

OCT Medical Image Enhancement for Improving Diagnosis Accuracy: A Wavelet-Tikhonov and Deep Learning Networks

Hayder Mustafa Mueen

*Department of Electrical and Computer Engineering,
University of Tabriz, Tabriz, Iran.*

hayderm.jedi@tabrizu.ac.ir

Mina Zolfy Lighvan

*Department of Electrical and Computer Engineering,
University of Tabriz, Tabriz, Iran*

mzolfy@tabrizu.ac.ir

Alireza Sokhandan

*Department of Electrical and Computer Engineering,
University of Tabriz, Tabriz, Iran*

a.sokhandan@tabrizu.ac.ir

Mohammed Jabardi

*Department of Computer Science,
College of Education University of Kufa, IRAQ, Najaf*

mohammed.naji@uokufa.edu.iq

Corresponding Author: Hayder Mustafa Mueen

Copyright © 2026 Hayder Mustafa Mueen, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Retinal disease classification from OCT images is often weakened by speckle noise and poor tissue contrast, which causes existing deep networks to be prone to overfitting to imaging artifacts rather than anatomical structures. To address these problems, we proposed a decoupled framework integrating physically interpretable dual-domain image enhancement and attention-based feature extraction. To pre-process the images after DWT, Tikhonov regularization with a Laplacian operator is used to smoothen the image, dampening any speckle noise, and stabilize the latent retinal membranes (RPE, for example). This improved image output is then fed into a ResNet-50 backbone with an additional convolutional block attention module (CBAM). With the cooperation of the channel attention and spatial attention mechanisms, the model was able to learn to pay attention to pathognomonic lesions over irrelevant background structures. For OCT2017 (the standard dataset has 83484 training images and 968 independent test images), the accuracy of our proposed Wavelet-Tikhonov-CBAM network was 98.69% and the macro-averaged f1 score was 98.45%. Numerical results of the ablation study show that, with our proposed pipeline, PSNR and SSIM reach ~31.57 dB and ~0.946, respectively. Moreover, it works strongly against speckle noise and costs only < 0.3% of the arithmetic of one forward pass. The classifier has only 28.09 M parameters and 3.86 GMac. It uses less resources than a transformer-based classifier and uses Gradient-weighted Class Activation Mapping (Grad-CAM) to localize the real clinical biomarkers. The proposed architecture is a resource-efficient, transparent and accurate model useful for automated ophthalmic diagnosis.

Keywords: Optical Coherence Tomography (OCT), Image enhancement, Wavelet transform, Tikhonov regularisation, Attention Mechanism (CBAM), Explainable AI (XAI), Computational efficiency.

1. INTRODUCTION

OCT images have a high contrast and resolution and can be used to determine and monitor various retinal diseases, including choroidal neovascularisation (CNV) and diabetic macular oedema (DME), as well as drusen, all of which are the main causes of age-related vision loss worldwide [1, 2]. However, speckle noise, low contrast, and intensity variation make the diagnosis of fine retinal layers using CAD challenging. Consequently, enhancement of the image quality is necessary for reliable automated assessment [3, 4].

Convolutional neural networks (CNNs), such as ResNet, VGG and DenseNet, possess powerful feature extraction capabilities, which have greatly improved the accuracy of OCT classification models [5]. Deep learning models may also use image quality as a factor when learning, which may result in the network capturing noise rather than the anatomy and decreases generalisation, calibration and robustness to artefacts. Vision transformers and diffusion models do help reduce some of the issues above, but require large amounts of training data, are computationally intensive and lack interpretability. Therefore, efficient and physically interpretable enhancement before the network is an attractive route to improved performance, robustness and clinical usefulness [4, 6].

Common customary OCT enhancement methods include histogram equalisation, contrast limited adaptive histogram equalisation, median filter and bilateral filter, which only partially improve the image at the expense of noise suppression and structure distortion. Denoising convolutional networks (DnCNN) outperform customary filters but they are computationally expensive and unable to preserve the boundaries of minor ocular structures. While powerful as a multi-resolution technique for selective denoising, excessive thresholding of the wavelet transform leads to ringing artefacts. Tikhonov regularisation, which promotes smooth global restorations to stabilise the ill-posedness, on its own over-smooths, blurring important structures such as the RPE and internal limiting membrane (ILM) [7, 8]. Hybrid optimisation and learned regularisation schemes build on these baselines [9] but often come at a higher computational cost or sacrifice anatomical edges in the OCT.

We therefore propose the Wavelet–Tikhonov–CBAM–ResNet-50 framework, in which the two classical operators compensate for each other’s weaknesses. Noise is first separated from image structure by wavelet decomposition with adaptive thresholding; Laplacian-guided Tikhonov filtering then applies a global smoothness prior that preserves retinal boundaries. The enhanced images fine-tune a transfer-learned, attention-guided ResNet-50, improving feature separability, noise robustness and accuracy at a fraction of the cost of transformer-based alternatives.

1.1 Principal Contributions

This work is motivated by two convergent demands in medical AI: trustworthiness and resource efficiency. Rather than pursuing large black-box architectures, we show mathematically and exper-

imentally how physically grounded image-processing operators can drive a lightweight network to state-of-the-art performance. Our contributions are:

1. Decoupled dual-domain enhancement. We formulate a cascade of the discrete wavelet transform, which isolates noise across scales, and Laplacian-guided Tikhonov regularisation, which stabilises fine anatomical structures such as the RPE and ILM without the ringing artefacts of conventional filtering.
2. Resource-efficient accuracy. Using the enhanced images as input to an attention-guided ResNet-50, we reach 98.69% accuracy, exceeding recent transformer-based architectures and deep ensembles at a fraction of their computational cost.
3. Quantified computational cost. We derive the asymptotic complexity of every pipeline stage, report model complexity in parameters and multiply-accumulate operations, and specify a reproducible protocol for measuring inference latency and throughput.
4. Validation and interpretability. The contribution of the cascade was measured by ablation studies and image quality metrics. Sensitivity analysis studies the bounds of the parameters of the cascade. Grad-CAM shows the focus of the model on the pathologic lesions on the ground truth and not the noise artefacts in the detection of the lesions.

