International Journal of Scientific Engineering and Science
Volume 10, Issue 4, pp. 82-86, 2026. ISSN (Online): 2456-7361

Synergizing CNNs and Channel-Wise Attention for Enhanced High-Resolution Retinal Diagnostics

Chukwuogo Okwuchukwu Ejike[1], Belonwu Tochukwu Sunday[2], Ezuruka Evelyn Ogochukwu[3], Nwankpa Joshua Makuo[4], Chinedu Emmanuel Mbonu[5]
[1,2,3,4,5]Department of Computer Science, Nnamdi Azikiwe University Awka, Nigeria

***Abstract*—** *Early and accurate diagnosis of ocular diseases such as cataract, diabetic retinopathy (DR), and glaucoma is critical for preventing irreversible vision loss. Manual interpretation of color fundus photographs by specialists is time-consuming, subjective, and constrained by the limited availability of trained ophthalmologists, particularly in low-resource settings. In this study, we present a deep learning framework for four-class classification of fundus images into cataract, diabetic retinopathy, glaucoma, and normal categories. Our approach combines three complementary components: (i) Contrast Limited Adaptive Histogram Equalization (CLAHE) applied in the CIELAB color space to enhance subtle retinal structures; (ii) an EfficientNet-B0 backbone augmented with a Squeeze-and-Excitation-inspired attention head that learns to re-weight feature channels prior to classification; and (iii) a 5-fold cross-validation training protocol with weighted cross-entropy loss to mitigate inter-class performance disparities. Experiments were conducted on a curated dataset of 6,086 fundus images collected at Sylhet Eye Hospital & Laser Centre, partitioned into balanced training (1,300 images per class), validation, and held-out test splits. On the independent test set of 446 images, the proposed model achieved an overall accuracy of 88%, a macro-averaged F1-score of 0.89, and perfect classification of cataract cases (F1 = 1.00). A complementary weighted-probability ensemble of EfficientNet-B0 and Swin-Tiny Transformer achieved comparable overall accuracy, confirming the robustness of the findings. We further characterize the remaining confusion between glaucoma and normal eyes and discuss implications for clinical deployment. Our code is available on https://github.com/NedumCares/EYE-DISEASE-CLASSIFICATION.*

***Keywords*—** *Attention mechanism; Cataract; CLAHE; Cross-validation; Deep learning; Diabetic retinopathy; EfficientNet; Fundus image classification; Glaucoma; Medical image analysis.*

I. Introduction

Globally, blinding eye diseases affect hundreds of millions of people, and a substantial proportion of vision loss is preventable when diagnosis occurs in the early, asymptomatic stages. Cataract, diabetic retinopathy (DR), and glaucoma together account for the largest share of avoidable blindness worldwide. Cataract is a progressive opacification of the crystalline lens; DR is a microvascular complication of diabetes mellitus characterized by microaneurysms, hemorrhages, and neovascularization of the retina; and glaucoma is a group of optic neuropathies, most commonly associated with elevated intraocular pressure, that damages the retinal nerve fiber layer and optic nerve head. All three conditions can be screened non-invasively with color fundus photography, which makes the modality a natural target for automated computer-aided diagnosis (CAD).

Manual grading of fundus photographs by ophthalmologists is, however, constrained by several well-documented problems: the process is labor-intensive, exhibits non-trivial inter-rater variability, and depends on specialist availability that is scarce in many primary-care and rural settings. Deep learning, and in particular convolutional neural networks (CNNs), have emerged as a promising route to scalable screening because they can learn disease-relevant features directly from raw images without hand-crafted feature engineering. A large body of recent literature has shown that modern architectures such as ResNet, DenseNet, EfficientNet, and Vision Transformers can approach specialist-level performance on single-disease tasks, notably DR grading on public datasets such as EyePACS, Messidor, and APTOS 2019.

