Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Automated Early Detection of Retinopathy of Prematurity Zones using SWIN Transformer

Nazar Salih Abdulhussein[1,*], Royida A. Ibrahem Alhayali[2], Mohammed Rashid Subhi[3], Nebras Hussein[4], Mohamed Ksantini[5], Amina Turki[5]

[1] Computer Science Department, Al-Imam Al-Adham University College, Baghdad, Iraq
[2] Department of Computer Engineering, College of Engineering, University of Diyala, Diyala, Iraq
[3] Department of Petroleum System Control Engineering, College of Petroleum Processes Engineering, Tikrit University, Tikrit, Iraq
[4] Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq
[5] Control and Energies Management Laboratory (CEM-Lab), National Engineering School of Sfax, University of Sfax, Sfax, Tunisia

ARTICLE INFO

Article history:
Received 7 January 2025
Received in revised form 24 February 2025
Accepted 15 July 2025
Available online 8 August 2025

Keywords:
Health care; deep learning; retinopathy of prematurity; fundus images; SWIN transformer

ABSTRACT

Retinopathy of prematurity (ROP) is known to be the primary cause leading to permanent vision loss in children, which calls for its diagnosis and treatment based on subjective assessment of retinal vascular characteristics; even though this traditional approach is practical, it takes much time and likely results in errors. Therefore, automation is required not only to enhance precision but also productivity. The study proposes an innovative approach to early detection of ROP zones on fundus images between 2015 and 2020. It will use the SWIN Transformer model, which has demonstrated superior precision and achieved a performance rate of 90.11%. This work denotes significant advancement in this field, emphasizing the potential of transformer-based architectures for the precise and efficient detection of ROP in clinical environments. The findings underscore the significance of utilizing state-of-the-art, comprehensive learning approaches to enhance early detection procedures, improving clinical outcomes for at-risk newborns.

1. Introduction

In 1940, Terry blazed the trail as the primary investigator to pin Retinopathy of Prematurity (ROP) down and describe it. He termed it retrolental fibroplasia due to detachment of the retina located behind the lens [1]. Later, it was widely acknowledged that this is indeed the major factor contributing to childhood blindness globally [2,3]. The survival rate for neonates delivered at a gestational age less than 37 weeks has increased with establishment of neonatal intensive care units: up to 15 million preterm births take place every year in all parts of the world [4]. Up to 15 million preterm births occur globally each year [5]. Today ROP is now a major public health issue [6]. There are two main issues leading to blindness caused by ROP: a scarcity of ophthalmologists knowledgeable enough to detect and treat the disease. Early treatment of ROP in high-risk individuals

[*] *Corresponding author*
E-mail address: nazarsalih@imamaladham.edu.iq

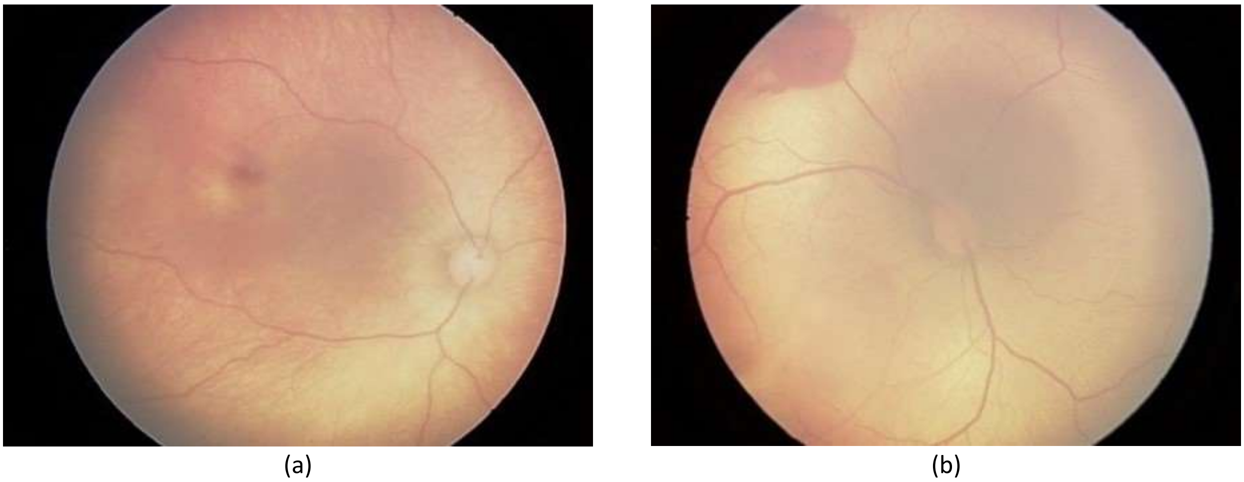
https://doi.org/10.37934/ard.141.1.217231

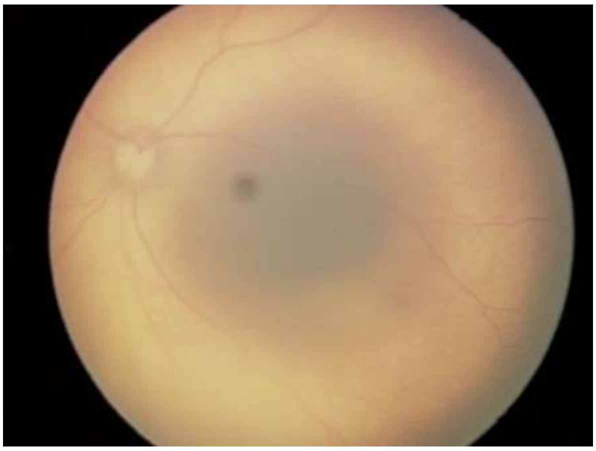
can help them preserve the majority, if not all, of their vision. Therefore, early-stage screening for ROP is crucial to prevent long-lasting visual impairment [7].

ROP is classified into stages 1-5 based on the severity of the sickness [8], zones 1-3 [7] and the presence of plus disease [9], according to the principles of the International Classification of Retinopathy of Prematurity (ICROP) established in 1984 [10], 1987 [11] and 2005 [12]. The first zone encompasses the entire visual field and has a radius twice the distance between the optic disc's centre and the macula's fovea. Zone 2 is an annular space, different from zone 1, with a radius that matches the distance between the optical disc and the serrated nasal border. Outside zones 1 and 2, the remaining crescent-shaped territories comprise zone 3. ICROP criteria for ROP severity are shown in Figure 1.