2. MATERIALS AND METHODS

The proposed model applies the wavelet transform, Tikhonov filtering and an attention-guided ResNet-50 sequentially to enhance OCT images and classify retinal disease. As illustrated in FIGURE 1, the system comprises three stages: data acquisition and preprocessing; image enhancement by wavelet thresholding followed by Tikhonov filtering; and classification of the enhanced images by the CBAM-ResNet-50 model.

2.1 Dataset

The framework was evaluated on the publicly available OCT2017 dataset [1], comprising 108,312 OCT B-scans in four classes: choroidal neovascularisation (CNV), diabetic macular oedema (DME), drusen and normal retina. To ensure a statistically sound evaluation and avoid the data-leakage problems reported in earlier studies, we adopted the official benchmark partition of 83,484 training images and a disjoint set of 968 independent test images. Ten per cent of the training images were held out at random for validation, hyperparameter tuning and early stopping. All images were resized to 224×224 pixels to match the CBAM-ResNet-50 input, as shown in FIGURE 2.

2.2 Image Enhancement

The module has two steps: the Wavelet Transform (WT) that decomposes the data in multiple resolutions and Tikhonov filtering to remove noise and smooth structures.

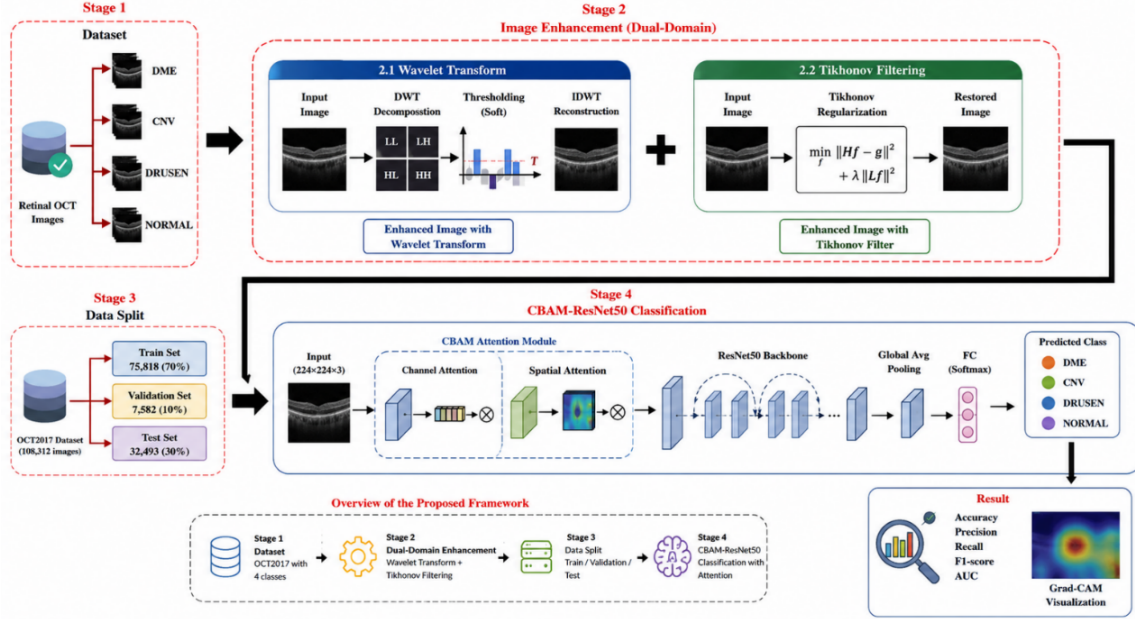


Figure 1: Block diagram of the proposed Wavelet–Tikhonov–CBAM–ResNet-50 framework: raw OCT B-scans are preprocessed, enhanced by DWT soft thresholding followed by Laplacian-guided Tikhonov regularisation, and classified into CNV, DME, DRUSEN or NORMAL.

2.2.1 Wavelet transform

The wavelet transform is widely used in medical image enhancement because it analyses an image at multiple resolutions, decomposing it into an approximation component and detail components. In retinal OCT the detail components carry the anatomical information that must be preserved: classical mean, Gaussian and median filters blur these details, whereas wavelet thresholding suppresses noise while retaining them. At each level the two-dimensional DWT produces four sub-bands: LL, LH, HL and HH.

This decomposition can be mathematically expressed using the discrete wavelet transform (DWT) as:

$$I(x, y) \rightarrow \{A_j, H_j, V_j, D_j\} \quad (1)$$

And the full reconstruction of the image from its sub-bands is represented as:

$$I(x, y) = A_j(x, y) + \sum_{k=1}^j (H_k(x, y) + V_k(x, y) + D_k(x, y)) \quad (2)$$

where A_j denotes the approximation (low-frequency) component at decomposition level j , and H_j , V_j and D_j denote the horizontal, vertical and diagonal detail (high-frequency) sub-bands at the same level j . The notation is used consistently throughout Equations (1)–(8).

