Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Enhancing Ocular Disease classification with ConvNeXt-Tiny

Fatimazahra IDOUZGARH

fatimazahra.idouzgar@enim.ac.ma

Ecole National des Mines de Rabat

Zineb FARAHAT

Ecole National des Mines de Rabat

Nabila ZRIRA

Université Internationale Abulcasis des Sciences de la Santé

Sajid RAHIM

McMaster University

Ibtissam Benmiloud

Ecole National des Mines de Rabat

Research Article

Keywords: ODIR-5K, Eye Disease, ConvNeXt-Tiny, Squeeze-and-Excitation Attention, Deep Learning, Multi-Class Classification, Fundus Image Analysis

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11318600/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Enhancing Ocular Disease classification with ConvNeXt-Tiny

Fatimazahra IDOUZGARH[1*], Zineb FARAHAT[1,2], Nabila ZRIRA[1], Sajid RAHIM[3], Ibtissam Benmiloud[1]

[1*]ADOS Team, LISTD Laboratory,National Superior School of Mines Rabat, Morocco.
[2]International Abulcasis University of Health Sciences (UIASS), Rabat, Morocco.
[3]Computing and Software Dept, McMaster University, Hamilton, Canada.

*Corresponding author(s). E-mail(s): fatimazahra.idouzgar@enim.ac.ma;
Contributing authors: farahatzineb1@gmail.com; zrira@enim.ac.ma;
rahims9@mcmaster.ca; benmiloud@enim.ac.ma;

Abstract

Early detection of eye diseases is critical to the prevention of vision loss and timely clinical intervention to improve clinical outcomes. In this study, we are proposing a new deep learning framework for multi-class retinal image classification that is based on the ConvNeXt-Tiny architecture. The proposed approach was evaluated using two publicly available datasets namely, ODIR-5K (8 classes) and Eye Disease (4 classes). We trained and compared two model configurations. First, a baseline ConvNeXt-Tiny and its' enhanced variant which was the augmented with Squeeze-and-Excitation (SE) attention blocks. The baseline model achieved **92%** accuracy on ODIR-5K and **91%** on Eye Disease datasets which are equal or better than present published results. With the addition of SE attention resulted in improved performance with **96%** accuracy on ODIR-5K and **95%** on Eye Disease. F1 scores and recall also improved with difficult classes such as Normal and Glaucoma. These results demonstrate that by integrating channel-wise attention into a lightweight CNN–Transformer hybrid backbone improves feature discrimination and generalization across datasets. Our approach offers a robust and computationally efficient solution for creating automated retinal disease screening.

Keywords: ODIR-5K, Eye Disease, ConvNeXt-Tiny, Squeeze-and-Excitation Attention, Deep Learning, Multi-Class Classification, Fundus Image Analysis

1 Introduction

Diabetes, a metabolic disorder, has shown a dramatic global increase from 108 million cases in 1980 to 422 million in 2014. Estimates predict 600 million cases by 2040 [1]. This condition is associated with several complications, some of which can be severe or life threatening. Among these, diabetic retinopathy (DR) stands out as a major complication and is the leading cause of preventable blindness worldwide. The number of effected individuals with Vision Threatening Diabetic Retinopathy (VTDR) is expected to increase from 37.3 million to 56.3 million by 2030. This poses a significant public health burden [2].

Early detection of DR is important in preventing vision loss. It requires effective management strategies which include lifestyle management and glycemic control. These have been shown to slow disease progression [3]. However, in advanced DR stages, treatments such as intra-vitreal injections, and laser treatments may be required [4]. Despite the clear benefits of timely diagnosis, DR screening remains limited especially in developing countries where there is a shortages of trained ophthalmologists, and inadequate healthcare facilties with relevant diagnostic equipment [5].

Artificial Intelligence (AI) has shown promising results for large scale DR screening. AI technologies when combined with fundus imaging, and Visual Acuity (VA) and Intraocular Pressure (IOP) measurements provides insight into evaluating DR [6]. With the fundus image captured of a dilated eye is able to provide a clear visualization of diabetic retinal lesions. This includes DR clinical features such as microaneurysms, hemorrhages, and exudates. With these features and measurements, an early detection and diagnosis is possible. By incorporating AI into clinical DR screening methods, the resource constraints can be overcome. This also includes improved accessibility and reduced burden on fragile health care systems.

The early stages of DR are characterized by several features. This includes the dilation of the retinal vessels, autoregulation impairment and variations between the macular and peripheral regions. Studying these patterns can offer deeper insights into the pathology of DR progression and its impact on the retina.

Vision impairment is also caused by other conditions [7], which include Age Related Macular Degeneration (AMD), Cataracts, Glaucoma, Hypertensive Retinopathy, and Myopia. AMD is prevalent in older adults. It leads to the loss of central vision resulting from macular damage. There are two types of AMD, dry and wet. The wet form of AMD is more severe [8]. Cataracts are caused by the clouding of the occular lens' resulting in blurred vision [9]. Cataracts are routinely treated with occular implants. Glaucoma results from a group of conditions that can result in optic nerve damage [10]. This is managed by regulating intraocular pressure. Hypertensive Retinopathy is retinal vascular damage that is caused by high blood pressure . Myopia, a very

common occular condition, causes blurred distance vision [11]. It can lead to severe complications, if untreated. It is essential for the early diagnosis and management of these conditions, so that the patients' vision is preserved and better clinical outcomes.

