STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **15**, February 2026, pp 1226–1235.
Published online in International Academic Press (www.IAPress.org)

Deep Learning-Based Classification of Retinal Pathologies

Kawtar NAIM*, Aziz DAROUICHI

Computer and Systems Engineering Laboratory (L2IS), FST, Cadi Ayyad University, Marrakech, Morocco

Abstract Age-related Macular Degeneration (AMD), Diabetic Macular Edema (DME), Epiretinal Membrane (ERM), Retinal Artery Occlusion (RAO), Retinal Vein Occlusion (RVO), and Vitreomacular Interface Disease (VID) are prevalent eye conditions that can lead to partial or complete vision impairment and blindness. Addressing these challenges in eye care necessitates advanced imaging technologies like Optical Coherence Tomography (OCT). The evolution of OCT from time-domain to frequency-domain techniques has significantly enhanced its utility in routine clinical procedures. This paper introduces a novel R50-CapsNet architecture designed to classify retinal diseases more accurately and reliably. Our approach aims to improve diagnostic accuracy for the OCTDL and Kermany datasets.

Keywords Ophthalmic Disease, Retinal, OCT, Classification, Deep Learning (DL), Artificial Intelligence (AI), CapsNet, ResNet50

AMS 2020 subject classifications: 68T05, 68T07, 68T45

DOI: 10.19139/soic-2310-5070-2767

1. Introduction

An estimated 1.1 billion individuals had vision impairment in 2020. Of this total, about 43 million (M) are blind, 295 M have moderate to severe disability, 258 M have mild disability, and 510 M with near vision impairment [1]. In 2023, the world has 2.2 billion vision-impaired individuals, according to WHO [2]. These issues can lead to conditions of blurry or distorted vision and perceiving shadow or spots. Vision problems are due to conditions such as Age-Related Macular Degeneration (AMD) [3] that lead to central vision loss in older people. Diabetic Macular Edema (DME) is macular edema due to diabetes [4]. Epiretinal Membrane (ERM) is the development of fibrous tissue over the retina [5]. Retinal Artery Occlusion (RAO) results from artery blockage and subsequent instantaneous blindness [6]. Retinal Vein Occlusion (RVO) is occlusion and inflammation of veins [7]. Vitreomacular Interface Disease (VID) is macula traction by vitreous [8]. Choroidal Neovascularization (CNV) is defined as the formation of blood vessels outside the retina beneath the retina [9] and Drusen are yellow material deposited under the retina [10]. Eye care is severely threatened throughout the globe, characterized by disparities in the number and quality of preventive, curative, and rehabilitation services [2]. To overcome such threats, advanced imaging devices have gained popularity. Because the medium in the eye is transparent, Optical Coherence Tomography (OCT) is an essential imaging modality that produces high-resolution cross-sectional pictures [11, 12] and it is particularly useful in the field of ophthalmology [13]. OCT technology has come a long way in the last 25 years, moving from time-domain to frequency-domain techniques. This modification has improved tissue contrast and image capture speed, allowing OCT to be used in routine clinical procedures [14].

The paper is organized as follows: Section 2 provides a review of relevant research and current work in the field, encompassing several Deep Learning models founded on OCT classification method. Datasets and pre-processing

*Correspondence to: Kawtar NAIM (Email:k.naim.ced@uca.ac.ma). Department of Computer Science, FST, Cadi Ayyad University. Bd. Abdelkrim El Khattabi , B.P. 549 Guéliz, Marrakech (40000), Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

techniques used in our study are explained in the third section 3. Section 4 presents our suggested R50-CapsNet model. Section 5 presents experimental results on 2 OCT datasets. Section 6 concludes this paper.