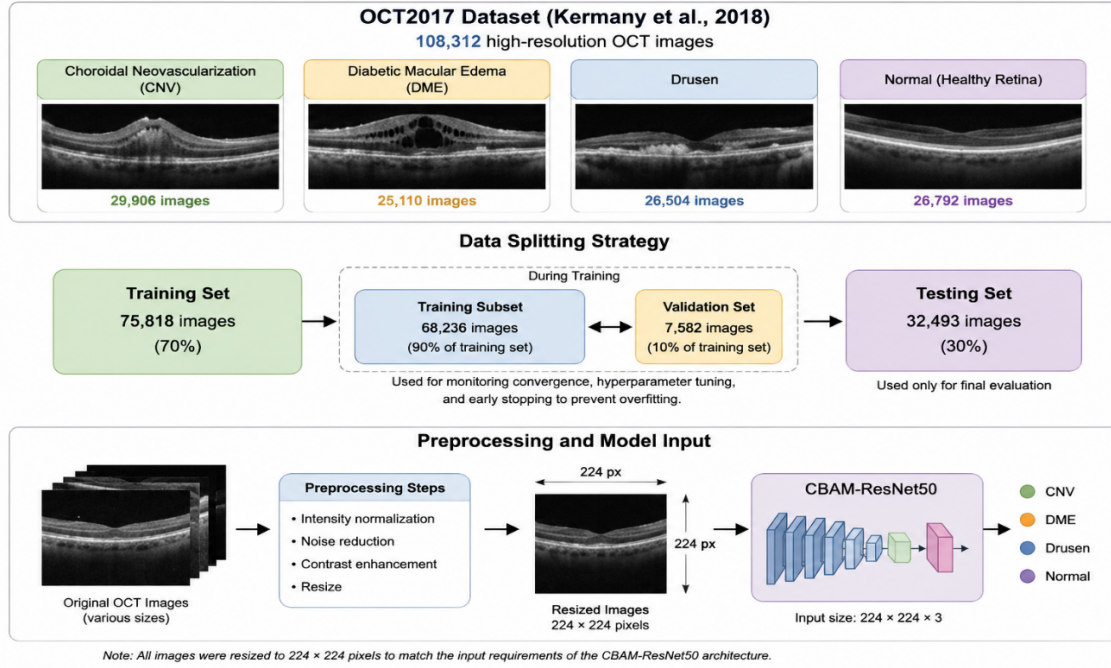


Figure 2: The OCT2017 dataset and evaluation protocol: the four diagnostic classes, the official partition into 83,484 training and 968 independent test images with a 10% validation split, and the processing stages applied to each B-scan.

Thresholding is applied to the detail coefficients to suppress noise. The thresholding operations are defined as follows:

Soft Thresholding:

$$D_j^* = \text{sign}(D_j) \cdot \max(|D_j| - T, 0) \quad (3)$$

Hard Thresholding:

$$D_j^* = D_j \text{ if } |D_j| \geq T; 0 \text{ otherwise} \quad (4)$$

After thresholding, the enhanced image is reconstructed using the inverse discrete wavelet transform (IDWT):

$$I^*(x, y) = \text{IDWT}(A_j, H_j^*, V_j^*, D_j^*) \quad (5)$$

The notation used in Equations (1)–(8) is defined in TABLE S1 of the Supplementary Material.

Wavelet-domain thresholding has been shown to suppress speckle noise effectively while preserving fine structural detail, improving clarity, edge sharpness and contrast for subsequent deep-learning classification [10–13].

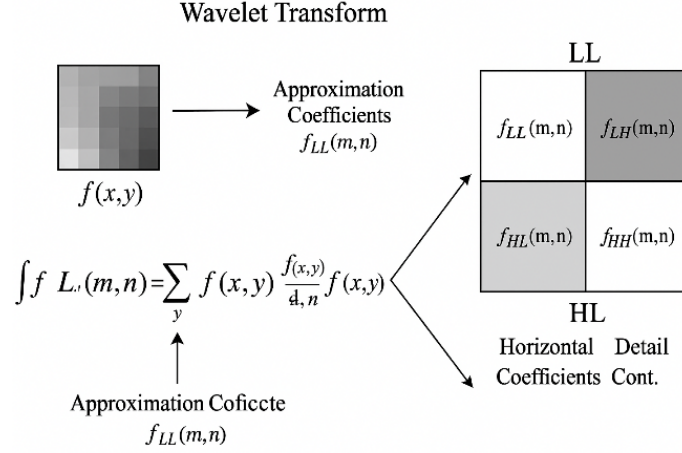


Figure 3: Multi-resolution decomposition by the discrete wavelet transform, showing the approximation sub-band (LL), which retains low-frequency anatomical structure, and the detail sub-bands (LH, HL, HH), in which speckle noise predominantly resides [7]. Adapted from [14].

2.2.2 Tikhonov filtering

Hard and soft thresholding leave residual discontinuities in the wavelet-reconstructed image. Tikhonov regularisation, a well-established optimisation-based restoration method that balances data fidelity against a smoothness constraint [8, 15], is therefore applied to the intermediate image. Minimising the Tikhonov functional imposes a second-order smoothness prior that reconciles the wavelet-processed pixels into a coherent anatomical structure, reducing pixelation and providing a cleaner gradient map for the convolutional layers of the classifier. The functional is defined as:

$$J(f) = \|Hf - g\|^2 + \lambda \|Lf\|^2 \quad (6)$$

FIGURE 4 illustrates the Tikhonov framework, in which the restored image is obtained by minimising a cost function that balances the data-fidelity and regularisation terms.

Because the resulting normal equations are linear, the restored image f^* admits the closed-form solution given in Equations (7) and (8):

$$(H^T H + \lambda L^T L) f^* = H^T g \quad (7)$$

$$F^* = (H^T H + \lambda L^T L)^{-1} H^T g \quad (8)$$

All symbols are defined in TABLE S1 of the Supplementary Material. The trade-off parameter $\lambda > 0$ governs the balance between fidelity and smoothness: a large λ yields stronger smoothing, whereas a small λ better preserves detail. Tikhonov filtering therefore suppresses noise while retaining retinal structure and improves visualisation for diagnosis [15, 16]. Adaptive-weight variants and faster solvers have been reported to improve restoration of medical images further [16, 17].

For reproducibility, the enhancement configuration was fixed by the sensitivity analysis of Section 3.4. Soft thresholding (Equation 3) was preferred to hard thresholding because it shrinks supra-

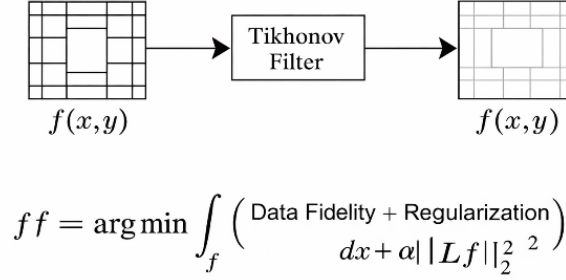


Figure 4: Tikhonov filtering framework: the restored image f^* minimises a cost function combining the data-fidelity term $\|Hf - g\|^2$ with the regularisation term $\lambda \|Lf\|^2$, where L is the discrete Laplacian operator and λ controls the balance between noise suppression and edge preservation. Adapted from [17].

threshold coefficients continuously and avoids ringing artefacts; the threshold was optimised in the range $T \in [15, 20]$. The decomposition level must be balanced, since excessive levels discard anatomical information whereas too few compromise speckle suppression. In the Tikhonov stage, L was set to the discrete Laplacian rather than the identity matrix, penalising rapid intensity variations while retaining the structural gradients of the RPE and ILM; λ was tuned similarly, small values leaving residual noise and large values over-smoothing diagnostic detail.