Retinal imaging is the gold standard for ROP diagnosis. Many ROP fundus examinations utilize the Retinal Camera (Retcam), a wide-angle optical retinal imaging equipment. It can take, store, produce and send fundus images in both directions. Furthermore, its structure is superior for educational purposes, clinical research, consulting and follow-up. There is now a plethora of morphological datasets available [13].

Artificial intelligence (AI) in health care has increased alongside advancements in AI technology. Applying deep learning models in medical diagnosis has proven beneficial [14]. These models have been employed in numerous computer vision applications, including image segmentation, object recognition [15], picture categorization and sickness detection. Due to the availability of big data sets and the development of deep network designs, AI has been proposed to help medical professionals deal with some of the strain. Image recognition and classification have been the target of several traditional machine-learning approaches. To finish the task, however, other methods, such as feature extraction and dimensionality reduction, are required, which prolong the process. However, losing critical features during the image matrix's conversion to a one-dimensional vector could reduce the models' efficacy.





(c)

Fig. 1. Retina pictures depicting the three zones of retinopathy of prematurity (ROP): (a) zone 1, (b) zone 2 (c) zone 3 [7]

Using these extensive databases, various techniques, such as the ophthalmoscope, can perform diagnostic analysis on ROP. However, via the utilization of AI, we can get optimal outcomes. This research aims to develop a unique way of identifying and classifying ROP zones very early using the SWIN Transformer [16], which could be applied to an already available fundus image dataset.

The proposed methodology introduces the SWIN Transformer, adopting a hierarchical vision transformer architecture and a shifted window mechanism to use pictures as an input data resource, allowing for more effective and reliable picture processing. The primary goal of our present research is to develop a new approach for the early identification and classification of ROP zones. The method will be tested on a selected collection of fundus images using a SWIN Transformer.

SWIN Transformer uses the hierarchical vision transformer architecture with the shifted window mechanism. This allows for a more effective analysis of resource consumption and precision since it processes images at various levels of detail and avoids overlapping between extracted features.

2. Related Works

Automated retinal diagnostics using Retinal fundus pictures facilitate the timely identification of several disease disorders. Applying low-level statistical characteristics in these diagnoses efficiently detects different retinal diseases [17]. Artificial intelligence, specifically transformer-based models such as SWIN Transformer, has been progressively used in medical imaging technologies to identify and diagnose different illnesses at an early stage [18]. A unique transformer-based SWIN-T ROP model has been created to accurately distinguish ROP from normal neonatal fundus pictures. This model has shown promising outcomes in the automated identification of ROP zones [19]. This technological development not only assists in the early detection of retinal illnesses but also enables a more accurate assessment of diseases, potentially enhancing patient outcomes [20]. Incorporating transformer-based models such as the SWIN Transformer into automated medical imaging systems has significant promise to improve the early identification and diagnosis of several disorders, including ROP zones [7]. This section critically examines the pertinent literature and imparts essential knowledge that forms the basis for the proposed methodology.

In 2021, Chioma *et al.,* [18] Presented SWINIR, a robust base model for picture restoration that uses the SWIN Transformer's capabilities. Human resource reconstruction units, deep feature extraction and shallow feature extraction make up SWINIR. By consistently outperforming the competition across six separate scenarios, the SWINIR model proved its mastery of every facet of image restoration.

In 2022, Liao *et al.,* [21] Presented the SWIN-PANet model, which incorporates a window-based self-attention mechanism utilizing the SWIN switch into a pre-existing supervision network. For melanoma diagnosis utilizing computer-aided diagnosis (CAD), the suggested SWIN PANet was used to take advantage of this change and increase segmentation accuracy. The model performed very well compared to newer models. However, it has limitations regarding transfer learning.

In 2022, Li *et al.,* [22] Proposed a continuous Wavelet sliding transformer called DnSWIN for real-world image denoising. It uses a convolutional neural network (CNN) encoder to extract bottom features from noisy input images, extracting high-frequency and low-frequency information and building frequency dependencies. Using a CNN decoder, the WSWT uses discrete wavelet transform, self-attention and inverse DWT to extract deep features and reconstruct them into denoised images. The proposed method outperforms state-of-the-art methods.

In 2022, Gu *et al.,* [23], presented a novel approach that integrates SWIN transformer blocks and a lightweight U-Net type model with a HarDNet blocks-based encoder-decoder structure to enhance the accuracy and speed of stroke diagnosis using MRI images. The STHarDNet model underwent evaluation using the ATLAS dataset, which consists of 229 T1-weighted MRI images depicting anatomical tracings of lesions following a stroke. The model attained superior performance compared to state-of-the-art models U-Net, SegNet, PSPNet, FCHarDNet, TransHarDNet, SWIN Transformer, SWIN UNet, X-Net and D-UNet, with Dice, IoU, precision and recall values of 0.5547, 0.4185, 0.6764 and 0.5286, respectively. This approach seeks to surpass the constraints of traditional models in MRI segmentation, enabling a more efficient and precise diagnosis of strokes.

In 2022, Hao *et al.,* [24] Suggested the two-stream SWIN transformer network (TSTNet) as a solution for remote sensing problems. The two streams that make up TSTNet are the edge stream and the original stream. Both streams use deep features from edges and images to make predictions. A SWIN transformer supports each stream and the edge stream incorporates a differentiable edge Sobel operator module (DESOM) for robust edge information suppression and adaptive learning. According to experimental results, TSTNet works better than cutting-edge techniques.

In 2023, Dihin *et al.,* [25], The research presented a new approach to automatically detecting the degree of diabetic retinopathy progression by integrating wavelet and multi-wavelet transformations with a SWIN Transformer. Using the multi-wavelet transform to glean helpful information is a groundbreaking innovation in this research. A novel approach is developed at the feature extraction stage by incorporating the resultant photos into the SWIN Transformer model. Using a dataset including 3662 photographs, the researchers conducted their investigation. Impressively, the experimental training accuracy was 97.78% and the test accuracy was 97.54%. A maximum of 98.09% accuracy was achieved throughout the training process.