The objective of the proposed study is to fill the gap that exists between artificial intelligence and the field of ophthalmology, specifically in the early and accurate diagnosis of diseases of the retina via the automatic interpretation of fundus images.The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents the medical background. Section 4 describes the proposed methodology. Section 5 reports the experimental results. Finally, Section 6 concludes the paper and outlines future research directions.

2 Related Work

Yang and Yi [12] presented a multi-class fundus disease classifier called DSRA-CNN. Their proposed framework is based on the Xception architecture. This architecture incorporated depth separable convolution modules, residual learning modules, as well as Squeeze-and-Excitation modules. The proposed network was tested using a fundus disease dataset called ODIR. Results showed that it achieved a high degree of accuracy, about 87.90%, when classifying eight different fundus diseases.

Babaqi et al. [13] developed a deep learning classification method for eye diseases, incorporating Convolutional Neural Networks and transfer learning. In their research, they targeted the classification of normal eyes and three common eye ailments, namely diabetic retinopathy, cataracts, and glaucoma. The authors performed a comparative analysis of a conventional CNN and a multiclass transfer learning classification. The findings obtained from their experiments clearly indicated the superiority of the transfer learning method, yielding a 94% accuracy in classification, while the conventional CNN attained a mere 84% accuracy. The findings clearly indicate the efficacy of transfer learning in improving the accuracy level for the classification of eye diseases.

Zhenzhen Lu et al. [14] developed a model with a classification framework for various fundus diseases based on a convolution neural network with a Squeeze-and-Excitation attention mechanism. The classification accuracy was evaluated on the ODIR dataset in terms of normal cases as well as seven major fundus diseases. The accuracy of the model was 94.27%.

Xo'jayev and Abdullayev [15] investigated the classification of eye diseases using lightweight deep learning architectures, specifically the MobileNetV3 and EfficientNet models. The study focused on identifying common eye diseases associated with diabetes, such as diabetic retinopathy, cataracts and glaucoma. The experimental results showed that MobileNetV3 achieved an accuracy of 73%, whereas EfficientNetB3 achieved a significantly higher accuracy of 94%.

Shumoos Al-Fahdawi et al. [1] developed a hybrid deep learning framework integrating several components—High-Resolution Network (HRNet), Attention Block, Squeeze-and-Excitation Network (SENet) Block, and a Discriminative Restricted Boltzmann Machine (DRBM) for the classification of eight ocular diseases. Their model, trained on the Ocular Disease Intelligent Recognition (ODIR) dataset, achieved

an F1-score of 88.56%, a Kappa score of 88.92%, an AUC of 99.76%, and an overall accuracy of 92.41%.

Fanyu Du et al. [16] utilized ResNet-50 and ResNet-101 architectures to classify multiple ophthalmological disorders using the ODIR dataset. After appropriate preprocessing and training, their approach attained an accuracy of 92%.

Al Rakib et al. [17] proposed the use of an EfficientNet-based convolutional neural network model for automatic classification of retinopathy diseases. This model provides for the classification of diabetic retinopathy, cataract, glaucoma, and normal retinal condition from fundus images. Through leveraging transfer learning and fine-tuning, the model achieves high accuracy of 93.8%, which proves a great balance between these performance metrics and computational power.

Karthikayan et al. [5] proposed a set of dilated ResNet models for retinal fundus image classification. By substituting standard convolutions with dilated convolutions in deeper network layers, they expanded the receptive field, improved classification performance, and reduced computational overhead. Experiments conducted on the ODIR dataset (comprising eight retinal disease classes) showed that dilated ResNet architectures outperformed standard ResNets across five variants (ResNet-18 to ResNet-152), achieving average F1-scores of 0.71, 0.70, 0.69, 0.67, and 0.70, respectively.

Alsohemi and Dardouri [18] introduced a deep learning solution for the classification of retinal disease using a pre-trained EfficientNetB3 architecture, which they fine-tuned using the retinal image Kaggle public data set. Their system classified images of four classes of retinal disease: cataract, retinal diabetic retinopathy, retinal glaucoma, and normal eyes. The system obtained an accuracy rate of 95.12% to confirm the efficiency of their solution in diagnosing retinal disease.

Vidivelli et al. [3] explored transfer learning strategies for ocular disease classification using the ODIR dataset. They evaluated several pretrained CNN architectures—MobileNet, VGG-16, ResNet, DenseNet, and Inception V3—under various optimization settings. Their experiments revealed that MobileNet, combined with the Adam optimizer, achieved the highest testing accuracy of 89.64%, demonstrating its effectiveness for automated eye disease diagnosis.

Sivaz and Aykut [19] presented an image-level and multi-label classification approach to retinal disease classification based on Swin Transformer V2 and Shunted Cross-Attention (SCA) classification heads. Their approach aims to learn global context and preserve multi-scale discriminative features that have been further improved by Adaptive Sharpness-Aware Minimisation and Scalable Neighbour Discriminative Loss to represent interdependence among labels. Final scores of 87.60% and 85.11% were reported based on an average of the Kappa statistic, F1 measure, and the Area Under the Curve (AUC) values associated with off-site and on-site settings, respectively, on the ODIR dataset.