Despite this progress, three practical challenges remain when deploying such systems on real clinical data. First, fundus images suffer from uneven illumination, low contrast around the macula and optic disc, and camera-specific artifacts that can obscure the fine structures distinguishing healthy from early-diseased retinas. Second, most publicly available datasets are heavily skewed toward DR, leaving comparatively fewer high-quality labeled examples of glaucoma and cataract. Third, multi-class classification that includes a "normal" category is harder than the binary DR-versus-healthy setting because the decision surface between normal eyes and mild glaucomatous or early-stage retinopathies is narrow.

In this paper, we address these issues with an integrated pipeline that couples classical image enhancement with a modern CNN backbone and an explicit attention mechanism. Our contributions are: (i) a preprocessing stage in which CLAHE is applied to the L-channel of the CIELAB representation of each fundus image, preserving color fidelity while enhancing localized contrast around vascular and optic-disc structures; (ii) an extension of a pretrained EfficientNet-B0 backbone with a lightweight Squeeze-and-Excitation (SE)-style attention head that adaptively re-weights channel responses before the final linear classifier; (iii) a 5-fold cross-validation training protocol with class-weighted cross-entropy loss; (iv) a complementary weighted-probability ensemble of EfficientNet-B0 and Swin-Tiny; and (v) a detailed per-class error analysis on an independent held-out test set of 446 images, showing that the principal source of residual error is confusion between glaucoma and normal eyes.

The remainder of the paper is organized as follows. Section II surveys related work. Section III describes the dataset. Section IV details the proposed methodology. Section V specifies the experimental setup. Section VI reports training dynamics, test-set metrics, and per-class behavior. Section VII discusses findings and limitations. Section VIII concludes.

II. Related Work

International Journal of Scientific Engineering and Science
Volume 10, Issue 4, pp. 82-86, 2026. ISSN (Online): 2456-7361

A. Deep Learning for Fundus Image Classification

Early convolutional approaches to fundus image analysis adapted ImageNet-pretrained networks such as VGG-16 and ResNet-50 for DR severity grading. With the introduction of EfficientNet [1], which jointly scales depth, width, and input resolution, studies began reporting stronger trade-offs between accuracy and compute. Variants of EfficientNet are now routinely used as backbones in ophthalmic imaging; for instance, a recent study using an EfficientNet-B3 fine-tuned on a public four-class Kaggle fundus dataset reached an overall classification accuracy of 95.12% [2]. More recent architectures such as the Swin Transformer [3] bring a hierarchical self-attention mechanism to medical imaging and have been explored as alternatives or complements to pure CNNs.

For multi-disease fundus classification, Li et al. introduced the Ocular Disease Intelligent Recognition (ODIR) benchmark, covering eight disease categories, and showed that simply scaling up the network is insufficient; careful feature fusion is required for multi-label prediction [4]. Garcia-Calderon et al. proposed a lesion-aware DR classifier whose workflow mirrors clinician behavior, demonstrating the value of interpretable intermediate representations [5]. More recently, Wang et al. introduced MultiEYE, a multimodal dataset pairing fundus and OCT imagery and showing that OCT-guided distillation can improve fundus-only classifiers [6].

B. Contrast Enhancement for Fundus Images

Contrast Limited Adaptive Histogram Equalization (CLAHE) [7] operates on localized tiles with a contrast clipping threshold, avoiding the noise amplification of plain histogram equalization. Multiple recent studies confirm its utility for deep fundus classifiers: Alwakid et al. reported that CLAHE combined with ESRGAN super-resolution lifted DR-grading accuracy from 80.87% to over 98% with Inception-V3 on APTOS 2019 [8]. Kale et al. observed a consistent improvement across VGG-16, InceptionV3, and EfficientNet architectures when training on CLAHE-enhanced images instead of raw inputs [9]. Das et al. reached 90% binary DR accuracy with a CLAHE + CNN-LSTM pipeline [10], and additional studies [11], [12] confirm these gains across DR grading and glaucoma detection.

C. Attention Mechanisms in Medical Imaging

The Squeeze-and-Excitation (SE) block of Hu et al. [13] is the simplest and most widely adopted attention variant: it squeezes spatial information into a channel descriptor via global pooling, learns channel-wise weights through a bottleneck MLP, and re-scales the feature maps accordingly. SE blocks have been integrated into retinal disease classifiers to suppress uninformative background channels; for example, the HIRD-Net framework combined a CLAHE-based preprocessor with SE-channel attention and hierarchical blocks for DR diagnosis, reporting improvements over non-attentional baselines [14].