In contrast, a testing accuracy of 82% was achieved when the multi-wavelet method was applied to multiclass classification; the validation and training accuracies were 91.60% and 82.42%, respectively. The results show that the multi-wavelet strategy performs better in the research than other methodologies. The training and test sets show that the model performs exceptionally well on binary classification tasks. It should be noted that the model's accuracy dropped in multiclass classification, highlighting the necessity for additional research and improvement to deal with a wider variety of classification tasks.

In 2023, Sankari *et al.,* [26] aim to separate retinal vessels from fundus pictures using SegNet and extract features using SURF and SIFT Feature Extraction techniques. Four traditional machine learning classifiers categorize normal and ROP retinal vessels. Based on the transformer architecture and SWIN-T, a unique ROP classification model is explicitly created to distinguish between ROP and normal Neonatal fundus pictures. The performance of the proposed QSVM model is evaluated in comparison to Resnet50, DarkNet19 and traditional classifiers. The research used a dataset of 200 fundus pictures comprising 100 normal newborn retinal images and 100 neonatal retinal images showing signs of ROP. The machine learning classifiers demonstrate 86.7%, 75%, 74% and 76.5% classification accuracies when distinguishing between ROP and normal retinal pictures. ResNet50 and DarkNet19, which are deep learning networks, attain 92.87% and 89% accuracy rates, respectively. The Quantum machine learning classifier surpasses traditional classifiers regarding classification accuracy, sensitivity and specificity. The suggested approach provides a precise diagnosis of ROP based on newborn fundus pictures, which might assist in point-of-care diagnosis in places with limited healthcare services.

In 2024, Haque *et al.,* [27] introduced the SWIN Transformer architecture to learn global context information. They applied it in the classification of fundus images into five different levels of diabetic retinopathy: no apparent retinopathy, mild non-proliferative DR (NPDR), moderate NPDR, severe NPDR, neovascularization and vitreous/pre-retinal haemorrhage (PDR). They used a publicly available dataset of fundus images with DR annotations for training and evaluation. They measured the model performance using accuracy and area under the ROC curve (AUC) for each category— where the SWIN Transformer model outperformed the previous leading research by 56.8% (obtained from American University in Cairo at 83.4%) on the discriminatory solid capacity for each category. This work surpasses other deep learning structures used in prior research, demonstrating the effectiveness of SWIN Transformers specifically for DR classification tasks. This work illustrates the efficacy of SWIN transformers in accurately and reliably classifying diabetes mellitus (DR) on eye-bed images. This method can enhance the automated DR screening systems, assist in prompt detection and rapid intervention, improve patient outcomes and avoid visual impairment.

3. Materials and Methods

3.1 Dataset

The photographs were captured in the Private Clinic Al-Amal Eye Centre in Baghdad, Iraq. The photos were obtained by skilled professionals utilizing a RetCam3 imaging device. This facility, dedicated to a specific purpose, has been offering ROP screening services for many years. A total of 1365 fundus images were obtained from ROP screening between 2015 and 2020.

3.2 Labelling

The study involves two experienced ophthalmologists specializing in ROP treatment with over 15 years of clinical experience. The specialists allocated three classification zones to each of the fundus photographs. Before comparing the photos, the three ophthalmologists individually classified them to identify any discrepancies in the labelling process, precisely to determine if the specialists assigned different labels to the same image. The labels were arranged collectively after a discussion among the experts and a specific label was assigned to the images.

3.3 Preprocessing

The fundus photographs had a resolution of 640 by 480 pixels. Nevertheless, their dimensions were reduced to 224x224 when inputted into our deep-learning models. We utilized data from a total of 1029 patients for training. The study excluded photographs that were indistinct, hazy or poorly lit. We examined fundus photos depicting different zones of ROP in a single infant, ensuring no overlaps between patients in the training and test datasets. The dataset used for training, evaluation and testing of the model is randomly divided, as outlined in Table 1.

Table 1
ROP zone dataset [7]

	Zone 1	Zone 2	Zone 3
Train set (70%)	305	286	364
Validation set (10%)	44	41	52
Test set (20%)	87	82	104
Total	436	409	520

3.4 Data Augmentation

Overfitting may arise during training when the model is developed with limited data. We employed data augmentation techniques to address this issue to create new retinal fundus images based on the existing training dataset. Data augmentation was utilized to generate additional datasets. In this inquiry, we utilized augmentation strategies like rotation range [3, 3], width shift range [0.1, 0.1], height shift range [0.1, 0.1], zoom range [0.85, 1.15] and horizontal flip. The training dataset was augmented by a factor of seven, resulting in a total of 18,808 images for training.

3.5 Training Procedure and Hyperparameters

In the training phase of the SWIN Transformer model for early detection of ROP in fundus images, specific hyperparameters were meticulously chosen to optimize the learning process. The learning rate of 0.001 facilitated the balanced convergence rate during improvement. Using the 64-image batch size (BS) for each frequency, it achieved a trade-off for the generalization of computational efficiency. The module was trained 200 times, ensuring comprehensive exposure to the data set for optimal learning of advantages. To assess the model performance and prevent overprocessing, a 15% cross-validation split was applied, with a distinct sub-cluster dedicated to performance evaluation during training. These super-markers, including the learning rate, the size of the batch, the number of Epochs and the validation partitioning, are adjusted through experience to balance effective convergence with strong dissemination, thus enhancing the effectiveness of the SWIN transformer in the early detection of ROP.

3.5.1 Proposed methodology

The SWIN Transformer architecture is intricately designed, as shown in Figures 2 and 3, commencing with the ‘Patch Partition’ block responsible for segmenting input images into smaller patches. Subsequently, four stages, each housing one or more SWIN Transformer blocks, are employed to correct and transform features iteratively. At the apex of each stage, patch merging or linear embedding is applied, termed ‘Patch Merging’ or ‘Linear Embedding’ exclusively in the initial layer. This process involves reducing the number of distinctive tokens by a factor of 4, leading to an effective down sampling of resolution by x2. Consequently, a pyramidal-shaped feature map emerges due to varying resolutions at each stage. The final block, denoted ‘Two Successive SWIN Transformer Blocks’, consists of Multi-Layer Perceptron (MLP), Layer Normalization (LN), Window-based Multi-Head Self-Attention (W-MSA) and Multi-Head Self-Attention Module with Regular Windowing (SW-MSA).