Belharar et Zrira [20]. proposed the classification system named DeepRetino to automatically classify ophthalmic diseases using the power of convolutional neural networks. Moreover, the authors considered the usage of transfer learning with the models of ResNet50, InceptionV3, Inception-ResNetV2, and EfficientNetB7.The proposed method was evaluated using the Ocular Disease Intelligent Recognition (ODIR) dataset, achieving a validation accuracy of 94.05% and an area under the curve (AUC) of 99.3%.

3 Medical Background

3.1 Retina Features

The retina is a thin, light-sensitive layer of tissue which lines the back of the interior of the eye. It comprises millions of photoreceptor cells. Their function is to capture visual information and convey it to the brain via the optic nerve. As illustrated in Figure 1, the retina consists of several key anatomical components. In the center of the retina, is the macula. It is responsible for visual details. Within the macula, lies a small depression known as the fovea. This region has the highest density of cone cells. These facilitate exact color discrimination and provide visual clarity. The optic disk (OD) is a region in the center. Here the retinal nerve fibers merge to make the optic nerve. The optic nerve communicates the signals from the eye to the brain. The venules are narrow veins that transport deoxygenated blood from the retinal tissue. The arterioles are small blood vessels whose function is to deliver oxygen and nutrients to the various retinal layers; this ensures proper cellular function and visual performance.

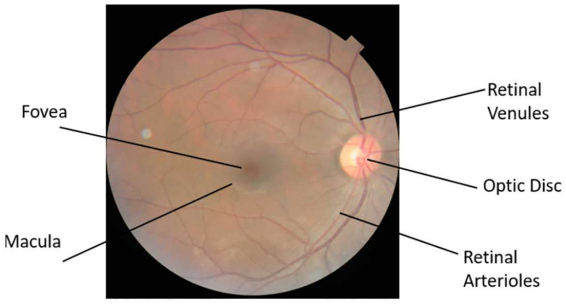


Fig. 1 Retina Features

3.2 Retinal Diseases

3.2.1 Age-Related Macular Degeneration (AMD)

Age Related Macular Degeneration (AMD) is a progressive ocular disorder. It is usually associated with aging [7]. It gradually impairs the central vision over time causing the loss of center vision. Symptoms include blurred vision, the appearance of dark or blank spots in the visual field 2[8]. Objects may appear as distorted in many affected individuals. This also affects colour perception which can appear less vibrant. AMD is characterised with the presence of yellowish deposits, known as drusen, underneath the retina. In advanced stages, there is possibility of abnormal blood vessel growth, This can lead to the swelling or bleeding within the retinal layer further deteriorating the vision. Fig. 2(B).

3.2.2 Glaucoma

Glaucoma, is an eye disorder which can result in irreversible vision loss [21]. It is a result of elevated intraocular pressure (IOP) that arises due to impaired drainage of aqueous humor. The resulting pressure from the fluid build up starts to compress the optic nerve and associated retinal blood vessels. Gradually the peripheral vision loss occurs which if left untreated will result in blindness [10]. Fig. 2(C).

3.2.3 Cataract

A cataract is a common vision impairment. It is primarily caused by aging as the eye's lens loses its ability be transparent [9]. As the lens is cloudy, the resulting image on the retina lacks clarity. When mild, no intervention may be necessary as this has no affect on the daily patient's life. However in advanced stages, cataract surgery may be the only option [9]. Fig. 2(D).

3.2.4 Myopia

Myopia is a common refractive error condition of the eye [22]. The eyeball of the patient is either too long or the cornea is too curved. The resulting image is projected incorrectly placed ahead of the retina. The near vision is clear but distance vision is blurry. It is projected that by 2050, nearly half of the global population will be affected. Myopia tends to be most prevalent in urban areas. Most cases are benign, the severe cases can cause serious complications such as retinal detachment, or glaucoma [11]. Management strategies normally include corrective lenses or refractive surgery. Fig. 2(E).

3.2.5 Diabetic Retinopathy

Diabetic Retinopathy (DR) is a serious eye complication. It is associated with sustained high blood sugar levels. This silently damages the small blood vessels in the retina. Early detection is key step in managing this condition. When the disease is advanced, vision loss can occur. DR is graded by the severity of retinal ischemia and its associated risk to vision loss. It begins as mild to moderate non-proliferative diabetic retinopathy. It can progress to more severe non-proliferative. DR characterized by extensive retinal ischemia. If treatment is not administered, it can progress to proliferative diabetic retinopathy. This appears as abnormal neo-vascularization on the retinal surface or optic disc which increases the risk of vision impairment [23]. Fig. 2(F).

3.2.6 Hypertensive Retinopathy

Hypertensive retinopathy is a retinal condition which is caused by prolonged high blood pressure [24]. This damaged the blood vessels supplying the retina. In the body of the eye, a nutrient rich fluid called aqueous humor is present. It is located between the cornea and the lens. It is replaced every two to three hours. The regulation results in an optimal intraocular pressure. This ranges between 10 to 20 mmHg. Prolonged,

high blood pressure alters this balance and disrupts the retinal circulation. This can lead to structural damage and vision loss.