2. Literature Review

In 2024, a recent article [8] introduces a newly established dataset featuring retinal images obtained through Optical Coherence Tomography (OCT), a pivotal medical imaging technology for detailed retinal visualization. This dataset, OCTDL, encompasses over 2,000 images sourced from patients with diverse retinal ailments such as AMD and DME. The primary aim is to leverage this dataset to advance artificial intelligence methods in OCT image analysis, thereby enhancing diagnostic capabilities in ophthalmology [15, 8]. In this article [8], the authors evaluated the effectiveness of the deep learning architectures VGG16 and ResNet50 using their dataset named OCTDL. VGG16 and ResNet50 are standard CNN architectures that have been used and assessed across multiple OCT datasets, thereby providing a robust benchmark for analyzing the performance of the OCTDL dataset with these models. Although VGG and ResNet are regarded as traditional architectures, they continue to demonstrate exceptional performance in numerous image classification tasks [8]. VGG16 [16] is a 16-layer architecture known for its simplicity, the architecture comprises thirteen convolutional blocks followed by 3 fully connected layers, as arranged in thirteen convolutional and three fully connected layers. The activation function for every layer is ReLU, and there are five max-pooling layers in the architecture. A softmax classification layer concludes the network. ResNet, the extension of the VGG architecture, introduced residual connections to solve the problem of the vanishing gradient. ResNet50 model [17] consisting of a total of 50 layers has 48 convolutional layers and one max-pooling layer and one average pooling layer (Figure 1).

In data preparation, the OCTDL dataset was divided into 3 sets: train, validation, and test such that images of a single patient were assigned to only one set. In all experiments, the authors employed the cross-entropy loss function and the Adam optimizer, utilizing a learning rate of 0.0005. Various data augmentation techniques were implemented, like random cropping, flipping (horizontal and vertical), rotation, translation, and Gaussian blur were applied [8]. In addition to detailing OCTDL, [8] offers a comparison between several widely used public OCT datasets. Notably, the Kermany dataset, also known as OCT2017, remains the largest, established in 2017 [18], encompassing over 200,000 OCT images categories such as CNV, DME, Drusen, and Normal cases. Other datasets like RETOUCH [19], OPTIMA [20], and Duke [21] contribute uniquely to retinal fluid classification and segmentation studies, collectively advancing OCT imaging techniques [8]. In addition, the OCTDL was integrated with the OCT2017 dataset for experimental purposes. The outcomes of training neural networks solely on OCTDL, as well as the results from merging it with the OCTID and OCT2017 datasets to address the classification challenge, were discussed. Confusion matrices illustrating the training of ResNet50 and VGG16 using the proposed dataset were included. Various metrics, including Accuracy (ACC), F1-score, and Area Precision (P), were employed, indicating strong and consistent performance across all classes in the OCTDL dataset (see Table 1).

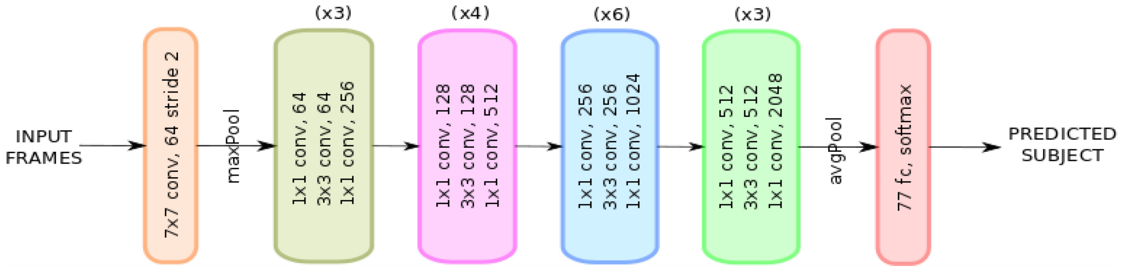


Figure 1. ResNet50 architecture [17]

Table 1. Performance Evaluation of Models [8]

