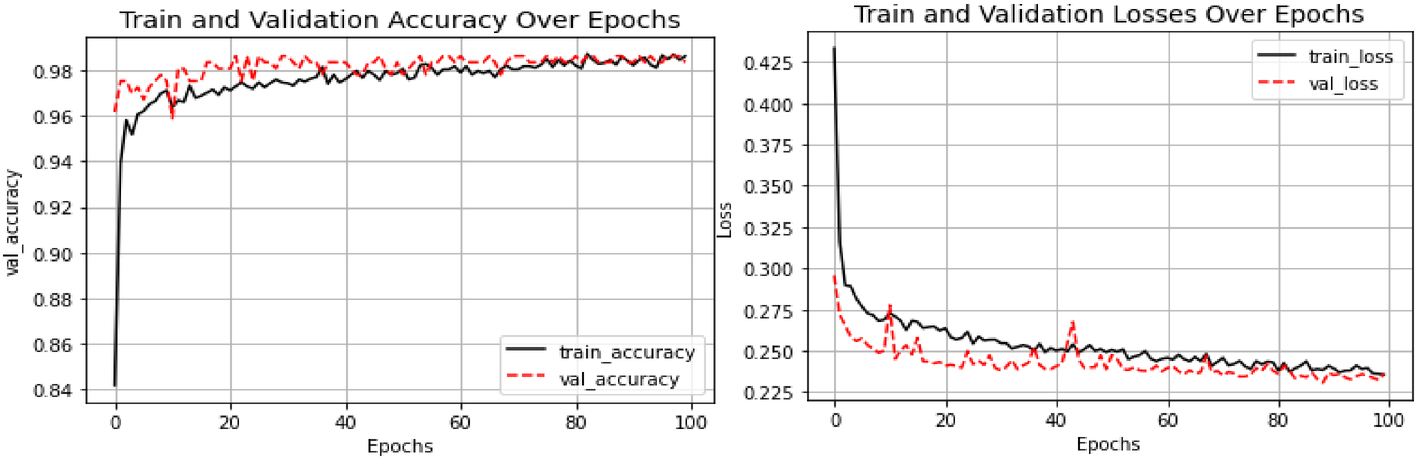
$$SP = \frac{TN}{FP + TN} \qquad (3)$$

$$F1Score = 2.\frac{precision * \mathrm{Re}\,call}{(precision + \mathrm{Re}\,call)} \qquad (4)$$

4-3: Research the different Swin transformer

4-3-1: Swin-T:

We used CLAHE and Gaussian in the enhancement phase on the image with size (224*224) with Swin-T transformer to DR classification to binary class as shown in Table 1, Figure 4. Table 1 shows the test accuracy, Sensitivity, specificity, and F1 Score, Figure 4 shows the raining and validation accuracy and loss between the training and valuation for 100 epoch, and Figure 5 displays the confusion matrix.



(a) **(b)**

Fig. 4 - Training and validation over epoch for APTOS 2019 dataset, (a)loss, (b) accuracy, (epochs=100), CLAHE, and Gaussian -Swin-T to binary class.

Table 1 -Test Accuracy and test loss of binary class for Swin-T

Class	Test Accuracy	Test loss	Sensitivity	Specificity	F1 Score
No-DR	96%	0.0077	0.9700	0.9700	0.9761
DR	96%		0.9698	0.9698	0.9766
Average	96%		0.9416	0.9416	96%

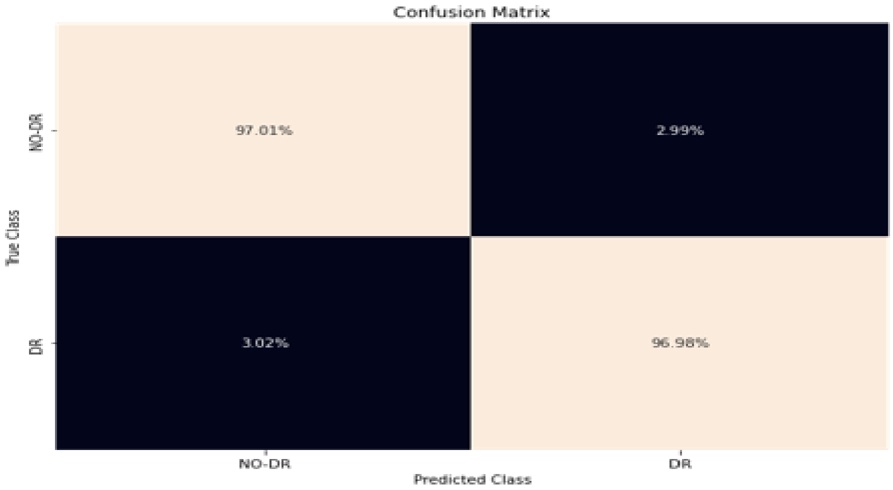


Fig. 5 - Confusion matrix for Swin-T binary class.