(A) Normal **(B) AMD** **(C) Glaucoma**

(D) Cataract **(E) Myopia** **(F) DR**

Fig. 2 Visualization of eye conditions [25].

4 Methodology

This section outlines the proposed methodology, encompassing dataset preparation, preprocessing, and model development. The architecture employs a pretrained ConvNeXt-Tiny backbone augmented with a Squeeze-and-Excitation (SE) block, followed by global average pooling, a dense GELU-activated layer with dropout regularization, and a softmax classifier for final prediction. The complete architecture of the framework proposed is shown below figure 3.

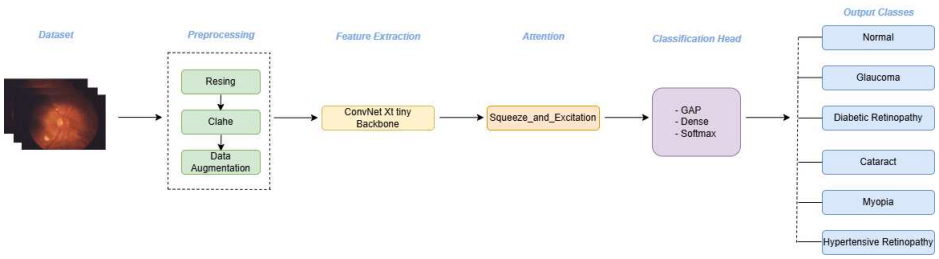


Fig. 3 End-to-end pipeline of the proposed retinal disease classification framework.

4.1 Used Datasets

The study utilizes two publicly available datasets from Kaggle: "Eye Diseases Classification" [26] and "Ocular Disease Intelligent Recognition (ODIR)" [27]. The first dataset contains retinal fundus images categorized into four classes namely Normal, Diabetic Retinopathy, Cataract, and Glaucoma—with approximately 1,000 images per class. These images were collected from various open-access sources, including IDRiD, Ocular Recognition, and HRF

The second dataset referred to as ODIR, is a large scale research opthalmic dataset with multiple ocular conditions to support research. It has data from 5,000 patients, across China. Each record has paired colour fundus photographs of the left and right eyes of each patient as well as their demographic information. The images are labeled into eight categories. These are Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age related Macular Degeneration (A), Hypertension (H), Pathological Myopia (M), and Other diseases/abnormalities (O).

4.2 Preprocessing

The preprocessing pipeline was designed to prepare the input data according to the requirements of convolutional neural networks (CNNs). In this study, image preprocessing was performed in three main stages. First, image quality was enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE). Second, the enhanced images were resized to match the model's input dimensions. Finally, data augmentation techniques were applied to the resized images to increase the variability and robustness of the training dataset.

4.2.1 Image Quality Improvement

The Contrast Limited Adaptive Histogram Equalization (CLAHE) technique was applied to adjust the image intensity distribution and enhance contrast. This method modifies the image's probability density function to produce a more uniform distribution of pixel values, effectively spreading them across the full intensity range (Fig. 4). The parameters used for this process are as follows:

- ClipLimit = 2
- TileGridSize = (8,8)

4.2.2 Image Resizing

Resizing images is an essential preprocessing step in deep learning. It ensures standardisation, consistency of images across the dataset while aiding the CNNs' computational efficiency. In this study, all fundus images were resized to 256×256 pixels using Lanczos interpolation, which preserves structural details while meeting the input requirements of the proposed model.



(a) (b)

Fig. 4 Original image (a). CLAHE image (b)

4.2.3 Data Augmentation

To improve model performance and prevent overfitting, a large and consistently balanced dataset is needed. To mitigate class imbalance and increase the training dataset, augmentation was used in the form of the Augmentor library, a Python package for generating synthetic images. This involved horizontal and vertical flipping transformations. Following augmentation, the dataset was divided into 70% for training, 20% for validation, and 10% for testing.

4.3 Used Model

In this study, we developed a custom classification model built upon the ConvNeXt-Tiny backbone pretrained on ImageNet. ConvNeXt [28] is a newer convolutional neural network (CNN) architecture. Introduced by Facebook AI Research that reimagines ResNet using key design principles from Vision Transformers (ViTs). It employs large convolution kernels (e.g., 7×7) to capture long range dependencies, adopts deeper and more uniform block structures for improved scalability. It replaces Batch Normalization with Layer Normalization for enhanced training stability. It utilizes depthwise separable convolutions to reduce computational complexity.