Models	Dataset	Categories	ACC (%)	F1 Score (%)	Precision (P) (%)
ResNet50 [8]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	84.6	86.6	89.8
VGG16 [8]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	85.9	86.9	88.8
ResNet50 [8]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	83.3	80.5	82.3
VGG16 [8]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	81.8	79.8	82.3

3. Dataset and Preprocessing

3.1. *Dataset*

The OCT is a method of imaging that does not require any invasive procedure and is commonly used in clinical practices by ophthalmologists. The early identification and ongoing assessment of retinal diseases are crucial because it shows different layers of the retina. The OCT uses light wave interference to produce microscopic images of the retinal layer and thus it can be useful in diagnosis for various ocular conditions. The OCTDL is Optical Coherence Tomography Dataset for Image-Based Deep Learning Methods [15], consists of sample 2064 images where are categorized and annotated based on disease groups and retinal pathologies (see Figure 2 and Table 2). Between 2013 and 2017, kermany and his team undertook the gathering of a significant dataset [18]. This dataset, known as the OCT2017 dataset and OCT images Balanced version comprises 33,032 OCT B-scan images. Its purpose is to classify various Categories such as CNV, DME, Drusen, and Normal scans. Examples are shown in Figure 3, with the per-class counts summarized in Table 3.

Table 2. Number of images per disease category in the OCTDL dataset [15].

Disease	Images
AMD – Age-related central vision loss.	1231
DME – Retinal swelling from diabetes.	147
ERM – Thin membrane on retina surface.	155
NO – Healthy retinal scan.	332
RAO – Arterial blockage causing sudden vision loss.	22
RVO – Vein blockage with retinal bleeding/swelling.	101
VID – Vitreous pulling on the macula.	76
Total	**2,064**

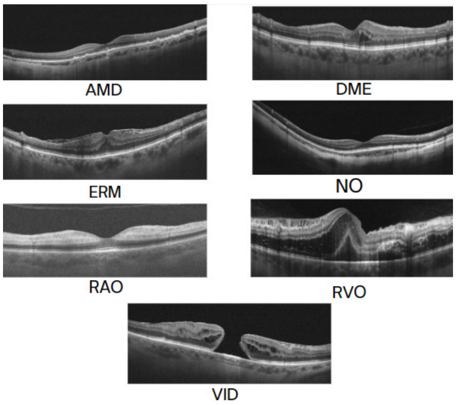


Figure 2. Visualization of retinal disease from the OCTDL dataset [15].

3.2. *Pre-processing*

To ensure optimal performance on the OCTDL dataset [15], combined with OCT2017, we applied a series of preprocessing and augmentation steps. All images were resized to 112×225×3 pixels and normalized to the [0, 255] range for consistency. To improve image quality, a bilateral filter was applied with parameters $d = 9$, $sigmaColor = 75$, and $sigmaSpace = 75$, which reduces noise while preserving retinal boundaries. Data augmentation included horizontal flips (50% probability), small geometric transformations (shifts up to 5%, scaling up to 10%, and rotations up to 10 degrees), brightness and contrast variations (±15%), and light blurring (Gaussian or motion blur with kernel size up to 5), see Table 4. Finally, the dataset was divided into training (90%) and test sets (10%), with the testing set additionally balanced to ensure an equal representation of all classes (see Table 5).

Table 3. Number of images per disease category in OCT2017 dataset [18].

Disease	Training	Testing
Normal [18]	8,016	242
CNV [18]	8,016	242
DME [18]	8,016	242
Drusen [18]	8,016	242
Total	**32,064**	**968**

Table 4. Training-time preprocessing and augmentation settings.