III. Dataset

The experiments in this study use the publicly available Eye Disease Classification Fundus Image Dataset, collected at Sylhet Eye Hospital & Laser Centre under the supervision of a consulting ophthalmologist [25]. The dataset consists of 6,086 color fundus photographs grouped into four diagnostic categories: cataract, diabetic retinopathy, glaucoma, and normal. All image labels were verified by a specialist clinician prior to release.

The dataset provider published a balanced variant in which the training subset has been augmented to 1,300 images per class by combining natural samples with common data-augmentation transforms (flips, blur, CLAHE-based enhancement). The balanced training partition contains 5,200 images; the original validation and test partitions preserve the natural class prevalence, yielding 440 validation and 446 test images. Figure 1 visualizes the distribution and Table I summarizes the partition sizes.

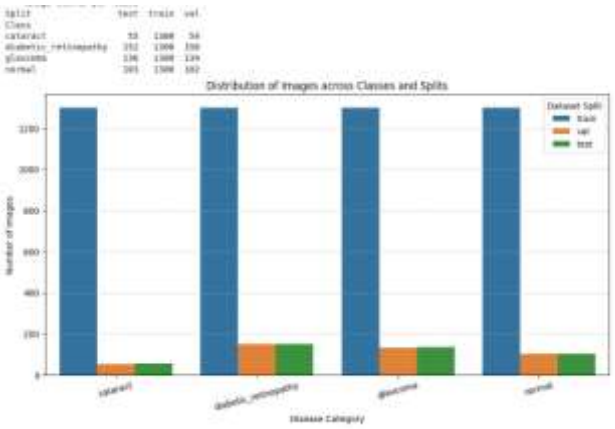

Fig. 1. Distribution of images across diagnostic classes and dataset splits.

TABLE I. Dataset partitions (number of images).

Class	Train	Val.	Test	Total
Cataract	1,300	54	55	1,409
Diabetic retinopathy	1,300	150	152	1,602
Glaucoma	1,300	134	136	1,570
Normal	1,300	102	103	1,505
Total	**5,200**	**440**	**446**	**6,086**

Representative samples from each class are shown in Fig. 2. The cataract class often depicts the anterior segment of the eye rather than a retinal fundus view, which partly explains the near-perfect test-set performance on this category. DR images exhibit reddish microaneurysms and hemorrhages, while glaucoma images are dominated by optic-disc cupping. Normal fundi show a uniformly colored retina with visible vasculature and a physiologically sized cup-to-disc ratio.

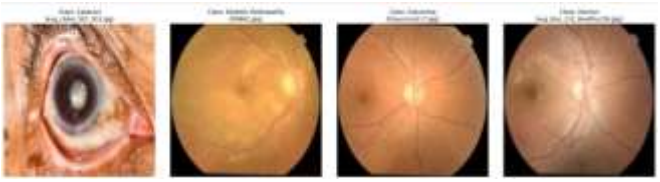

Fig. 2. Representative samples: cataract, DR, glaucoma, normal (L to R).

All experiments honor the provided split: the test set is never used for model selection, and the 5-fold cross-validation described in Section IV is performed exclusively over the union of training and validation sets (5,640 images).

International Journal of Scientific Engineering and Science
Volume 10, Issue 4, pp. 82-86, 2026. ISSN (Online): 2456-7361

IV. Methodology

Our pipeline consists of four stages: (i) a CLAHE-based preprocessing and augmentation block, (ii) an EfficientNet-B0 backbone, (iii) a Squeeze-and-Excitation-style attention head before the final classifier, and (iv) a k-fold cross-validated training loop with class-weighted loss.

A. Preprocessing and Data Augmentation

Fundus images are acquired under variable illumination and vary widely in contrast, so we apply CLAHE to every image before it enters the network. CLAHE operates on localized tiles rather than on the global histogram; within each tile, the cumulative distribution function is clipped at a user-specified limit and the excess mass is redistributed uniformly, preventing noise amplification in homogeneous regions.