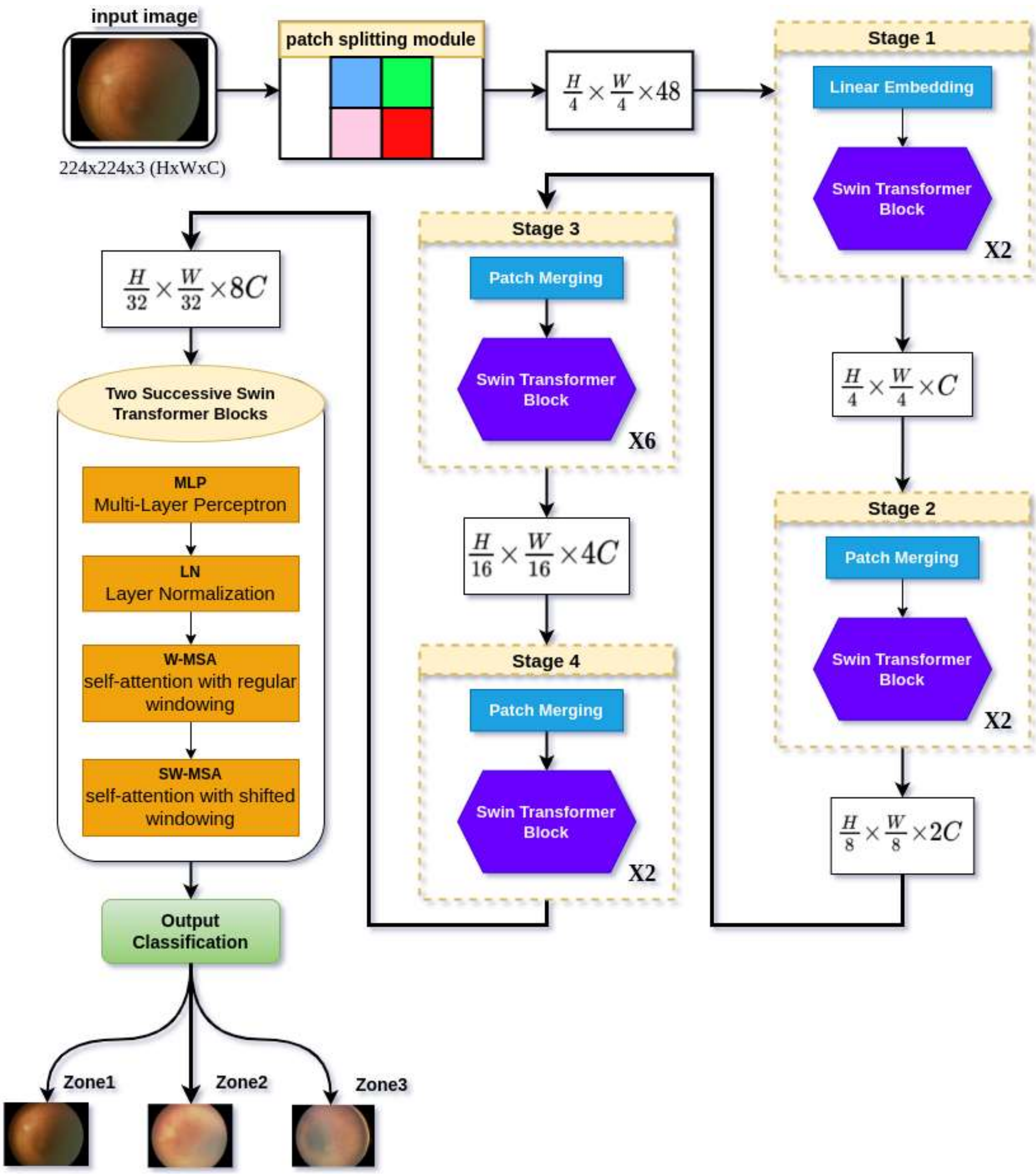


Fig. 2. The suggested method

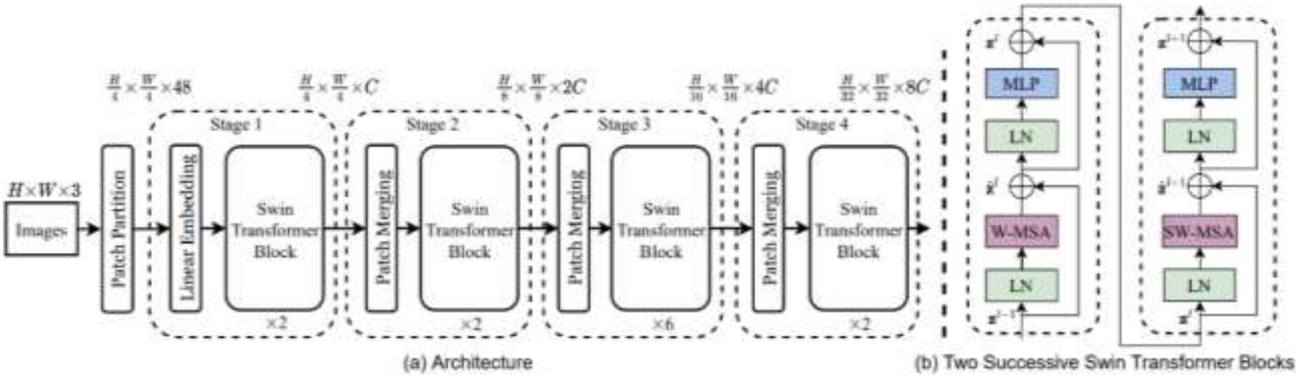


Fig. 3. SWIN Transformer architecture

3.5.2 Evaluation metrics

The study included various metrics, including precision, recall, F-measure and area under the curve, to assess and compare the effectiveness of each trained model. Following the completion of this investigation, the data obtained from the various classifiers were integrated into the voting classifier to predict accuracy. This study evaluated the precision and recall of the trained model in classifying the ROP zones. Regardless of the veracity of these numbers, precision is determined by the degree of accuracy and refers to the proximity between two or more qualities.