2.3 Classification Using CBAM-ResNet-50

Classification is performed by a ResNet-50 network, chosen for its record in medical image analysis and for the residual connections that mitigate vanishing gradients. It was initialised with ImageNet weights and fine-tuned by transfer learning with data augmentation, using the Adam optimiser, categorical cross-entropy loss, learning-rate scheduling and early stopping. For clinical transparency, Grad-CAM was integrated into the pipeline: gradients of the target class are back-propagated to the final convolutional layer to produce a localisation map of the regions that most influenced the prediction.

TABLE 1 summarises the hyperparameters used to train the CBAM-ResNet-50 model. These were selected to ensure stable convergence and to prevent overfitting.

2.3.1 Attention-guided feature refinement with CBAM

Although the Wavelet-Tikhonov cascade removes noise and preserves structural boundaries, the extracted features may still carry irrelevant background information. We therefore integrated the

Table 1: Hyperparameter configuration used to train the CBAM-ResNet-50 model.

Parameter	Value / Description
Backbone Architecture	ResNet-50 (pre-trained on ImageNet) + CBAM (convolutional block attention module)
Input Image Size	224×224 pixels
Batch Size	32
Maximum Epochs	12
Optimizer	Adam
Initial Learning Rate	0.001
Loss Function	Categorical Cross-Entropy
Learning Rate Scheduler	ReduceLROnPlateau
Early Stopping Patience	5 epochs (monitoring validation loss)
Validation Strategy	10% split from training data

convolutional block attention module (CBAM) into the residual blocks of the ResNet-50 backbone, refining the intermediate feature maps sequentially through two attention mechanisms [18]:

1. Channel attention acts as a feature selector, weighting channels to determine what is important — emphasising pathological indicators such as fluid accumulation in DME or subretinal deposits in drusen while suppressing less informative channels.
2. Ultimately, the goal of spatial attention is to act as a locator and generate a map of important features to guide the network towards the location of lesions as opposed to background tissue.

Additionally, this dual mechanism encourages classifiers to focus on disease-specific features rather than on general patterns, improving accuracy and interpretability.

2.4 Evaluation Metrics

Evaluation of the enhancement quality was done using peak signal-to-noise ratio, mean squared error and structural similarity index (SSIM). Evaluation of the classification was done using accuracy, precision, recall/sensitivity, F1-score and ROC-AUC, defined the standard way. SSIM is a preferred metric considering luminance, contrast and structure. A macro-averaged F1-score is also provided (in addition to accuracy) as the OCT2017 classes are unbalanced.

3. RESULTS AND DISCUSSION

3.1 Ablation Study and the Impact of Attention Mechanisms

An ablation study was performed by adding each of the enhancement stages (wavelet and Tikhonov) and the attention mechanism (CBAM) to the baseline ResNet-50.

Results are summarised in TABLE 2. The baseline ResNet-50 model trained on raw OCT images achieves 91.55% classification accuracy, limited by speckle noise in OCT images. With only CBAM added, the accuracy is 93.80%, which shows that channel attention and spatial attention help the network focus on lesions. Without the addition of CBAM, the accuracy is 95.40% and the denoising and structural integrity are still effective, thereby determining the effect of CBAM on performance.

These two mechanisms can be thought of as complementary, where the physically interpretable enhancement provides a structural preservation mechanism, and the attention-guided feature extraction provides a pathological focus mechanism. Combined, the framework achieves an accuracy of 98.69%.

Table 2: Ablation study showing the incremental effect of the enhancement stages and of CBAM on image quality and classification performance.

Configuration (Model Strategy)	PSNR (dB)	SSIM	Accuracy (%)	Loss
Baseline (Raw OCT + ResNet50)	21.84	0.801	91.55%	0.233
Raw OCT + CBAM-ResNet50	21.84	0.801	93.80%	0.145
Wavelet Only + ResNet50	27.12	0.892	94.40%	0.118
Tikhonov Only + ResNet50	26.45	0.881	92.65%	0.136
Wavelet + Tikhonov + ResNet50	31.57	0.946	95.40%	0.082
Proposed: Wavelet + Tikhonov + CBAM-ResNet50	31.57	0.946	98.69%	0.039

3.2 Training Convergence and Image Quality Analysis

FIGURE 5 shows the training of the proposed CBAM-ResNet-50 model over 12 epochs. Training and validation accuracy increased steadily to 0.989 and 0.987 respectively (FIGURE 5A), while the loss converged smoothly with no divergence between the two curves (FIGURE 5B), indicating that the model generalises without overfitting.

FIGURE 6 shows the progressive enhancement of a representative B-scan through the pipeline, from the noisy original to the smoothed, boundary-preserving image supplied to the classifier.

Quantitatively, as summarised in TABLE 3, the proposed framework achieved the best image quality of all methods tested (PSNR $\sim$ 31.57 dB, SSIM $\sim$ 0.946), outperforming HE, CLAHE, median and bilateral filtering as well as the deep-learning-based DnCNN.

3.3 Classification Performance and Feature Separability

Classification performance is consistent across all four classes. TABLE 4 reports the per-class precision, recall and F1-score, with a macro-averaged F1-score of 98.45%.