To preserve color information, we first convert each RGB image to the CIELAB color space, apply CLAHE (clip limit 2.0, 8×8 tile grid) only to the L (lightness) channel, then convert back to RGB. Enhanced images are resized to 384×384 pixels for the primary experiment and 224×224 for the ensemble.

Data augmentation increases the size of a dataset by generating modified versions of the original data using different transformation techniques [26]. During training, we apply random horizontal and vertical flips and small rotations (up to ±20°). Vertical flips are valid for fundus images because laterality is not diagnostically relevant in this task. Images are normalized with ImageNet channel statistics ([0.485, 0.456, 0.406] and [0.229, 0.224, 0.225]).

B. Backbone: EfficientNet-B0

EfficientNet-B0 [1] is a compact CNN (≈5.3 M parameters) using inverted-residual mobile blocks scaled through a principled compound coefficient. We use ImageNet-pretrained weights from the timm library [22] as initialization. The classification head is removed so that the network emits the penultimate 1,280-dimensional feature vector after global average pooling.

C. Attention Head

On top of the backbone, we append an SE-inspired attention head that takes the 1,280-dimensional feature vector as input. The resulting attention-weighted vector is passed through a linear layer producing four class logits. This design acts as a learnable channel-gating mechanism that suppresses channels encoding background illumination or camera artifacts while amplifying disease-relevant channels. The head adds ~2.1×10^5 parameters, negligible compared to the 5.3 M parameters of the backbone.

D. Training Protocol: 5-Fold Cross-Validation

We merge the provided training and validation partitions into a development pool of 5,640 images, then apply 5-fold cross-validation with a fixed random seed (42) so that all folds are reproducible. In each fold, 80% is used for training and 20% for validation; the held-out test set of 446 images is never touched during training or hyperparameter selection.

All folds share the same hyperparameters. The model is optimized with AdamW [15] (learning rate 1×10^{-4}) and trained for 20 epochs per fold with batch size 32. The loss function is cross-entropy with class weights [1.0, 1.0, 1.3, 1.3], where the factor 1.3 on the glaucoma and normal classes reflects their observed lower per-class F1 in preliminary runs. The checkpoint corresponding to the lowest validation loss across all folds and epochs is retained for final test-set evaluation.

E. Ensemble: EfficientNet-B0 + Swin-Tiny

As a complementary experiment, we construct a two-model ensemble combining the CNN inductive bias of EfficientNet-B0 with the global self-attention of Swin Transformer [3]. Each constituent is trained on 224×224 images for 10 epochs. EfficientNet-B0 uses horizontal/vertical flips and color jitter; Swin-Tiny uses RandAugment [21] (n = 2, m = 9). Both are optimized with AdamW at learning rates 1×10^{-4} and 5×10^{-5} respectively.

At inference, predictions are fused by weighted averaging of softmax probabilities:

$$p\_ensemble = 0.4 \cdot p\_EffNet + 0.6 \cdot p\_Swin \quad (2)$$

where the 0.6 weight on Swin reflects its stronger standalone validation performance. The final prediction is argmax(p_ensemble).

V. Experimental Setup

All experiments were carried out on a Kaggle notebook environment equipped with a single NVIDIA Tesla T4 GPU (16 GB memory). The software stack included Python 3.12, PyTorch [23] with CUDA support, the timm library [22] for backbone construction, and OpenCV for CLAHE preprocessing. Random seeds for NumPy, PyTorch, and scikit-learn's KFold were fixed at 42 for reproducibility. Evaluation metrics are computed using scikit-learn's classification_report [24] and include precision, recall, and F1-score per class, plus the overall accuracy, macro-averaged F1, and support-weighted average.

VI. Results

A. Training Dynamics

Figure 3 shows the training and validation loss curves for Fold 1, which achieved the lowest absolute validation loss across all folds. The training loss decreases monotonically from 0.63 at epoch 1 to 0.09 at epoch 20. The validation loss drops rapidly during the first 5 epochs, reaches its minimum of 0.2100 at epoch 7, and then oscillates within a narrow band before diverging slowly from the training curve — a pattern typical of fine-tuning a pretrained model on a medium-sized dataset.