Accuracy (ACC) Eq. (1): The proportion of correctly identified samples to total samples:

Accuracy (Acc): ((TP+TN))/((TP+TN+FP+FN)) (1)

Precision (Prec) Eq. (2): Precision is defined as the division of truly positive cases among all examples that we projected to be positive:

Precision (Prec): ((TP))/((TP+FP)) (2)

Recall Eq. (3): the proportion of Positive samples accurately identified as Positive to the total number of Positive models:

Recall: ((TP))/((TP+FN)) (3)

F1 score Eq. (4): The F1 score is the harmonic mean of precision and sensitivity:

F1 Score: 2×(Precision×Recall)/((Precision+recall)) (4)

Area under the curve (AUC): The ROC curve is called the Receiver Operating Characteristics. The integral of the curve is a crucial performance measure that demonstrates the model's ability to distinguish between multiple classes properly. Keep in mind that increasing the elevation of this region will result in a more precise model for identification purposes. The Receiver Operating Characteristic (ROC) curve is computed based on the true positive rate (TPR) and false positive rate (FPR) as in Eq. (5).

$$FPR=FP/(FP+TN) \tag{5}$$

Where, $TP=True\ Positives$, $TN=True\ Negatives$, $FP=False\ Positives$ and $FN=False\ Negatives$.

4. Results and Discussion

The purpose of this research is to find out the zones of ROP in premature newborns. We evaluated the performance of our models in a fine way differentiating ROP from fundus photographs in three different zones. The data sets were fed into four unique classifiers which aimed to forecast the irregularities in the data based on precision, recall, F1-measure and area under curve: though finally SWIN Transformer integrated each classifier's output into one composite accuracy measurement.

4.1 Experimental Setup

The 200 Epochs at a rate of 0.001 per repeat and the size of the 64th batch were what the modules trained with. Adam was the optimizer, with cross-entrepreneur loss as the loss function. The training images were supplemented by data by random flipping and lateral recycling to make the training data set more diverse and less susceptible to over computation— Model training has become more efficient using GPU acceleration: SWIN Transformer underwent comprehensive training independently and was evaluated using the F1 score, recall, accuracy and precision. SWIN Transformer was the most effective in detecting precipitous retinal malformation early, with the best accuracy, precision, recall and F1 degree. The analysis was conducted on an Intel Core i7 computer with a random-access memory (RAM) of 8 gigabytes and a central processing unit at a speed of 2.7 gigahertz. Scikit-learn is an open-source machine learning program based on Python. To make the study analysis more efficient and accessible, we used Google Colab, a free web-based and open-source basic system, to create, share and cooperate in real-time with reports, images, equations and encrypted prose.

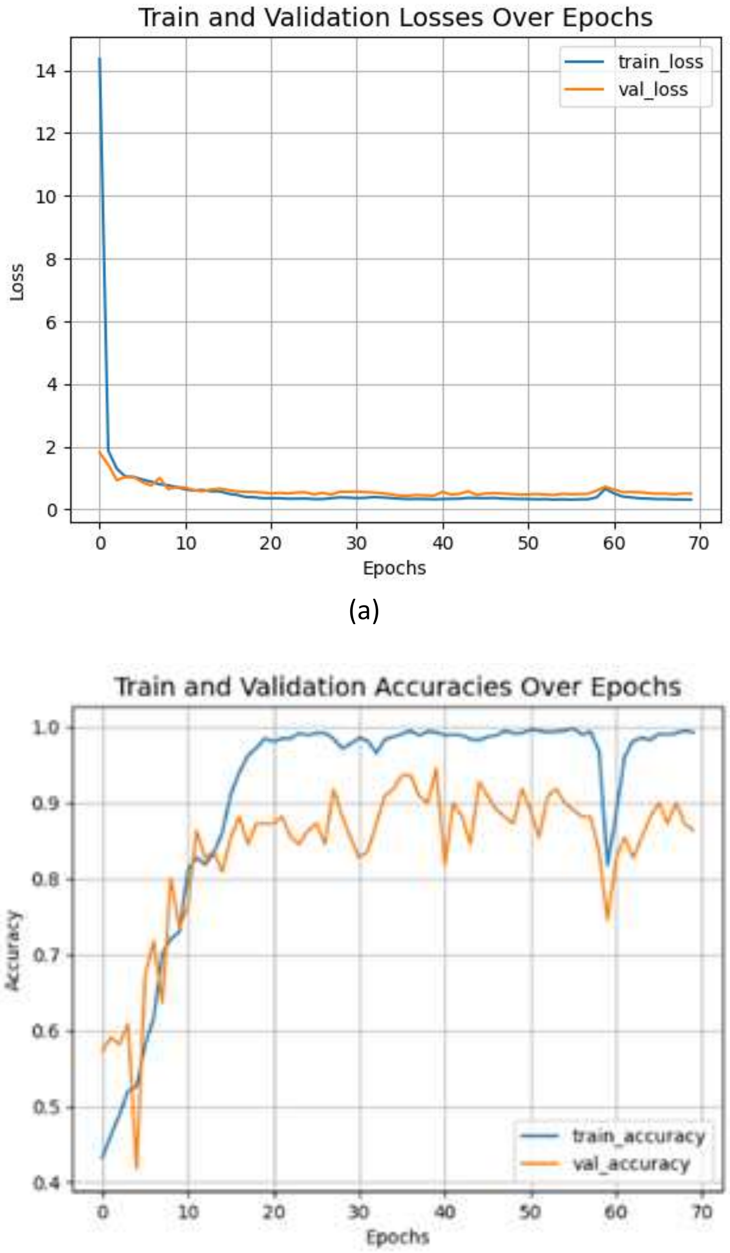
In addition, the models were thoroughly evaluated and validated using different data sets and standards to ensure their robustness and dissemination. The training process involved careful control of super-teachers to improve performance and mitigate potential biases or overprocessing. Furthermore, extensive experiments have explored different architectural designs and formations to achieve the best possible results.

Training has been distributed using more computational methods and involves multiple GPUs and parallel computing. The use of this strategy resulted in the control of expense and effort rationalized to strike an equilibrium with a proper addressing on effective grounds plus managing large dataset sizes, saving time during the training reduction. New models are developed based on refined algorithms and updated organization techniques— with the aim to achieve optimal convergence and prevent misuse from taking place at all.

Large models are expensive to train due to their computational requirements. Such an adoption has spurred the use of distributed training that works on multiple graphic processing units plus a parallel computing framework. This paves way for the effective development of methods and ensures quick data production leading to minimizing time consumption during model training. The algorithms were improved; the models were trained with optimization for convergence in mind: without any wastage.