The base model is used without the top classification head, enabling transfer learning while preserving ConvNeXt's powerful feature extraction capabilities. To improve the representational capacity, we integrated a Squeeze and Excitation (SE) block. This adaptively recalibrates channel wise feature responses. SE block also performs global average pooling, that is followed by two fully connected layers with a bottleneck structure. This is a reduction ratio of 16. A sigmoid activation is used to generate channel attention weights. These weights are then applied to the feature maps via element-wise multiplication. This allows the network to emphasise the informative channels. The features are next aggregated through global average pooling, and passed through a fully connected layer with GELU activation and L2 regularization. It is regularized with dropout. Lastly, a softmax layer outputs the class probabilities corresponding to the number of classes in the ocular datasets. This architecture combines the efficiency of ConvNeXt with the adaptive feature recalibration of SE attention,

improving the model's ability to focus on relevant visual patterns. The whole architecture of ConvNeXt-Tiny model enhanced with Squeeze-and-Excitation (SE) attention is depicted in Figure 5.

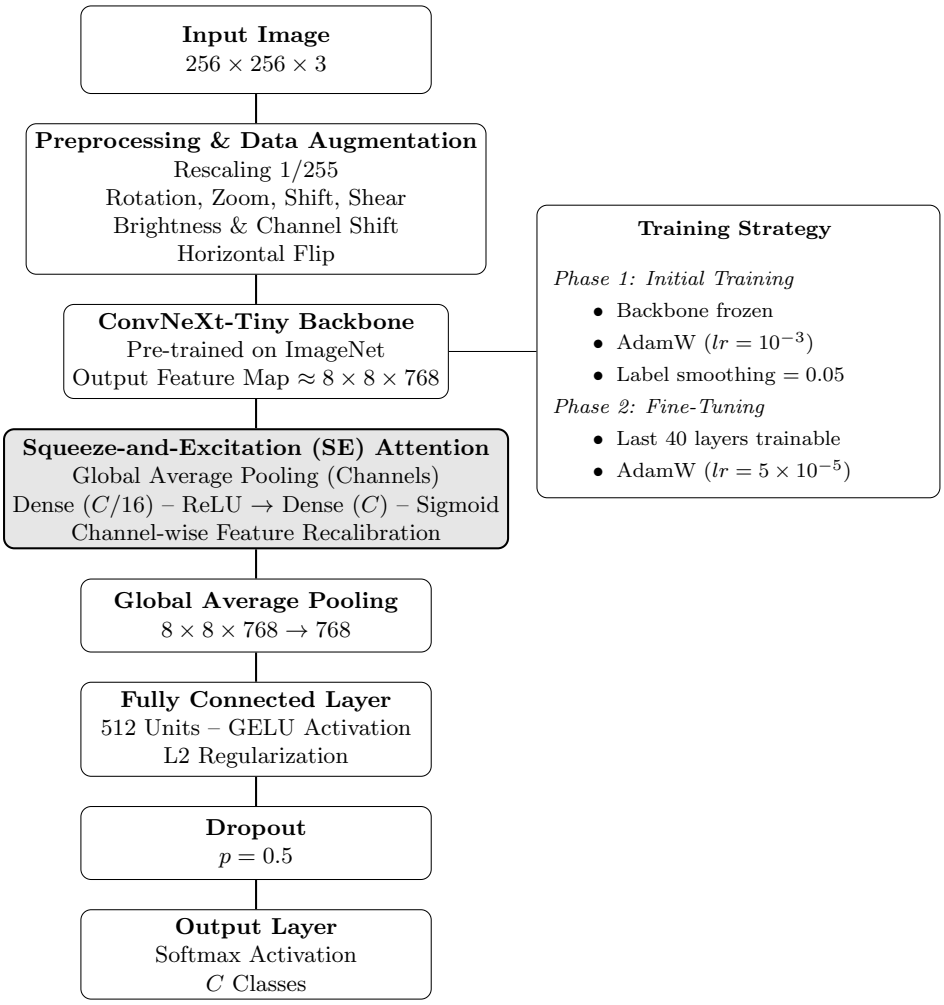


Fig. 5 Architecture of the proposed ConvNeXt-Tiny model enhanced with Squeeze-and-Excitation (SE) attention.

5 Experimental Results

The backbone employed in our experiments is **ConvNeXt-Tiny**, a modern convolutional neural network architecture inspired by transformer design principles. It

incorporates depthwise convolutions, inverted bottleneck blocks, and Layer Normalization, providing an effective balance between accuracy and computational efficiency. By combining the strengths of convolutional networks and vision transformers, this architecture captures both local and global features from retinal fundus images.

To enhance feature representation, two model configurations were evaluated: a baseline ConvNeXt-Tiny and an enhanced version that integrates **Squeeze-and-Excitation (SE) attention blocks** at each stage to adaptively recalibrate channel-wise feature responses.

5.1 Experimental Setup

The model was run on Kaggle. The model training was performed using the following parameters as noted in Table 1.

Table 1 Hyperparameter configuration for model training.

Parameter	**Value**
Input size	$256 \times 256 \times 3$
Batch size	32
Learning rate (initial)	1×10^{-3}
Learning rate (fine-tuning)	5×10^{-5}
Optimizer	AdamW
Epochs	45
Loss function	Label Smoothing Cross-Entropy ($\epsilon = 0.05$)
Output activation	Softmax

5.2 Results on ODIR-5K Dataset

On the ODIR-5K dataset, which comprises eight classes (A: Age-related Macular Degeneration, C: Cataract, D: Diabetes, G: Glaucoma, H: Hypertension, M: Pathological Myopia, N: Normal, O: Other abnormalities), the baseline ConvNeXt-Tiny model achieved a validation accuracy of **92%** and a macro F1-score of **0.92**. The primary challenge occurred in classifying the Normal (N) category, where the recall was 0.56, with frequent misclassifications involving diabetic and other disease classes.