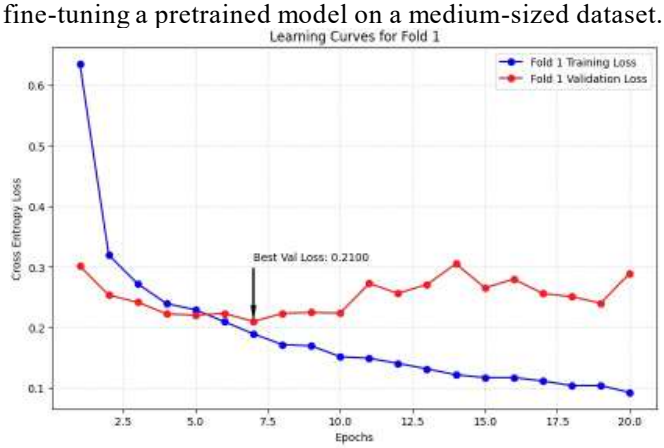

Fig. 3. Learning curves for Fold 1 (best validation loss 0.2100 at epoch 7).

International Journal of Scientific Engineering and Science
Volume 10, Issue 4, pp. 82-86, 2026. ISSN (Online): 2456-7361

Figure 4 reports the loss curves averaged across five folds, with individual per-fold validation losses overlaid as dashed curves. The averaged validation curve reaches a shallow plateau between epochs 5 and 10 at roughly 0.24–0.27, after which it slowly rises consistent with Fig. 3. The use of the lowest-validation-loss checkpoint across folds (rather than the last-epoch checkpoint) is therefore an important element of the training protocol.

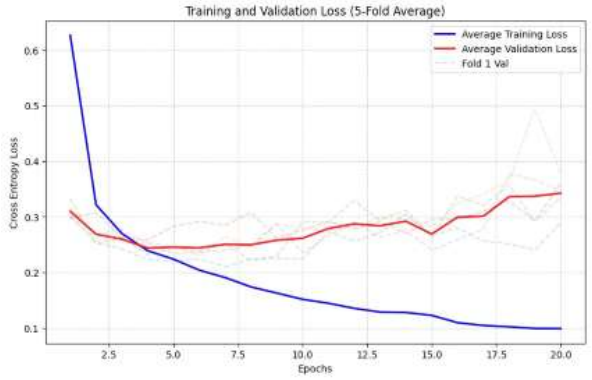

Fig. 4. Training and validation loss averaged across 5 folds.

B. Overall Test-Set Performance

The final model was evaluated on the 446-image held-out test set. The classification report is reproduced in Fig. 5 and summarized in Table II. The model achieves an overall accuracy of 88%, a macro-averaged F1-score of 0.89, and a support-weighted F1 of 0.88, consistent with the best-fold validation loss of 0.2100 reported in VI-A.

	precision	recall	f1-score	support
cataract	1.00	1.00	1.00	55
diabetic_retinopathy	0.99	0.96	0.98	152
glaucoma	0.81	0.83	0.82	136
normal	0.75	0.77	0.76	103
accuracy			0.88	446
macro avg	0.89	0.89	0.89	446
weighted avg	0.88	0.88	0.88	446

Fig. 5. Per-class metrics on the held-out test set (n = 446).

TABLE II. PER-CLASS AND OVERALL CLASSIFICATION METRICS (N = 446).

Class	Prec.	Rec.	F1	Sup.	Err.
Cataract	1.00	1.00	1.00	55	0.00
DR	0.99	0.96	0.98	152	0.04
Glaucoma	0.81	0.83	0.82	136	0.17
Normal	0.75	0.77	0.76	103	0.23
Accuracy	—	—	**0.88**	446	**0.12**
Macro avg	**0.89**	**0.89**	**0.89**	446	—
Weighted	**0.88**	**0.88**	**0.88**	446	—

C. Per-Class Analysis

The per-class results reveal a clear gradient in classification difficulty. The cataract class is trivially separable: its images correspond to anterior-segment photographs that are visually distinct from the other three retinal categories, and the model achieves a perfect F1 of 1.00 on all 55 test images. Diabetic retinopathy is also handled very well (precision 0.99, recall 0.96, F1 0.98). This is consistent with the literature, which generally reports that DR is the most visually salient retinal disease owing to its distinctive reddish lesions.