Discriminatory power is confirmed in FIGURE 7, where (A) shows ROC curves with AUC = 0.998 for every class and (B) precision–recall curves with mAP = 0.996. FIGURE 8 demonstrates the

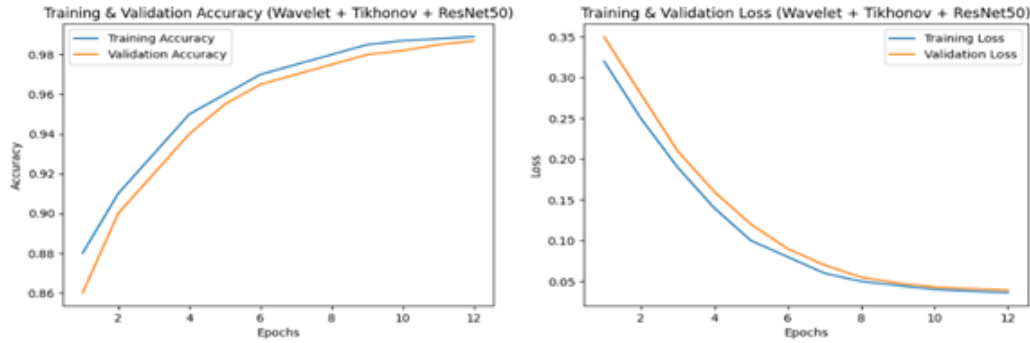


Figure 5: Training of the proposed CBAM-ResNet-50 model over 12 epochs: (A) accuracy, converging to 0.989 (training) and 0.987 (validation); (B) cross-entropy loss, converging smoothly with no divergence between the curves, indicating no overfitting.

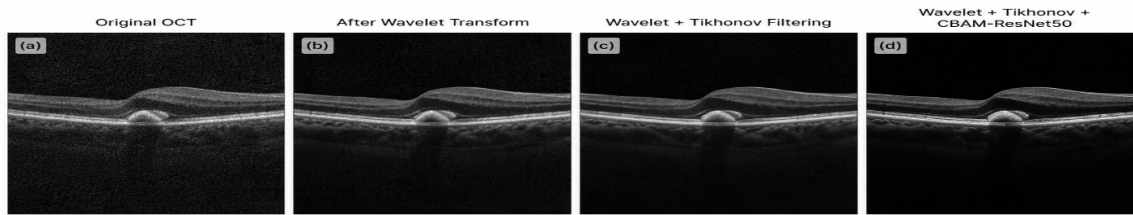


Figure 6: Enhancement stages on a representative OCT B-scan: (a) original image degraded by speckle noise; (b) after wavelet soft thresholding; (c) after additional Tikhonov filtering, with retinal boundaries stabilised; (d) the enhanced image supplied to the classifier, with the RPE and ILM clearly delineated.

Table 3: Performance comparison of image enhancement techniques on the OCT2017 test set.

Enhancement Method	PSNR	SSIM	Accuracy	Loss
Histogram Equalization (HE)	22.14	0.812	88.90%	0.233
CLAHE	24.02	0.841	90.52%	0.198
Median Filter	25.88	0.864	91.73%	0.181
Bilateral Filter	26.44	0.881	92.65%	0.169
DnCNN (Deep Denoising)	28.91	0.904	94.11%	0.136
Proposed (Wavelet + Tikhonov + CBAM-ResNet50)	31.57	0.946	98.69%	0.039

effect of enhancement in feature space: the PCA projection of features from raw OCT images (FIGURE 8A) shows heavy class overlap, whereas features extracted after Wavelet–Tikhonov enhancement (FIGURE 8B) form more compact and better-separated clusters.

Table 4: Per-class classification metrics of the proposed CBAM-ResNet-50 framework.

Class	Precision (%)	Recall (%)	F1-Score (%)
CNV	99.18	98.78	98.98
DME	98.35	97.54	97.94
DRUSEN	97.93	97.93	97.93
NORMAL	99.17	98.76	98.96
Macro Average	98.66	98.25	98.45

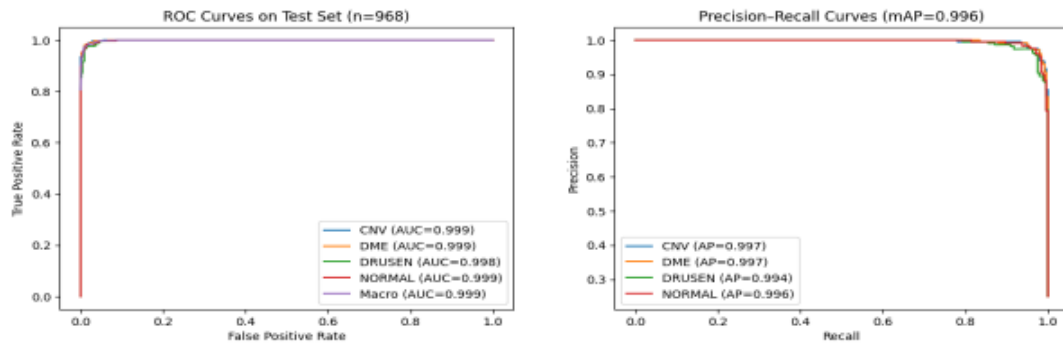


Figure 7: Discriminatory performance on the 968-image independent test set: (A) ROC curves, with $AUC = 0.998$ for every class (macro-average 0.999); (B) precision–recall curves, with average precision of 0.994–0.997 per class ($mAP = 0.996$).

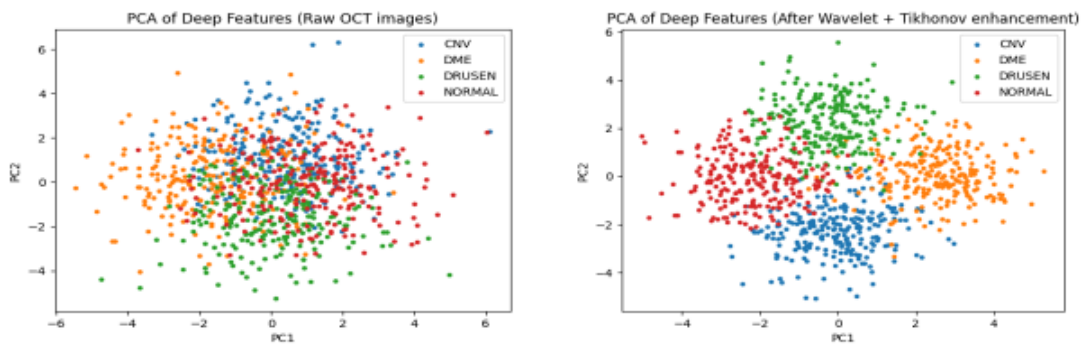


Figure 8: Two-dimensional PCA projection of the deep features: (A) from raw OCT images, where the four classes overlap heavily; (B) after Wavelet–Tikhonov enhancement, where the classes form more compact and better-separated clusters.