After integrating Squeeze-and-Excitation (SE) attention blocks, the model's validation accuracy improved to 96%, and the macro F1-score increased to 0.96, with a notable improvement in the Normal class (recall rising to 0.74). A detailed performance comparison between ConvNeXt-Tiny with and without SE attention on ODIR-5K is presented in Table 2. As shown in Figure 6, the confusion matrix for the SE-enhanced model exhibits a cleaner diagonal and significantly fewer misclassifications.

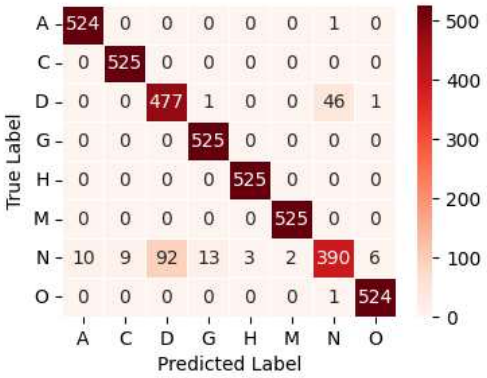


Fig. 6 Confusion Matrix of ODIR-5K Dataset

Table 2 Performance comparison of ConvNeXt-Tiny with and without SE attention on ODIR-5K.

Class	**Without SE**			**With SE**		
	Precision	Recall	F1-score	Precision	Recall	F1-score
A	0.97	0.99	0.98	0.98	1.00	0.99
C	0.97	1.00	0.99	0.98	1.00	0.99
D	0.75	0.86	0.80	0.84	0.91	0.87
G	0.98	0.99	0.98	0.97	1.00	0.99
H	0.99	1.00	0.99	0.99	1.00	1.00
M	1.00	0.98	0.99	1.00	1.00	1.00
N	0.79	0.56	0.65	0.89	0.74	0.81
O	0.91	0.99	0.95	0.99	1.00	0.99
macro avg	0.91	0.91	0.91	0.96	0.96	0.96
weighted avg	0.92	0.92	0.92	0.96	0.96	0.95
Accuracy			0.92			0.96

5.3 Results on Eye Disease Dataset

We also evaluated the model on the Eye Disease dataset. This contained four classes namely, Cataract, Diabetic Retinopathy, Glaucoma, and Normal. The baseline model achieved a validation accuracy of **91%** and a macro F1-score of **0.91**. Glaucoma and Normal achieved lower performance of F1 (0.87) and F1 (0.88) respectively due to several misclassifications between these two classes. After adding SE attention, the validation accuracy increased to **95%** and the macro F1-score to **0.96**. The Normal class showed the most improvement, as recall rose from 0.86 to 0.98, leading to far fewer false negatives.A detailed performance comparison between ConvNeXt-Tiny with and without SE attention on Eye Disease is presented in Table 3. Figure 7 shows the confusion matrix with SE attention, confirming the overall improvement.

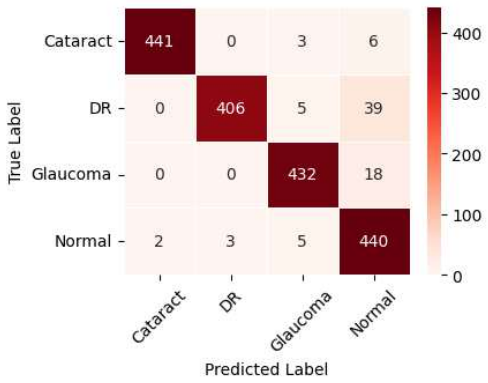


Fig. 7 Confusion Matrix of Eye Disease Datase

Table 3 Performance comparison of ConvNeXt-Tiny with and without SE attention on the Eye Disease dataset.

Class	**Without SE**			**With SE**		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Cataract	0.90	0.95	0.92	1.00	0.98	0.99
DR	0.97	0.98	0.97	0.99	0.90	0.95
Glaucoma	0.88	0.86	0.87	0.97	0.96	0.97
Normal	0.90	0.86	0.88	0.87	0.98	0.92
macro avg	0.91	0.91	0.91	0.96	0.96	0.96
weighted avg	0.91	0.91	0.91	0.96	0.95	0.96
Accuracy			0.91			0.955

5.4 Comparative Study

State-of-the-art methods presented in the following tables (Table 4 for ODIR-5K and Table 5 for Eye Disease). Our baseline model using ConvNeXt-Tiny without attention achieves results that are roughly in line with existing work, reaching **92%** accuracy on ODIR-5K and **91%** on Eye Disease. This is comparable to the typical range of 87–92% and 90–94% reported in the literature. However, upon the integration of SE attention blocks, our model clearly surpasses these benchmarks, achieving **96%** on ODIR-5K and **95%** on Eye Disease, together with higher macro F1-scores. This acknowledges that while, our baseline model performs similarly to current approaches, the use of channel wise attention significantly boosts classification performance. This provides a noticeable advantage over prior work, in particular by improving the recall for challenging classes such as Normal and Glaucoma.