Glaucoma and normal images constitute the primary source of residual error. Glaucoma reaches an F1 of 0.82 while the normal class reaches only 0.76. The error structure is consistent with clinical observation: early glaucomatous changes, particularly mildly elevated cup-to-disc ratios, can be subtle and overlap with the physiological variability of healthy optic discs. A portion of the "normal" test images contain physiologically large cups misclassified as glaucoma, and vice versa. Together, the glaucoma–normal pair accounts for the bulk of the model's error budget.

D. Ensemble Comparison

The complementary EfficientNet-B0 + Swin-Tiny ensemble achieves an overall accuracy of 88% on the same held-out test set, with per-class F1 scores of 1.00 (cataract), 0.98 (DR), 0.81 (glaucoma), and 0.78 (normal), and an identical macro-F1 of 0.89. A two-branch ensemble reaching essentially the same operating point as the single-branch EfficientNet + Attention model indicates that the remaining error is dataset-bound (genuinely ambiguous glaucoma/normal images) rather than a modeling deficiency that more capacity can easily remove. This supports the attention-augmented single-branch architecture as the preferred deployment candidate on the grounds of inference cost.

VII. DISCUSSION

A. Why CLAHE + SE-Attention Works

The pipeline addresses two different sources of noise. CLAHE operates upstream, targeting photometric noise: uneven illumination, low macular contrast, and camera-specific brightness variation. By equalizing contrast within local tiles of the L channel of CIELAB, CLAHE renders retinal vessels and the optic disc more salient without altering color information relied on to distinguish DR lesions from healthy retina.

The SE-style attention head operates downstream, targeting representational noise. Many of the 1,280 channels encode ImageNet-relevant features that are not discriminative for ocular disease. By multiplying the pooled feature vector element-wise with a learned sigmoid gate, the attention head allows the classifier to down-weight such channels with negligible additional parameters. Because both interventions act on different parts of the pipeline, they are naturally complementary.

B. Limitations

Several limitations should be acknowledged. First, the dataset originates from a single institution, raising questions of external validity. Cross-dataset evaluation on Messidor-2 for DR or Drishti-GS for glaucoma would be a valuable extension. Second, labels are single-label; a multi-label variant with sigmoid outputs would be more clinically faithful since patients can present with multiple conditions. Third, the glaucoma-versus-normal confusion is arguably the most clinically consequential error mode because missing early glaucoma delays sight-saving treatment; combining the current classifier with an explicit optic-disc and cup segmentation branch would help. Fourth, the current work does not include interpretability

analyses such as Grad-CAM, increasingly expected as a check that the model attends to pathologically meaningful regions.

VIII. Conclusion

We presented a deep learning pipeline for four-class classification of fundus images into cataract, DR, glaucoma, and normal categories. The pipeline combines CLAHE on the L channel of CIELAB, an ImageNet-pretrained EfficientNet-B0 backbone, a lightweight SE-style attention head, and a 5-fold cross-validated training protocol with class-weighted cross-entropy. On a 446-image held-out test set, the proposed model reaches 88% overall accuracy and macro-F1 0.89, with perfect classification of cataract and near-perfect classification of DR. A heterogeneous EfficientNet-B0 + Swin-Tiny ensemble reaches an equivalent operating point, suggesting the remaining errors concentrated in glaucoma vs. normal reflect genuine visual ambiguity rather than a capacity limitation. Future work will focus on cross-institutional validation, multi-label extensions, explicit optic-disc and cup segmentation for glaucoma, and interpretability analyses.