Preprocessing & augmentation (training only)	
Bilateral filter	$d = 9$, $\sigma_{\text{Color}} = 75$, $\sigma_{\text{Space}} = 75$
Geometric	H-flip (50%), shifts $\leq 5\%$, scale $\leq 10\%$, rot $\leq 10°$, shear $\leq 5\%$
Photometric	Brightness/contrast $\pm 15\%$
Blur	Gaussian or motion (kernel ≤ 5)
Split	90% train / 10% test

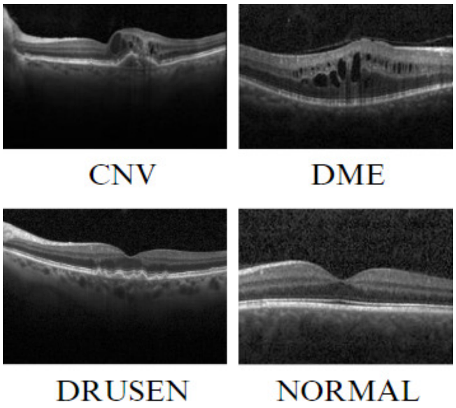


Figure 3. Visualization of OCT2017 classes: CNV, DME, Drusen, Normal [18].

Table 5. Data Splits Overview: OCTDL & OCT2017 Datasets [8].

Dataset	Labels	Training	Testing
OCTDL [8]	AMD, DME, ERM, NO, RAO, RVO, VID	6621	826
OCT2017 + OCTDL [8]	CNV, DME, Drusen, NO	21942	2108

4. Methodology and Implementation

This work is being completed on Kaggle, where free use of NVIDIA TESLA P100 GPUs such as the Nvidia P100 GPU is available on its platform. This GPU supports HBM2 (High Bandwidth Memory) with 16 GB capacity and a memory bus width of 4096 bits. The memory band of this GPU is 732 GB/s, and the per-image compute profile are summarized in Table 6.

Table 6. Reproducibility & runtime environment (Kaggle) with per-image compute metrics.

Kaggle runtime environment	
GPU (visible)	NVIDIA Tesla P100 (16 GB HBM2)
CUDA home	`/usr/local/cuda`
TensorFlow	`2.18.0`
CUDA/cuDNN (TF build)	`12.5.1/9`
TF CUDA compute capabilities[a]	`sm_60, sm_70, sm_80, sm_89, compute_90`
OS / NVIDIA driver	`<from nvidia-smi>`
Compute metrics (per image, forward pass)	
FLOPs	$\sim$ 1.63 GFLOPs ($\approx$ 0.82 GMACs)[b]
Latency (batch = 1, P100)	$\sim$ 33 ms
Profiler	TF Profiler (`order_by = float_ops`)

[a] Capabilities compiled into the TF wheel (device here: P100, `sm_60`).
[b] Rule of thumb: 1 MAC $\approx$ 2 FLOPs.

4.1. Capsule Network

Capsule network (CapsNet) is an image parsing network consisting of capsules, i.e., groups of neurons or any operation that attempts to predict the instantiation presence and parameters (position, size, orientation, etc.) of a given object in a given image region. We represent the entity presence probability using the length of the activity vector and its direction representing instantiation parameters [22]. CapsNet enhances the image analysis by employing a squash activation function that compresses capsule output to maintain orientation information within a limited range. It uses a dynamic routing algorithm that allows capsules to communicate with each other, achieving consensus across layers for better feature representation. CapsNet also employs Margin Loss, a special function used to guide network training based on measurement of difference between ground truth and predicted values, for detailed information refer to [22]. CapsNet relies on three main components:

(1) Squash Function: ensures output vectors have length in $(0, 1)$ while preserving orientation:

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1 + \|\mathbf{s}_j\|^2} \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|}.$$

(2) Dynamic Routing: capsules communicate through routing-by-agreement:

$$\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{j|i}, \quad \hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij} \mathbf{u}_i,$$

with coupling coefficients

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}.$$

(3) Margin Loss: encourages correct class capsules to have long vectors:

$$L_k = T_k \max(0, m^+ - \|\mathbf{v}_k\|)^2 + \lambda(1 - T_k) \max(0, \|\mathbf{v}_k\| - m^-)^2,$$

where $T_k = 1$ if class k is present, otherwise 0.