Moreover, advanced data processing techniques are addressed in the training line, including the computation of advantages and normalization measurement before the dimension reduction. These preprocessing steps helped solidly combine the input data and extract features that would benefit model learning to make accurate predictions later. Moreover, the training line involves advanced data processing techniques such as measuring benefits, normalization and dimensionality reduction. These preparatory steps assisted in solidifying the input data and bringing out features that would benefit effective learning by models developed to ensure accurate predictions were made.

The merger of advanced model structures, strict training procedures, state-of-the-art improvement algorithms and advanced computational resources generally resulted in exceptional performance effectiveness in early mesh detection. The knowledge gained from this study contributes significantly to the future development of an analysis of medical photography, which means that more ideas can be obtained from such research. This is an up-and-coming area that demonstrates the ability of automated learning to enhance healthcare outcomes. The visual representation of the training process is shown in Figure 4.



(a)

(b)

Fig. 4. Training and validation over epochs for ROP zones dataset: (a) loss (b) accuracy

As shown in Table 2, the model demonstrates significant learning and generalization improvements throughout the epochs. In the initial epochs (10-40), training accuracy rises from 87.27% to above 98%, with training loss dropping from 0.5556 to around 0.3164. Validation accuracy improves from 80.00% to 83.64%, while validation loss decreases overall from 0.6393 to around 0.5589. In the middle epochs (50-100), training accuracy remains high at around 99%, with training loss slightly decreasing and validation accuracy peaking at 86.36%, though validation loss fluctuates around 0.5 to 0.56. Training accuracy is nearly perfect in the later epochs (150-200), between 99.39% and 99.90%, with a low training loss of around 0.2998 to 0.3121. Validation accuracy stabilizes around 85.45% to 86.36% and validation loss fluctuates but stabilizes around 0.54 to 0.56. These results collectively indicate an efficiently performing model with strong learning capabilities and effective generalization to novel, unseen data, as evidenced by robust accuracy and low loss values across training and validation datasets.

Table 2
Accuracy and loss for SWIN transformer

Epoch	Train-Acc	Train-Loss	Val-Acc	Val-Loss
10	0.8727	0.5556	0.8000	0.6393
20	0.9868	0.3319	0.8091	0.5736
30	0.9949	0.3164	0.8182	0.5480
40	0.9837	0.3483	0.8364	0.5589
50	0.9949	0.3144	0.8636	0.5062
60	0.9929	0.3344	0.8636	0.5620
70	0.9949	0.3170	0.8545	0.4970
80	0.9949	0.3094	0.8636	0.5067
90	0.9949	0.3407	0.8455	0.5400
100	0.9929	0.3240	0.8273	0.5461
150	0.9939	0.3121	0.8636	0.5564
200	0.9990	0.2998	0.8545	0.5416

4.2 Confusion Matrix

A confusion matrix is a tabular representation that provides a concise summary of the number of accurate positive predictions, inaccurate positive forecasts, accurate negative predictions and inaccurate negative predictions for each class. Figure 5 displays the confusion matrix of the SWIN transformer. The table's rows correspond to the data's true labels, while the columns correspond to the anticipated labels. The figures in the table indicate the frequency with which the algorithm accurately or inaccurately identified each incident. The confusion matrix in this image displays the predicted labels as "zone1", "zone2" and "zone3", but the actual labels also correspond to "zone1", "zone2" and "zone3".

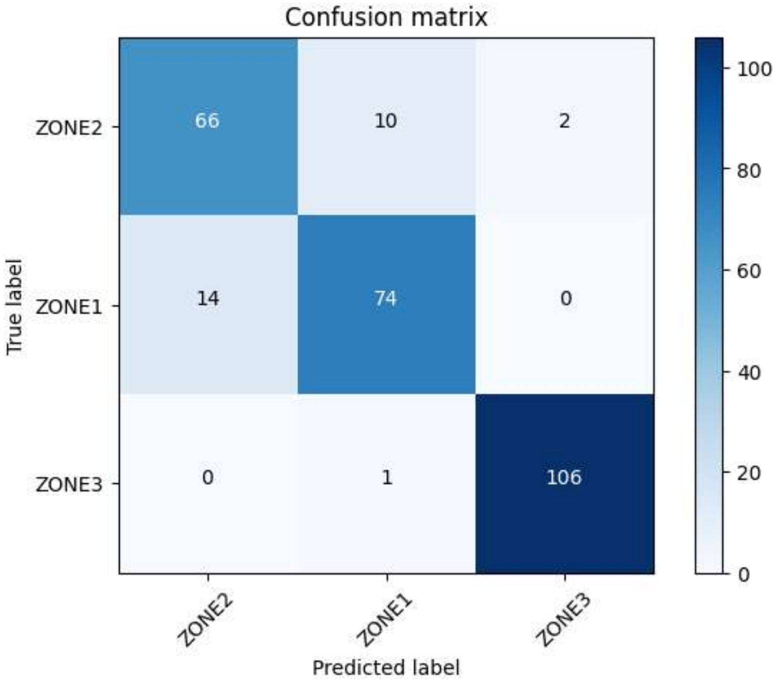


Fig. 5. Confusion matrix of SWIN transformer

The structure of the SWIN Transformer was designed in a complex manner, starting with the "Patch Partition" mass responsible for splitting the images into more minor corrections. First, it optimally ensures the distribution of inputs for further processing. The structure consists of four stages after this initialization, where each stage has one or more SWIN transformer blocks that make the features right and then change them into another domain. These blocks are the core of architecture; they help to collect information from different levels by capturing them properly and act as a means to integrate the advantages linearly at later stages (at the top). This process is called Patch Merging or Linear Embedding, performed only in the primary layer where correction is merged so that distinctive symbols can be reduced, which is 4x less than their actual number, leading to a reduction of accuracy by x2. This ensures structural adaptability towards images with varying dimensions while maintaining mathematical efficacy post-reductions.

Consequently, a pyramidal-shaped feature map emerges, reflecting the different resolutions at each stage. The final block, "Two Successive SWIN Transformer Blocks," incorporates two SWIN Transformer blocks in sequence. This arrangement allows for an even more intricate transformation of features and enhances the overall expressiveness of the architecture. Two essential components are significant within these blocks: Multi-Layer Perceptron (MLP) and Layer Normalization (LN). The MLP layers are responsible for updating distinctive token features after self-attention computation. This type of artificial neural network, characterized by multiple layers of linear and non-linear transformations, contributes to the architecture's ability to capture complex patterns and dependencies in the visual data.