4-3-2: Swin-B:

In the image enhancement phase for DR classification, we applied a combination of Contrast Limited Adaptive Histogram Equalization (CLAHE) and Gaussian filter on images with a size of (224 x 224) using the Swin-T transformer. The results are presented in Table 2 and Figure 6. Table 2 displays the test accuracy, sensitivity, specificity, and F1 Score, while Figure 6 illustrates the comparison of the training and validation accuracy and loss after 100 epochs. Additionally, Figure 7 presents the confusion matrix.

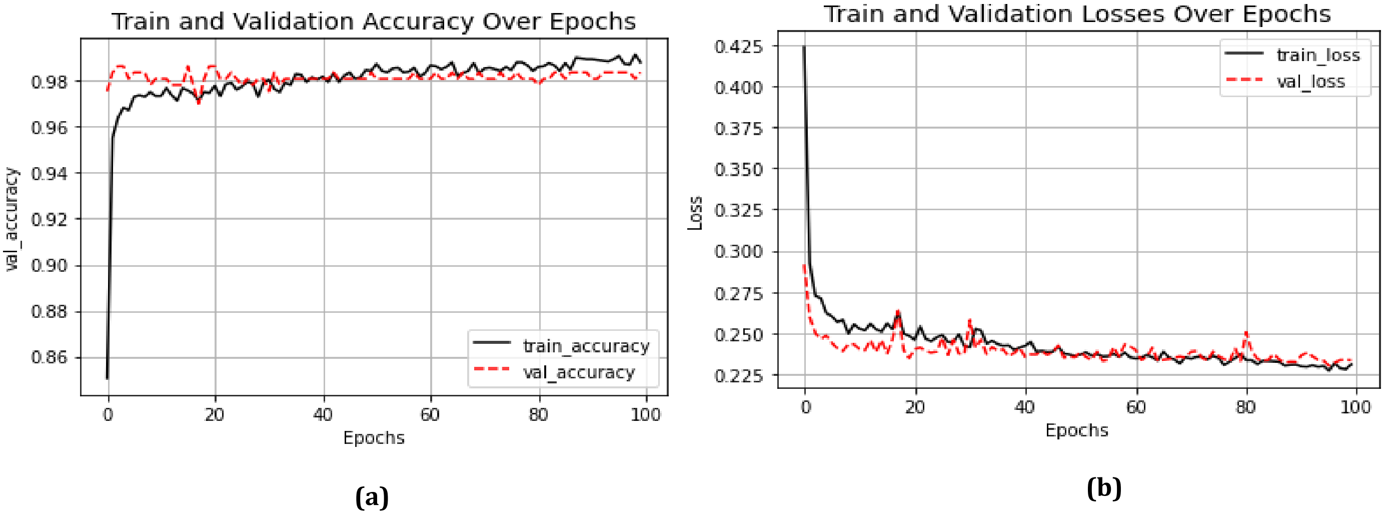


Fig. 6 - Training and validation over epoch for APTOS 2019 dataset, (a)loss, (b) accuracy, (epochs=100), CLAHE, and Gaussian -Swin-B to binary class.

Table 2- Test Accuracy and test loss of binary class for Swin-B

Class	Test Accuracy	Test loss	Sensitivity	specificity	F1 Score
No-DR	97%	0.0077	0.9753	0.9753	0.9814
DR	97%		0.9789	0.9789	0.9841
Average	97%		0.9773	0.9773	0.9885