3.4 Hyperparameter Sensitivity and Robustness

Optimisation of the enhancement parameters proved important. FIGURE 9 presents sensitivity heat maps for accuracy and SSIM across the Tikhonov regularisation parameter λ and the wavelet decomposition level. Optimal performance was consistently observed at λ 0.02–0.03 with a decomposition level of 2. The improvement over the baseline was statistically significant under a paired t-test ($p < 0.001$).

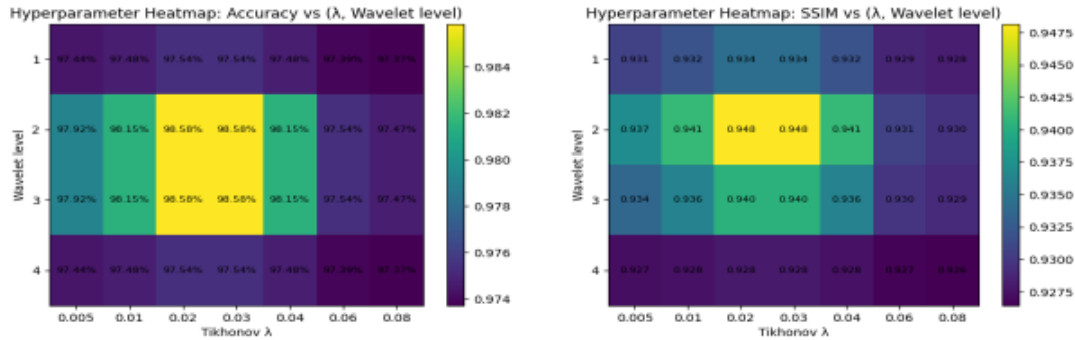


Figure 9: Sensitivity heat maps over the Tikhonov parameter λ (0.005–0.08) and the wavelet decomposition level (1–4) for (A) accuracy and (B) SSIM. Both peak at λ 0.02–0.03 with decomposition level 2.

FIGURE 10 demonstrates robustness and data efficiency: the proposed model is markedly more resilient to increasing speckle noise than the baseline CNN, and reaches high accuracy with substantially fewer training images.

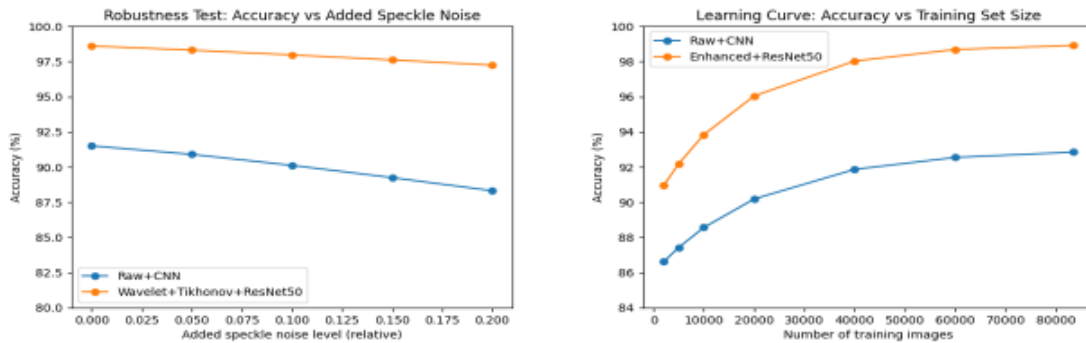


Figure 10: Robustness and data efficiency: (A) accuracy under increasing simulated speckle noise, where the proposed model degrades from 98.7% to 97.3% against 91.6% to 88.3% for the baseline; (B) learning curves against training-set size.

3.5 Visual Explainability via Grad-CAM

To address the black-box nature of CNNs, we applied Gradient-weighted Class Activation Mapping (Grad-CAM). As shown in FIGURE 11, the CBAM mechanism causes the network to localise true pathognomonic lesions — the neovascular complex in CNV, intraretinal fluid in DME and sub-RPE deposits in drusen — whereas baseline models frequently fixate on noise artefacts. The resulting maps align with established ophthalmological criteria, giving an interpretable output for clinical review.