Table 4 Comparison of Different Models and Methods for ODIR-5K Classification

Article	Year	Model / Method	Result
Yang and Yi [12]	2022	Xception + Residual + SE	Accuracy = 87.90%
Lu et al.[14]	2023	CNN + Squeeze-and-Excitation Attention	Accuracy = 94.27%
Al-Fahdawi et al.[1]	2024	HRNet + Attention + SENet + DRBM	Accuracy = 92.41%
Du et al. [16]	2024	ResNet-50 + ResNet-101	Accuracy = 92%
Karthikayan et al. [5]	2024	Dilated ResNet-18/34/50/101/152	F1 = 0.67–0.71
Vidivelli et al. [3]	2025	MobileNet + Adam	Accuracy = 89.64%
Sivaz and Aykut [19]	2025	Swin Transformer V2 + SCA	F1/Kappa/AUC ≈ 85–88%
Belharar and Zrira [20]	2022	EfficientNetB7	Accuracy = 94.05%
Proposed method	**2025**	**ConvNeXt-Tiny + SE attention**	**Accuracy = 96%**

Table 5 Comparison of Different Models and Methods for Eye Disease Classification (4-class datasets)

Article	Year	Model / Method	Result
Babaqi et al.[13]	2023	CNN vs Transfer Learning (VGG, ResNet)	Accuracy = 94%
Xo'jayev and Abdullayev [15]	2023	MobileNetV3 / EfficientNetB3	Best = 94%
Al Rakib et al.[17]	2024	EfficientNet	Accuracy = 93.8%
Alsohemi and Dardouri [18]	2025	EfficientNetB3	Accuracy = 95.12%
Proposed method	**2025**	**ConvNeXt-Tiny + SE attention**	**Accuracy = 95.5%**

6 Conclusion

In this study, we investigated the use of **ConvNeXt-Tiny** for automated classification of retinal fundus images, focusing on two widely used datasets: **ODIR-5K** and **Eye Disease (4-class)**. Our results indicate that the baseline ConvNeXt-Tiny model, without attention modules, performs on par with current state-of-the-art methods, highlighting its strong feature representation capabilities.

The addition of **Squeeze-and-Excitation (SE)** attention modules consistently improved performance across both datasets, achieving **96%** accuracy on ODIR-5K and **95%** on the Eye Disease dataset, with notable gains in recall for challenging classes such as Normal and Glaucoma. These improvements underscore the value of channel-wise feature recalibration in capturing subtle pathological patterns.

Beyond achieving high accuracy, this work demonstrates that lightweight CNN–Transformer hybrid architectures can be effectively enhanced with attention mechanisms to deliver strong performance without excessive computational cost. This

makes the approach well-suited for deployment in clinical decision-support systems, where both reliability and efficiency are critical. Future research may explore extending this framework to additional retinal pathologies, evaluating it on larger and more diverse datasets, and integrating alternative attention mechanisms or ensemble strategies to further improve robustness.

In this study, we explored the use of **ConvNeXt-Tiny** for automatic classification of retinal fundus images using two widely used datasets namely, **ODIR-5K** and **Eye Disease (4-class)**. Our results demonstrate that the baseline model without attention performs comparably to existing state-of-the-art approaches, thereby confirming the performance of ConvNeXt-Tiny. With the integration of **Squeeze-and-Excitation (SE) attention modules** boosted performance across both datasets, achieving **96%** accuracy on ODIR-5K and **95%** on Eye Disease. Noticeable gains were achieved in recall of challenged classes namely, Normal and Glaucoma. These improvements highlight the importance of channel-wise feature recalibration in distinguishing subtle pathological patterns from normal images. In addition to superior results, this work also demonstrates that lightweight CNN–Transformer hybrid architectures can benefit using attention mechanisms resulting in higher accuracy without excessive computational cost. This makes our framework suitable for real-world use in clinical decision support systems, where accuracy, reliability and efficiency are critical. Future work may involve extending these frameworks across newer retinal pathologies, evaluating performance on larger and more diverse datasets, and exploring other attention mechanisms or ensemble strategies to further enhance robustness.

Declarations

Competing Interests

The authors declare no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets analysed during the current study are publicly available. The ODIR-5K dataset is available at https://www.kaggle.com/code/gpreda/ocular-disease-recognition-eda and the Eye Diseases Classification dataset is available at https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification

Ethical Approval

Not applicable. This study used publicly available retinal image datasets and did not involve direct recruitment of human participants or new data collection.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

References

[1] Al-Fahdawi, S., Al-Waisy, A.S., Zeebaree, D.Q., Qahwaji, R., Natiq, H., Mohammed, M.A., Nedoma, J., Martinek, R., Deveci, M.: Fundus-deepnet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images. Information Fusion **102**, 102059 (2024)

[2] Roglic, G.: Who global report on diabetes: A summary. International Journal of Noncommunicable Diseases **1**(1), 3–8 (2016)