However, LN layers act by normalizing inputs in the feature dimension, which can be a great way to maintain stable input distribution. In turn, it improves training stability and thus boosts overall model performance. Moreover, SWIN Transformer has two self-attention mechanisms: Window-based Multi-Head Self-Attention (W-MSA) and Multi-Head Self-Attention Module with Regular Windowing (SW-MSA). W-MSA differs from the standard multi-head self-attention mechanism as it computes attention scores within local windows— capturing local dependencies based on different token features that help in efficient information exchange within limited contexts.

On the other hand, SW-MSA represents a variant of W-MSA, where windows are shifted by half their size along the image's original height and width dimensions. By enabling communication across different windows, SW-MSA layers excel in capturing long-range dependencies in distinctive token features. This capability proves particularly valuable in scenarios where contextual understanding across the entire image is necessary. These components' comprehensive structure and integration contribute to the SWIN Transformer's efficacy in handling complex visual tasks. These tasks can range from image classification, where the architecture accurately predicts the class labels of images, to image segmentation, where it accurately identifies and delineates objects or regions of interest within an image. The SWIN Transformer's ability to capture both local and long-range dependencies, combined with its effective down sampling and feature refinement operations, makes it a suitable choice for a wide range of real-world applications requiring sophisticated image analysis and understanding.

4.3 Comparison with Previous Studies

The research paper stems from the predecessor study based on the Private Clinic Al-Amal Eye Centre data set in Baghdad, Iraq. It contained 1365 ROP screening fundus images taken from 2015 through 2020. The previous investigation used deep learning algorithms and the voting classifier [7]. In the proposed method, using three steps— namely, image preprocessing, feature extraction through deep learning models and classification by voting classifier— is reported to result in 88.82% accuracy for predicting ROP zones. Instead, this work applied the SWIN Transformer model, which is a learning structure that is developed for the classification of images.

As a result, our model achieved an accuracy of 90.11%, a notable improvement over the previous study's best accuracy of 88.82% [7]. The advancement highlights the capability of transformer-based models in the analysis of medical images that help in early diagnosis of retinopathy of prematurity, which is typically noted for preterm infants. The significant performance improvement demonstrated by transformer-based models underscores the potential use of these systems to promote automated diagnostic systems and bring about a complete change in early diagnosis and treatment— which would then lead to improved health results for preterm infants at risk ROP.

5. Conclusion

This research paper provides a powerful and effective methodology for detecting ROP zones in premature infants using deep learning models, specifically SWIN Transformer's structure. This study emphasizes artificial intelligence-based systems' central and transformative role in revolutionizing health care, especially in complex and challenging areas such as paediatrics.

However, as with any scientific study, it is essential to recognize and address the limitations inherent in our research. One such limitation is the data set size used for training and evaluation purposes. Expanding the data set to include a more diverse range of cases and integrating external verification of diverse populations and clinical settings would significantly enhance the strength and generality of our findings. These future endeavours will undoubtedly improve the credibility and applicability of our proposed methodology, ensuring its smooth integration into real-world clinical functioning and increasing advanced healthcare systems. Despite these recognized limitations, our proposed methodology offers a promise and tremendous potential for strengthening existing healthcare systems by providing an early and accurate diagnosis of ROP, thereby facilitating timely interventions and ultimately improving healthcare outcomes for premature children at risk affected by ROP and other visually threatening situations.

Given the future, our future research will focus primarily on expanding the data set to a wider range of situations, enabling us to further enhance the performance of our models through rigorous testing and validation of external data sets. In addition, we will seriously emphasize the smooth integration of our artificial intelligence system into the clinical process, ensuring its smooth adoption and use by medical specialists in real-world scenarios. By tirelessly pursuing these research methods, we aspire to contribute to artificial intelligence-based healthcare solutions continued and ruthless progress. Ultimately, by harnessing the power of advanced technology, we seek to move this area forward, achieving improved diagnostic accuracy, facilitating timely interventions and ensuring better healthcare outcomes for vulnerable children at risk and struggling with retinal and other diseases. It's likely the stressful conditions that threaten vision.

In the future, we will concentrate on the continued development of algorithms, more methodologies and the production of a larger training dataset, all of which will aid in advancing medical reform in the current circumstances.

Acknowledgement

This research was not funded by any grant.

References

[1] Berrocal, Audina M., Kenneth C. Fan, Hasenin Al-Khersan, Catherin I. Negron and Timothy Murray. "Retinopathy of prematurity: advances in the screening and treatment of retinopathy of prematurity using a single center approach." *American journal of ophthalmology* 233 (2022): 189-215. https://doi.org/10.1016/j.ajo.2021.07.016

[2] Chiang, Michael F., Graham E. Quinn, Alistair R. Fielder, Susan R. Ostmo, RV Paul Chan, Audina Berrocal, Gil Binenbaum *et al.,* "International classification of retinopathy of prematurity." *Ophthalmology* 128, no. 10 (2021): e51-e68. https://doi.org/10.1016/j.ophtha.2021.05.031

[3] International Committee for the Classification of Retinopathy of Prematurity. "The international classification of retinopathy of prematurity revisited." *Archives of Ophthalmology (Chicago, Ill.: 1960)* 123, no. 7 (2005): 991-999. https://doi.org/10.1001/archopht.123.7.991

[4] Agrawal, Ranjana, Sucheta Kulkarni, Rahee Walambe and Ketan Kotecha. "Assistive framework for automatic detection of all the zones in retinopathy of prematurity using deep learning." *Journal of Digital Imaging* 34, no. 4 (2021): 932-947. https://doi.org/10.1007/s10278-021-00477-8

[5] World Health Organization. "Preterm birth." https://www.who.int/news-room/fact-sheets/detail/preterm-birth

[6] Bancalari, Aldo and Ricardo Schade. "Update in the treatment of retinopathy of prematurity." *American journal of perinatology* 39, no. 01 (2022): 022-030. https://doi.org/10.1055/s-0040-1713181

[7] Salih, Nazar, Mohamed Ksantini, Nebras Hussein, Donia Ben Halima, Ali Abdul Razzaq and Sohaib Ahmed. "Prediction of ROP zones using deep learning algorithms and voting classifier technique." *International Journal of Computational Intelligence Systems* 16, no. 1 (2023): 86. https://doi.org/10.1007/s44196-023-00268-9