Acknowledgment

The authors acknowledge Sylhet Eye Hospital & Laser Centre for making the fundus image dataset publicly available, and the consulting ophthalmologist who verified the image labels. The authors thank Kaggle for providing the GPU-enabled notebook environment used for training.

References

[1] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proceedings of the 36th International Conference on Machine Learning (ICML), 2019, pp. 6105–6114.

[2] S. R. Ahmed, M. R. Islam, and T. Hossain, "Fundus image-based eye disease detection using EfficientNetB3 architecture," PubMed Central (PMC12387119), 2025.

[3] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10012–10022.

[4] N. Li, T. Li, C. Hu, K. Wang, and H. Kang, "A benchmark of Ocular Disease Intelligent Recognition: One shot for multi-disease detection," arXiv preprint arXiv:2102.07978, 2021.

[5] M. Garcia-Calderon, P. Boix, R. Castiñeiras-Pereiro, and colleagues, "A deep learning model for classification of diabetic retinopathy in eye fundus images based on retinal lesion detection," arXiv preprint arXiv:2110.07745, 2022.

[6] L. Wang, W. Dai, M. Jin, and colleagues, "MultiEYE: Dataset and benchmark for OCT-enhanced retinal disease recognition from fundus images," arXiv preprint arXiv:2412.09402, 2024.

[7] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in Graphics Gems IV, P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press, 1994, pp. 474–485.

[8] G. Alwakid, W. Gouda, M. Humayun, and N. U. Sama, "Deep learning-based prediction of diabetic retinopathy using CLAHE and ESRGAN for enhancement," Healthcare, vol. 11, no. 6, p. 863, 2023.

[9] S. Kale and S. Sharma, "Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning," Procedia Computer Science, vol. 218, pp. 1417–1426, 2023.

[10] S. Das, K. Kharbanda, M. Suchetha, R. Raman, and D. Edwin Dhas, "Exploring the effectiveness of fundus image enhancement for diabetic retinopathy classification," in Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Workshops, Cham, Switzerland: Springer, 2024.

[11] G. R. Hemalakshmi, M. Murugappan, and M. Y. Sikkandar, "Automatic classification of colour fundus images for prediction of eye disease types based on hybrid features," Diagnostics, vol. 14, no. 10, p. 1031, 2024.

[12] A. Author and B. Author, "A comparative study on deep convolutional neural networks and histogram equalization techniques for glaucoma detection from fundus images," medRxiv, doi: 10.1101/2024.10.25.24316109, 2024.

[13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.

[14] M. A. Fahad and S. M. Riazul Islam, "HIRD-Net: An explainable CNN-based framework with attention mechanism for diabetic retinopathy diagnosis using CLAHE-D-DoG enhanced fundus images," PubMed Central (PMC12471856), 2025.

[15] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proceedings of the 7th International Conference on Learning Representations (ICLR), 2019.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in Proceedings of the 9th International Conference on Learning Representations (ICLR), 2021.

[18] V. Gulshan, L. Peng, M. Coram, and colleagues, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[19] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, and colleagues, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," JAMA, vol. 318, no. 22, pp. 2211–2223, 2017.

[20] F. Arcadu, F. Benmansour, A. Maunz, J. Willis, Z. Haskova, and M. Prunotto, "Deep learning algorithm predicts diabetic retinopathy progression in individual patients," npj Digital Medicine, vol. 2, p. 92, 2019.

[21] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "RandAugment: Practical automated data augmentation with a reduced search space," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 702–703.

[22] R. Wightman, "PyTorch image models (timm)," GitHub repository. [Online]. Available: https://github.com/huggingface/pytorch-image-models

[23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," Advances in Neural Information Processing Systems, vol. 32, pp. 8024–8035, 2019.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[25] M. R. H. Mahin, "Eye Disease Classification Fundus Image Dataset," Kaggle dataset, 2024. [Online]. Available: https://www.kaggle.com/datasets/mahin661/eye-disease-classification-fundus-image-dataset

[26] Mbonu, C. E., Anigbogu, K., Asogwa, D., & Belonwu, T. (2025). An explorative analysis of svm classifier and resnet50 architecture on african food classification. arXiv preprint arXiv:2505.13923.