4.2. Proposed R50-CapsNet Architecture

The R50-CapsNet structure, as depicted in Figure 4, was tailored for ophthalmic disease classification from retinal images. The pre-trained ResNet50 extracts feature maps, followed by convolutional layers with a 5×5 kernel and 148 filters. A 2D max-pooling and 50% dropout enhance feature selection and reduce overfitting. The use of 3×3 and 5×5 kernels enables multiscale feature capture, from fine lesions to broader retinal structures.

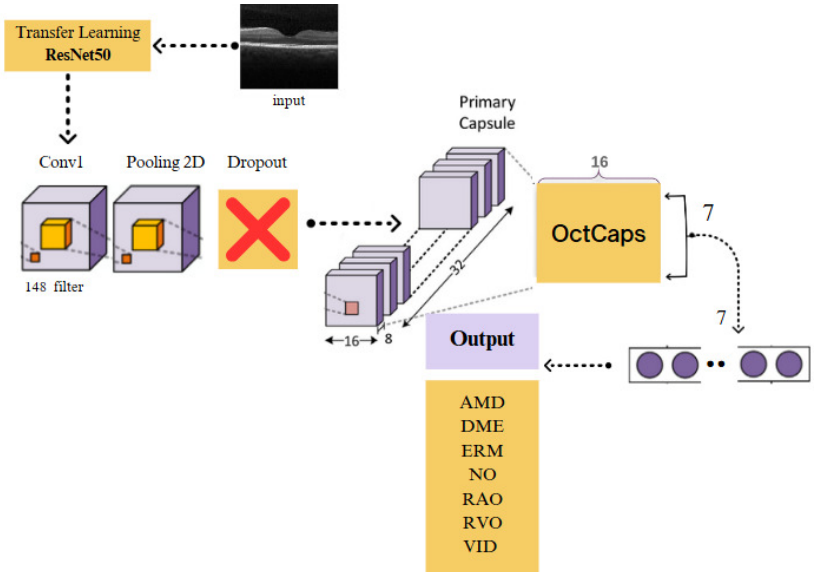


Figure 4. The architecture of our proposed R50-CapsNet approach.

The Primary Caps layer applies 328 filters with 5×5 kernels and a 2-pixel stride to generate 32 capsule maps, each with 8D capsules. The 8D dimension balances representation power and efficiency, consistent with prior CapsNet studies [22].

Finally, the OctCaps output layer represents each of the seven classes with 16D capsules, providing sufficient capacity to model intra-class variability while maintaining interpretability, since each capsule corresponds to one disease category. Dynamic routing aggregates information from lower capsules to support accurate classification.