[8] Salih, Nazar, Mohamed Ksantini, Nebras Hussein, Donia Ben Halima, Ali Abdul Razzaq and Sohaib A. Mahmood. "Detection of Retinopathy of Prematurity Stages Utilizing Deep Neural Networks." In *Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 1*, pp. 699-706. Singapore: Springer Nature Singapore, 2022. https://doi.org/10.1007/978-981-19-1607-6_62

[9] Salih, Nazar, Mohamed Ksantini, Nebras Hussein, Donia Ben Halima and Sohaib Ahmed. "Deep learning models and fusion classification technique for accurate diagnosis of retinopathy of prematurity in preterm newborn." *Baghdad Science Journal* 21, no. 5 (2024): 21. https://doi.org/10.21123/bsj.2023.8747

[10] American Academy of Pediatrics. "An international classification of retinopathy of prematurity." *Pediatrics* 74, no. 1 (1984): 127-33. https://doi.org/10.1542/peds.74.1.127

[11] Aaberg, Thomas, Isaac Ben-Sira, Steve Charles, John Clarkson, Ben Zane Cohen, John Flynn, Robert Foos *et al.,* "An international classification of retinopathy of prematurity: II. The classification of retinal detachment." *Archives of ophthalmology* 105, no. 7 (1987): 906-912. https://doi.org/10.1001/archopht.1987.01060070042025

[12] International Committee for the Classification of Retinopathy of Prematurity. "The international classification of retinopathy of prematurity revisited." *Archives of Ophthalmology (Chicago, Ill.: 1960)* 123, no. 7 (2005): 991-999. https://doi.org/10.1001/archopht.123.7.991

[13] Phelps, Dale L. and ETROP Cooperative Group. "The Early Treatment for Retinopathy of Prematurity study: better outcomes, changing strategy." *Pediatrics* 114, no. 2 (2004): 490-491. https://doi.org/10.1542/peds.114.2.490

[14] Obaid, Ahmed Mahdi, Aws Saad Shawkat and Nazar Salih Abdulhussein. "Original Research Article A powerful deep learning method for skin cancer detection." *Journal of Autonomous Intelligence* 7, no. 1 (2024). https://doi.org/10.32629/jai.v7i1.1156

[15] Obaid, Ahmed Mahdi, Aws Saad Shawkat and Nazar Salih Abdulhussein. "Exploring the potential of A-ResNet in person-independent face recognition and classification." *Int J Adv Netw Monit Controls* 8, no. 2 (2023): 12-9. https://doi.org/10.2478/ijanmc-2023-0052

[16] Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin and Baining Guo. "SWIN transformer: Hierarchical vision transformer using shifted windows." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012-10022. 2021. https://doi.org/10.1109/ICCV48922.2021.00986

[17] Cen, Ling-Ping, Jie Ji, Jian-Wei Lin, Si-Tong Ju, Hong-Jie Lin, Tai-Ping Li, Yun Wang *et al.,* "Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks." *Nature communications* 12, no. 1 (2021): 4828. https://doi.org/10.1038/s41467-021-25138-w

[18] Chioma, Roberto, Annamaria Sbordone, Maria Letizia Patti, Alessandro Perri, Giovanni Vento and Stefano Nobile. "Applications of artificial intelligence in neonatology." *Applied Sciences* 13, no. 5 (2023): 3211. https://doi.org/10.3390/app13053211

[19] Salih, Nazar, Mohamed Ksantini, Nebras Hussein, Donia Ben Halima, Ali Abdul Razzaq and Sohaib Ahmed. "An Advanced Approach for Predicting ROP Stages: Deep Learning Algorithms and Belief Function Technique." *Iraqi Journal of Science* (2024). https://doi.org/10.24996/ijs.2024.65.7.39

[20] Hassan, Bilal, Hina Raja, Taimur Hassan, Muhammad Usman Akram, Hira Raja, Alaa A. Abd-Alrazaq, Siamak Yousefi and Naoufel Werghi. "A comprehensive review of artificial intelligence models for screening major retinal diseases." *Artificial Intelligence Review* 57, no. 5 (2024): 111. https://doi.org/10.1007/s10462-024-10736-z

[21] Liao, Zhihao, Kai Xu and Neng Fan. "SWIN transformer assisted prior attention network for medical image segmentation." In *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*, pp. 491-497. 2022. https://doi.org/10.1145/3532213.3532287

[22] Li, Hao, Zhijing Yang, Xiaobin Hong, Ziying Zhao, Junyang Chen, Yukai Shi and Jinshan Pan. "DnSWIN: Toward real-world denoising via a continuous Wavelet Sliding Transformer." *Knowledge-Based Systems* 255 (2022): 109815. https://doi.org/10.1016/j.knosys.2022.109815

[23] Gu, Yeonghyeon, Zhegao Piao and Seong Joon Yoo. "STHarDNet: SWIN transformer with HarDNet for MRI segmentation." *Applied Sciences* 12, no. 1 (2022): 468. https://doi.org/10.3390/app12010468

[24] Hao, Siyuan, Bin Wu, Kun Zhao, Yuanxin Ye and Wei Wang. "Two-stream SWIN transformer with differentiable sobel operator for remote sensing image classification." *Remote Sensing* 14, no. 6 (2022): 1507. https://doi.org/10.3390/rs14061507

[25] Dihin, Rasha Ali, Ebtesam AlShemmary and Waleed Al-Jawher. "Diabetic retinopathy classification using SWIN transformer with multi wavelet." *Journal of Kufa for Mathematics and Computer* 10, no. 2 (2023): 167-172. https://doi.org/10.31642/JoKMC/2018/100225

[26] Sankari, VM Raja, U. Umapathy, Sultan Alasmari and Shabnam Mohamed Aslam. "Automated detection of retinopathy of prematurity using quantum machine learning and deep learning techniques." *IEEE Access* 11 (2023): 94306-94321. https://doi.org/10.1109/ACCESS.2023.3311346

[27] Haque, Md Mominul, Sweety Akter and Adnan Ferdous Ashrafi. "SWINMedNet: Leveraging SWIN Transformer for Robust Diabetic Retinopathy Classification from the RetinaMNIST2D Dataset." In *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, pp. 1286-1291. IEEE, 2024. https://doi.org/10.1109/ICEEICT62016.2024.10534544