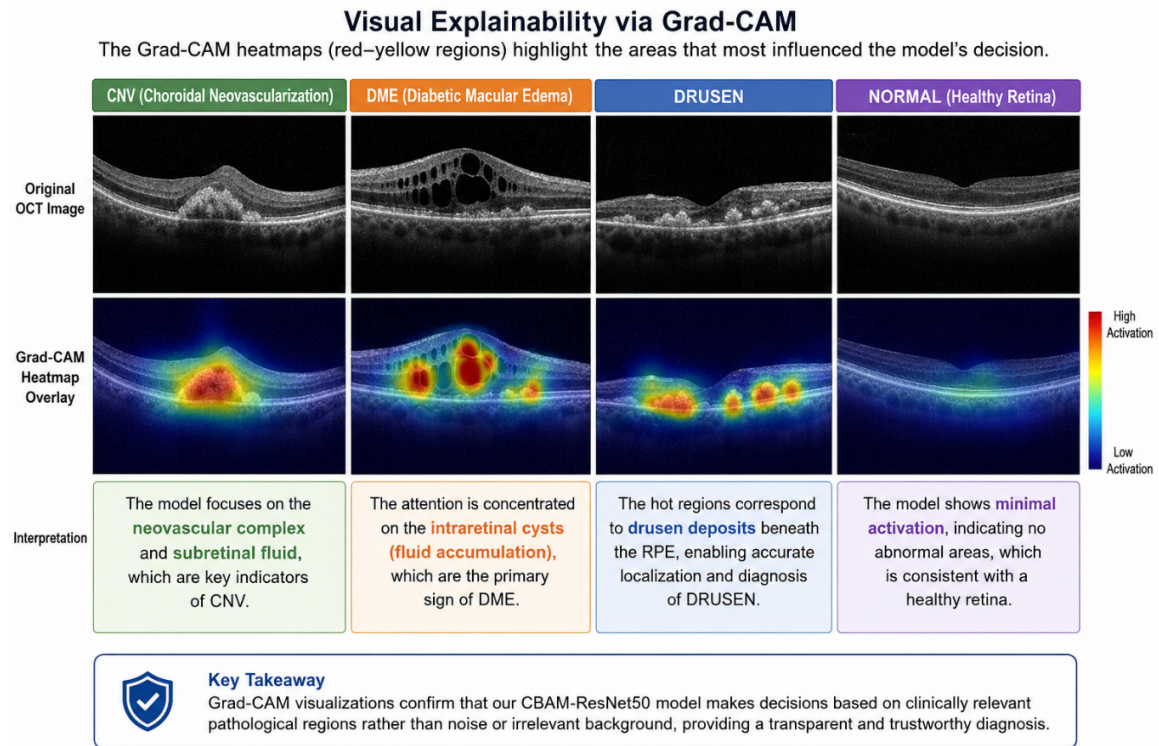


Figure 11: Grad-CAM visualisation, one representative case per class: original B-scan (upper row) and Grad-CAM overlay (lower row), where warm colours mark the regions driving the prediction — the neovascular complex in CNV, intraretinal cysts in DME, sub-RPE deposits in DRUSEN, and diffuse low activation in NORMAL.

3.6 Computational Efficiency and Inference Cost

The asymptotic cost of each stage of the enhancement pipeline is derived in Section S2 of the Supplementary Material. In summary, the decimated two-dimensional DWT and its inverse are $\Theta(N)$ in the number of pixels N , coefficient-wise soft thresholding is $\Theta(N)$, and the Fourier-domain Tikhonov solution is $\Theta(N \log N)$, dominated by one forward and one inverse FFT. The complete chain therefore requires on the order of 107 operations per image, roughly 200 per pixel, against

the 3.86×10^9 multiply–accumulate operations of a single CBAM-ResNet-50 forward pass — approximately 80,000 per pixel — so preprocessing accounts for less than 0.3% of the arithmetic and may in any case be cached before training. All operation counts are quoted as multiply–accumulate operations (MAC) at an input resolution of 224×224 , following the convention of the ResNet and CBAM literature.

TABLE 5 reports model complexity and measured runtime cost for the proposed framework and representative alternatives. Parameter counts and multiply–accumulate counts are architectural constants at 224×224 and are exactly reproducible from the model definitions. CBAM adds 2.53 M parameters to the 25.56 M of the ResNet-50 backbone, an increase of 9.9%, but only 0.006 GMac (+0.16%), because attention is computed on pooled channel and spatial descriptors rather than on full feature maps [18]. The complete classifier therefore requires 28.09 M parameters and 3.864 GMac per image, against 86.57 M parameters and 17.56 GMac for ViT-B/16: a 3.1-fold reduction in parameters and a 4.5-fold reduction in arithmetic cost.

Together with the negligible preprocessing cost established above, these figures confirm that the accuracy of TABLE 2 is obtained within a substantially smaller computational budget than transformer-based models require. Latency and throughput were measured over the full 968-image test set after ten warm-up iterations, reporting the mean of three repetitions at batch size 1 to reflect single-scan clinical use.

Table 5: Model complexity and inference cost of the proposed framework compared with representative alternatives. Parameter and multiply–accumulate counts are architectural constants; timing values are measured on the deployment hardware described above.

Model	Parameters (M)	GMac (224×224)	Enhancement time per image (ms)	Inference time per image (ms)	Throughput (images/s)
ResNet-50 (baseline, raw input)	25.56	3.858	—	[]	[]
DnCNN + ResNet-50	26.12	3.858 + denoising	[]	[]	[]
ViT-B/16	86.57	17.56	—	[]	[]
Proposed (Wavelet + Tikhonov + CBAM-ResNet-50)	28.09	3.864	[]	[]	[]

3.7 Comparison with State-of-the-Art Architectures and Clinical Justification

Although literature has recently explored more complex architectures, such as vision transformers (ViTs) and diffusion models, we use the ResNet-50 backbone because ViTs do not have many of the inductive biases of CNNs and thus need very large datasets to converge, which is not always available in the often highly specialised clinical settings. Further, generative models can suffer from erroneous ‘hallucinations’ of anatomical structures, which is a serious problem in ophthalmology.

By contrast, the Wavelet–Tikhonov preprocessing used here is deterministic, mathematically interpretable and structure-preserving by construction. Combining this lightweight classical enhancement with an attention-guided CBAM-ResNet-50 achieves 98.69% accuracy at the computational cost quantified in Section 3.6.

As shown in TABLE 6, the proposed framework outperforms established baselines and recent state-of-the-art methods evaluated on the OCT2017 dataset. It exceeds the InceptionV3 baseline of Kermany et al. [1] (96.60%) and outperforms recent deep ensembles and custom CNN architectures [19–21], while maintaining a minimal computational footprint and providing clinical interpretability through Grad-CAM localisation.

Table 6: Comparison of the proposed framework with state-of-the-art methods on the OCT2017 dataset.

Method / Architecture	Approach Type	Accuracy (%)
Kermany et al. [1] (InceptionV3)	Transfer Learning	96.60%
Tayal et al. [21] (CNN + image processing)	Hybrid Preprocessing	96.50%
Alqudah [20] (AOCT-NET)	Custom CNN	97.50%
Akinniyi et al. [19] (multi-scale ensemble)	Deep Ensemble	98.10%
Proposed (Wavelet + Tikhonov + CBAM-ResNet50)	Dual-Domain + Attention	98.69%

3.8 Limitations and Future Work

Several limitations should be acknowledged. First, validation was confined to a single public dataset. Although OCT2017 is the standard benchmark for this task and its official partition provides a disjoint independent test set, its images originate from a limited number of acquisition sites and a single class of spectral-domain device. Performance under different scanner vendors, signal-to-noise regimes and patient populations therefore remains to be established, and the reported accuracy should be read as an upper bound on what would be observed in unrestricted multi-centre deployment.

We therefore specify the external validation protocol to be applied in follow-up work. The model trained on OCT2017 will be evaluated, without retraining or re-tuning, on at least two independent cohorts acquired on different hardware: the Duke SD-OCT dataset of Srinivasan et al. and the OCT-C8 collection, which extends the label space to eight retinal conditions. Enhancement parameters will be held fixed at the values of Section 3.4, and generalisation will be reported as the change in macro-averaged F1-score relative to the internal test set, together with per-class expected calibration error, so that domain shift can be quantified before and after enhancement.