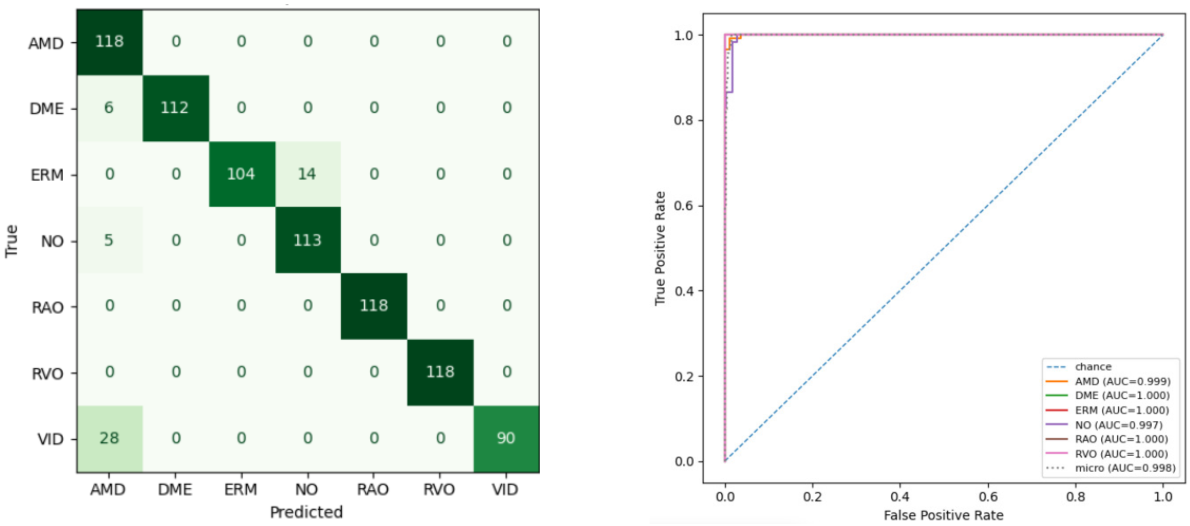
5. Results and Discussions

The R50-CapsNet model stands out for its high performance in classifying ophthalmic diseases, attributed to several factors. Firstly, it leverages the pre-trained weights of ResNet50 from the ImageNet database. Initializing the weights from a model already trained on a vast array of diverse data allows it to benefit from prior knowledge of general visual features. In addition, the use of Capsules provides a significant advantage, as they capture spatial relationships within OCT images more effectively than traditional convolutional techniques—an essential property for ophthalmic disease classification.

To ensure reliable evaluation, we applied 5-fold cross-validation, training the model for 50 epochs with an adaptive learning rate of 0.001 and a batch size of 50. On average, training required around 193 milliseconds per sample. The architecture, with 81 million parameters, is well-proportioned to handle the added complexity introduced by Capsule Networks while fully exploiting ResNet50's feature extraction capabilities.

5.1. *OCTDL Dataset*

5.1.1. Confusion matrix The confusion matrix for the OCTDL dataset indicates that while there are some classification errors, the majority of the 826 retinal image samples are accurately predicted. The results are satisfactory across all seven disease categories, see Figure 5a. The ROC curves (Fig. 5b) are near perfect macro-AUC 0.999; class AUCs: AMD 0.999, DME/ERM/RAO/RVO 1.000, NO 0.997 showing very high true-positive rates with very low false-positive rates.



True \ Predicted	AMD	DME	ERM	NO	RAO	RVO	VID
AMD	118	0	0	0	0	0	0
DME	6	112	0	0	0	0	0
ERM	0	0	104	14	0	0	0
NO	5	0	0	113	0	0	0
RAO	0	0	0	0	118	0	0
RVO	0	0	0	0	0	118	0
VID	28	0	0	0	0	0	90

(a) Confusion matrix

(b) ROC-AUC: macro = 0.999 — micro = 0.998

Figure 5. OCTDL classification results: confusion matrix and ROC curve.

5.1.2. Classification Report This report demonstrates the performance of the R50-CapsNet model. For the OCTDL dataset, the model achieves a maximum precision of 100% for the DME, ERM, VID, RVO and RAO classes, while the minimum precision is 75.2% for the AMD class, see Table 7.

To better understand the impact of the hybrid design, Table 8 presents an ablation study that compares R50-CapsNet with its individual components, ResNet50 and CapsNet. The results show a clear advantage for the combined model. While ResNet50 alone reaches an accuracy of 84.6% and CapsNet achieves 87.6%, the

Table 7. Classification Report for OCTDL dataset

Classes	Precision	Recall	F1-score	Support
AMD	0.752	1.000	0.858	118
DME	1.000	0.949	0.974	118
ERM	1.000	0.881	0.937	118
NO	0.890	0.958	0.922	118
RAO	1.000	1.000	1.000	118
RVO	1.000	1.000	1.000	118
VID	1.000	0.763	0.865	118
Accuracy			0.936	826
Macro Avg	0.949	0.936	0.937	826
Weighted Avg	0.949	0.936	0.937	826