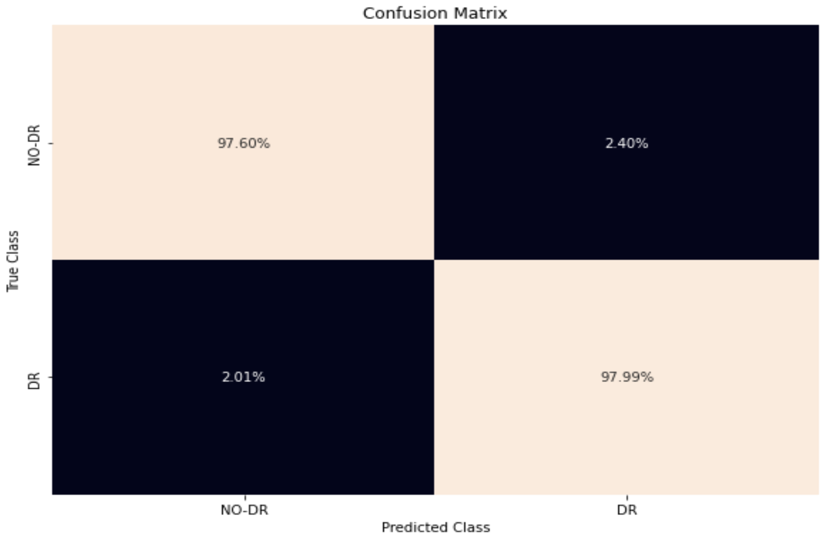


Fig. 7 - Confusion matrix for Swin-T binary class

5. Conclusion:

In this paper, we used the Contrast Limited Adaptive Histogram Equalization (CLAHE) and Gaussian filter in the enhancement phase and apply Swin-T and Swin-B DR classification where change on the last layers in the Swin transformer, where the accuracy for training is 0.9863 and 0.9836 for validation and the best accuracy is 0.9863 and the test accuracy is 96%, Sensitivity of 96% and F1 Score of 96% for Swin-T while for Swin-B is the training accuracy is 0.9881, while the validation accuracy is 0.9836 and the best accuracy is 0.9863, while the test accuracy is 97%, Sensitivity of 98% and F1 Score of 96% for binary DR classification.

Acknowledgments

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.

References

[1] F. Eigelshoven et al., “Advances on Smart and Soft Computing,” Inf. Process. \& Manag., vol. 12, no. 2, pp. 3–18, 2020.
[2] T. Nazir, M. Nawaz, J. Rashid, and R. Mahum, “Detection of Diabetic Eye Disease from Retinal Images Using a Deep Learning Based CenterNet Model Tahira,” MDPI, vol. 21, no. 5283, 2021.
[3] Z. Liu, Y. Lin, Y. Cao, H. Hu, and Y. Wei, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows arXiv:2103.14030v2,” arXiv:2103.14030v2, 2021.
[4] J. Liang, J. Cao, G. Sun, and K. Zhang, “SwinIR: Image Restoration Using Swin Transformer,” arXiv:2108.10257v1, 2021.
[5] S. Sanjana, N. S. Shadin, and M. Farzana, “Automated Diabetic Retinopathy Detection Using Transfer Learning Models,” 2021.
[6] Y. Li, Z. Song, S. Kang, S. Jung, and W. Kang, “Semi-Supervised Auto-Encoder Graph Network for Diabetic Retinopathy Grading,” IEEE Access, vol. 9, pp. 140759–140767, 2021, doi: 10.1109/ACCESS.2021.3119434.

[7] S. Gungor Koba, N. Baygin, E. Yusufoglu, and M. Baygin, “Automated Diabetic Retinopathy Detection Using Horizontal and Vertical Patch Division-Based Pre-Trained DenseNET with Digital Fundus Images,” MDPI, 2022.

[8] D. Chen, W. Yang, L. Wang, S. Tan, J. Lin, and W. Bu, “PCAT-UNet: UNet-like network fused convolution and transformer for retinal vessel segmentation,” PLoS ONE, vol. 17, no. 1 1. 2022, doi: 10.1371/journal.pone.0262689.

[9] S. Hao, B. Wu, K. Zhao, and Y. Ye, “Two-Stream Swin Transformer with Differentiable Sobel Operator for Remote Sensing Image Classification,” Remote Sens., 2022.

[10] R. Fan, X. Li, S. Lee, and T. Li, “Smart Image Enhancement Using CLAHE Based on an F-Shift Transformation during Decompression,” MDPI, vol. 12, no. 1374, 2020.

[11] B. Tymchenko, P. Marchenko, and D. Spodarets, “Deep Learning Approach to Diabetic Retinopathy Detection Borys,” arXiv:2003.02261v1, 2020.

[12] R. S. R, “Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network,” J. Trends Comput. Sci. Smart Technol., vol. 3, no. 2, 2021.

[13] C.-Y. Tsai, C.-T. Chen, G.-A. Chen, and C.-T. K. Kuo, “Necessity of Local Modification for Deep Learning Algorithms to Predict Diabetic Retinopathy Ching-Yao,” Int. J. Environ. Res. Public Health, 2022.