Second, the enhancement parameters are fixed rather than learned; an end-to-end differentiable formulation predicting the threshold and regularisation strength per scan is a natural extension. Third, the analysis operates on individual two-dimensional B-scans and does not exploit the volumetric context of a full OCT cube. Fourth, the framework classifies but neither segments nor quantifies lesions, limiting its direct use for disease staging. Finally, inference latency is hardware-dependent, and prospective clinical evaluation remains necessary before adoption.

4. CONCLUSION

This study presents an attention-guided dual-domain framework combining the discrete wavelet transform, Tikhonov regularisation and an attention-enhanced ResNet-50 for retinal OCT classification, decoupling image enhancement from feature extraction. The enhancement module outperforms conventional and state-of-the-art filters (HE, CLAHE, DnCNN), achieving a PSNR of 31.57 dB and an SSIM of 0.946, with optimal settings at decomposition level 2, $T \in [15, 20]$ and $\lambda \in [0.02, 0.03]$. Feeding these images to the dual-attention network yielded 98.69% accuracy and a macro-averaged F1-score of 98.45% on OCT2017. The framework proved robust to simulated speckle noise, degrading only to 97.3% where the baseline CNN fell to 88.3%, and converged with substantially fewer training samples. These results were obtained with 28.09 M parameters and 3.86 GMac per image, with an enhancement stage accounting for under 0.3% of that arithmetic. Grad-CAM confirmed that attention is directed towards genuine anatomical lesions rather than noise or filtering artefacts. The principal limitation is validation on a single public dataset; external multi-centre validation following the protocol of Section 3.8, together with adaptive or learnable parameters for the Wavelet–Tikhonov cascade, forms the focus of ongoing work.

References

- [1] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *cell*. 2018;172:1122-1131.
- [2] Varghese M, Varghese S, Preethi S. Revolutionizing Medical Imaging: A Comprehensive Review of Optical Coherence Tomography (OCT). *J Opt*. 2025;54:1178-1195.
- [3] Xu H, Zhang X, Lv R. SAFP-Net: A Spatially Augmented Feature Pyramid Network for Automatic Detection of Microaneurysms. *Biomed. Signal Process. Control*. 2025;100:106998.
- [4] Ahmed H, Zhang Q, Donnan R, Alomainy A. Denoising of Optical Coherence Tomography Images in Ophthalmology Using Deep Learning: A Systematic Review. *J Imaging*. 2024;10:86.
- [5] Naik A, Pavana BS, Sooda K. Retinal Disease Classification From Retinal-Oct Images Using Deep Learning Methods. *Machine Intelligence and Data Science Applications*. Springer. 2022:95-104.
- [6] Abbasi SA, Mei D, Wei Y, Xu C, Abbasi SM, Shakil S, Yuan W. Deconvolution Techniques in Optical Coherence Tomography: Advancements, Challenges, and Future Prospects. *Laser Photonics Rev*. 2025;19:2401394.
- [7] Guo T, Zhang T, Lim E, Lopez-Benitez M, Ma F, et al. A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities. *IEEE Access*. 2022;10:58869-903.
- [8] Adiazola J. On the Role of Tikhonov Regularizations in Standard Optimization Problems. 2022. ArXiv preprint: <https://arxiv.org/abs/2207.01139>
- [9] Lepcha DC, Goyal B, Dogra A, Wang SH, Chohan JS. Medical Image Enhancement Strategy Based on Morphologically Processing of Residuals Using a Special Kernel. *Expert Syst*. 2022:e13207.

- [10] Yang Y, Su Z, Sun L. Medical Image Enhancement Algorithm Based on Wavelet Transform. *Electron. Lett.* 2010;46:1.
- [11] Li J, Yang YF, Wang PX, Li B. A Novel Medical Image Enhancement Algorithm Based on Dyadic Wavelet Transform. *Applied Mechanics and Materials.* 2014;513:2726-9.
- [12] Xia KJ, Wang JQ, Cai J. A Novel Medical Image Enhancement Algorithm Based on Improvement Correction Strategy in Wavelet Transform Domain. *Clust Comput.* 2019;22:10969-10977.
- [13] Kumar A, Tomar H, Mehla VK, Komaragiri R, Kumar M. Stationary Wavelet Transform Based Ecg Signal Denoising Method. *ISA Trans.* 2021;114:251-262.
- [14] Luo G, Zhou RG, Luo J, Hu W, Zhou Y, Ian H. Adaptive LSB Quantum Watermarking Method Using Tri-Way Pixel Value Differencing. *Quantum Inf. Process.* 2019;18:49.
- [15] Khattab MM, Zeki AM, Alwan AA, Badawy AS. Regularization-Based Multi-Frame Super-Resolution: A Systematic Review. *J King Saud Univ Comput Inf Sci.* 2020;32:755-762.
- [16] J Zhang, L Zhang, Z Liu, Y Zhang, D Liu, et al, Tikhonov Regularization-Based Extended Kalman Filter Technique for Robust and Accurate Reconstruction in Diffuse Optical Tomography. *J Opt Soc Am A.*2023;40:10-20.
- [17] Ferreira HH, Gastal ES. Recursive Filtering 2D Tikhonov Regularization. In 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE. 2022;1:168-173.
- [18] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. *European conference on computer vision (ECCV).*Springer. 2018:3-19.
- [19] Akinniyi O, Rahman MM, Sandhu HS, El-Baz A, Khalifa F. Multi-Stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture. *Bioengineering.* 2023;10:823.
- [20] Alqudah AM. AOCT-NET: A Convolutional Network Automated Classification of Multiclass Retinal Diseases Using Spectral-Domain Optical Coherence Tomography Images. *Med Biol Eng Comput.* 2020;58:41-53.
- [21] Tayal A, Gupta J, Solanki A, Bisht K, Nayyar A, Masud M. DL-CNN-Based Approach With Image Processing Techniques for Diagnosis of Retinal Diseases. *Multimed Syst.* 2022;28:1417-1438.