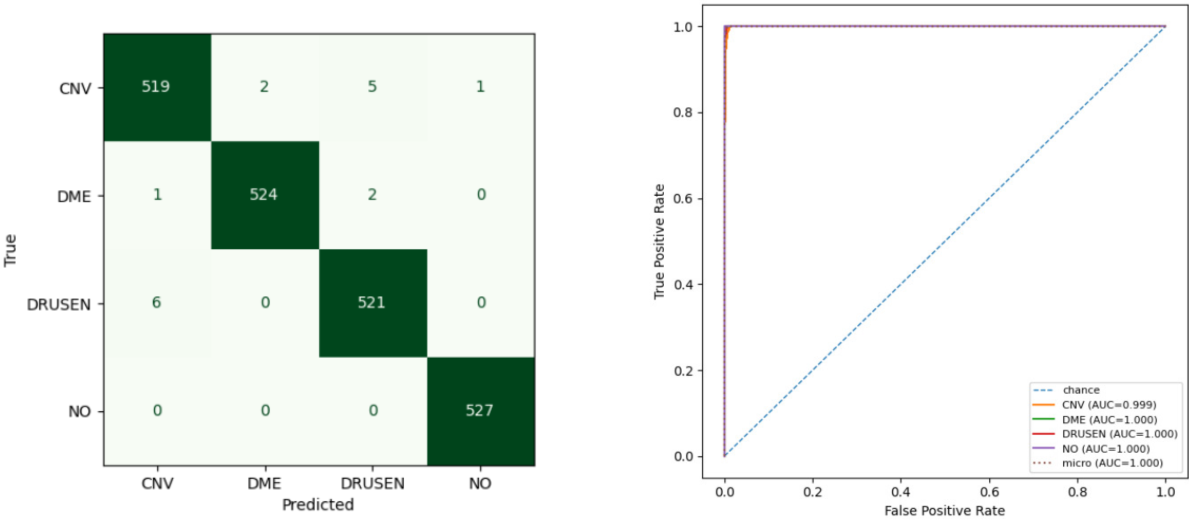
Table 8. Ablation Metric Comparison on OCTDL

Architecture	Accuracy	Precision	Recall	F1-score	AUC
ResNet50	84.6	86.6	89.8	84.6	98.8
CapsNet	87.6	91.0	87.6	88.2	97.2
R50-CapsNet	93.6	94.9	93.6	93.7	99.9

integration of the two pushes the accuracy up to 93.6%. The same trend appears across other metrics: precision, recall, F1-score, and AUC all improve when the models are combined. This demonstrates that R50-CapsNet effectively merges the strong feature extraction of ResNet50 with the spatial representation strengths of CapsNet, resulting in a more powerful and generalizable architecture.

5.2. *OCTDL + OCT2017 Datasets*

5.2.1. Confusion matrix In the combined OCTDL + OCT2017 dataset, the confusion matrix reflects occasional misclassifications, yet the model largely predicts the majority of retinal images accurately. Across all four disease classes, the outcomes are generally satisfactory, see Figure 6a. The ROC curves (Fig. 6b) are perfect AUC = 1.000.



(a) Confusion matrix

(b) ROC-AUC: macro = 1.000 — micro = 1.000

Figure 6. OCTDL + OCT2017 classification results: confusion matrix and ROC curve.

5.2.2. Classification Report This report demonstrates the performance of the R50-CapsNet model. For hybrid OCTDL and OCT2017 dataset, the model achieves a maximum precision of 99.8% for the Normal class while the minimum precision is 98.7% for the DRUSEN and CNV, see Table 9.