[3] Vidivelli, S., Padmakumari, P., Parthiban, C., DharunBalaji, A., Manikandan, R., Gandomi, A.H.: Optimising deep learning models for ophthalmological disorder classification. Scientific Reports **15**(1), 3115 (2025)

[4] Pektaş, M.: Performance analysis of efficient deep learning models for multi-label classification of fundus image. Artificial Intelligence Theory and Applications **3**(2), 105–112 (2023)

[5] Karthikayan, P.N., V, Y.S.V., Kattamuri, H.G., Jayaraman, U.: Explainable AI: Comparative Analysis of Normal and Dilated ResNet Models for Fundus Disease Classification (2024). https://arxiv.org/abs/2407.05440

[6] Babaqi, T., Jaradat, M., Yildirim, A.E., Al-Nimer, S.H., Won, D.: Eye Disease Classification Using Deep Learning Techniques (2023). https://arxiv.org/abs/2307.10501

[7] Yorston, D.: Retinal diseases and vision 2020. Community Eye Health **16**(46), 19–20 (2003)

[8] Hobbs, S.D., Tripathy, K., Pierce, K.: Wet age-related macular degeneration (amd). In: StatPearls [Internet]. StatPearls Publishing, ??? (2024)

[9] Lam, D., Rao, S.K., Ratra, V., Liu, Y., Mitchell, P., King, J., Tassignon, M.-J., Jonas, J., Pang, C.P., Chang, D.F.: Cataract. Nature reviews Disease primers **1**(1), 1–15 (2015)

[10] Rhee, D.J.: Overview of Glaucoma. Merck Manual (2025). https://www.merckmanuals.com/professional/eye-disorders/glaucoma/overview-of-glaucoma

[11] Marcus, M.W., Vries, M.M., Montolio, F.G.J., Jansonius, N.M.: Myopia as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. Ophthalmology **118**(10), 1989–1994 (2011)

[12] Yang, X.-l., Yi, S.-l.: Multi-classification of fundus diseases based on dsra-cnn. Biomedical Signal Processing and Control **77**, 103763 (2022)

[13] Babaqi, T., Jaradat, M., Yildirim, A.E., Al-Nimer, S.H., Won, D.: Eye disease classification using deep learning techniques. arXiv preprint arXiv:2307.10501 (2023)

[14] Lu, Z., Miao, J., Dong, J., Zhu, S., Wu, P., Wang, X., Feng, J.: Automatic multilabel classification of multiple fundus diseases based on convolutional neural network with squeeze-and-excitation attention. Translational Vision Science & Technology **12**(1), 22–22 (2023)

[15] Xo'jayev, O., Abdullayev, H.: Classification of eye diseases with mobilenetv3 and efficientnetb0 models. DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE **1**(1), 92–96 (2023)

[16] Du, F., Zhao, L., Luo, H., Xing, Q., Wu, J., Zhu, Y., Xu, W., He, W., Wu, J.: Recognition of eye diseases based on deep neural networks for transfer learning and improved ds evidence theory. BMC Medical Imaging **24**(1), 19 (2024)

[17] Rakib, A., Billah, M., Ahamed, A., Imamul, H., Masum, M.: Efficientnet-based model for automated classification of retinal diseases using fundus images. European Journal of Computer Science and Information Technology **12**(8), 48–61 (2024)

[18] Alsohemi, R., Dardouri, S.: Fundus image-based eye disease detection using efficientnetb3 architecture. Journal of Imaging **11**(8), 279 (2025)

[19] Sivaz, O., Aykut, M.: Image-level multi-label retinal disease classification with a novel classification head. Computers and Electrical Engineering **124**, 110410 (2025)

[20] Zahra Belharar, F., Zrira, N.: Deepretino: Ophthalmic disease classification from retinal images using deep learning. In: 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), pp. 392–399 (2022). https://doi.org/10.1109/SETIT54465.2022.9875570

[21] Mahabadi, N., Zeppieri, M., Tripathy, K.: Open angle glaucoma. 2024 mar 7. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing (2025)

[22] Bourne, R.R., Flaxman, S.R., Braithwaite, T., Cicinelli, M.V., Das, A., Jonas,

J.B., Keeffe, J., Kempen, J.H., Leasher, J., Limburg, H., *et al.*: Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. The Lancet Global Health **5**(9), 888–897 (2017)

[23] Shukla, U.V., Tripathy, K.: Diabetic retinopathy.[updated 2021 aug 29]. StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing (2024)

[24] Mandal, R., Saha, S.: Systemic Diseases And Ocular Manifestations: An Academic Guide: As Per National Commission For Allied And Healthcare Professions (NCAHP) Syllabus. OrangeBooks Publication, ??? (2025)

[25] A leading researcher explains how to protect your vision — cspi.org. https://www.cspi.org/article/leading-researcher-explains-how-protect-your-vision. [Accessed 07-12-2025]

[26] Doddi, G.V.: Eye Diseases Classification. Kaggle Dataset. https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification

[27] Shanggong Medical Technology Co., Ltd.: Ocular Disease Intelligent Recognition (ODIR-5K). ODIR-2019 Grand Challenge (2019). https://www.kaggle.com/code/gpreda/ocular-disease-recognition-eda

[28] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986 (2022)