Table 10 presents the ablation study on the hybrid dataset. ResNet50 and CapsNet alone achieve accuracies of 95.7% and 95.8%, respectively, while their combination in R50-CapsNet raises accuracy to 99.2%. The hybrid

Table 9. Classification Report for OCTDL + OCT2017 dataset

Classes	Precision	Recall	F1-score	Support
CNV	0.987	0.985	0.986	527
DME	0.996	0.994	0.995	527
DRUSEN	0.987	0.989	0.988	527
NO	0.998	1.000	0.999	527
Accuracy			0.992	2108
Macro Avg	0.992	0.992	0.992	2108
Weighted Avg	0.992	0.992	0.992	2108

Table 10. Ablation Metric Comparison on OCTDL + OCT2017

Architecture	Accuracy	Precision	Recall	F1-score	AUC
ResNet50	95.7	95.4	95.7	95.5	99.6
CapsNet	95.8	95.9	95.8	95.8	99.6
R50-CapsNet	99.2	99.2	99.2	99.2	100

model also improves precision, recall, and F1-score, and reaches a perfect AUC of 100, confirming its reliability and strong generalization ability.

5.3. Discussions

Table 11 showcases the performance evaluation of several models in the realm of OCT image classification, including ResNet50, VGG16, DenseNet, EfficientNet-v2, ViT, Swin Transformer, and the proposed R50-CapsNet. Two distinct datasets were utilized for evaluation: the OCTDL dataset and a combined dataset merging OCTDL with OCT2017's classes (CNV, DME, Drusen, and NO). Across both datasets, R50-CapsNet consistently outperforms the other models. On OCTDL, it improves accuracy by about 7.7% compared to the best baseline, while on the combined dataset it shows a much larger gain of 0.43%. These results highlight the effectiveness of the hybrid architecture in delivering both higher accuracy and more stable performance across different datasets. For MobileNetV3 achieves a per-image latency of 6.95 ms, which is faster than R50-CapsNet but with lower accuracy.

Table 11. Performance Evaluation of Models

Model	Dataset	Categories	ACC (%)	F1 Score (%)	Precision (P) (%)	AUC (%)
VGG16 [8]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	85.9	86.9	88.8	86.9
VGG16 [8]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	81.8	79.8	82.3	99.6
EfficientNet-v2 [23]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	73.28	72.64	77.16	98.12
EfficientNet-v2 [23]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	98.34	98.34	98.35	99.90
ViT [24]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	60	50.99	44.79	73.50
ViT [24]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	90.11	90.09	90.13	98.12
DenseNet121 [25]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	73.44	73.40	82.35	98.84
DenseNet121 [25]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	98.77	98.77	98.78	99.93
Swin Transformer [26]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	80.15	79.91	81.41	98.47
Swin Transformer [26]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	98.23	98.23	98.24	99.89
MobileNetV3 [27]	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	65.79	65.52	77.25	96.46
MobileNetV3 [27]	OCT2017 + OCTDL	CNV, DME, Drusen, NO	98.32	98.32	98.32	99.89
Our R50-CapsNet	OCTDL	AMD, DME, ERM, NO, RAO, RVO, VID	93.6	93.7	94.9	99.9
Our R50-CapsNet	OCT2017 + OCTDL	CNV, DME, Drusen, NO	99.2	99.2	99.2	100

To see what the network is actually keying on, we overlaid Grad-CAM heatmaps on representative OCT B-scans. As shown in Fig. 7, the hotspots fall on the same structures clinicians look for: in AMD/Drusen they trace undulating RPE/Bruch's elevations (drusen); in CNV they outline pigment-epithelial detachment margins with nearby pockets of sub- or intraretinal fluid; in DME they light up cystoid intraretinal spaces (sometimes with a thin SRF layer); in ERM they run along the inner retinal surface where a hyper-reflective membrane distorts the foveal contour; in RAO they emphasize the acutely hyper-reflective inner layers; in RVO they pick out cystoid macular edema (± subfoveal fluid); and in VID/VMT they cluster at points of vitreomacular adhesion producing a "tented" fovea. These class-specific maps don't replace segmentation, but they make the decision process more transparent. Overall, they show that R50-CapsNet attends to anatomically meaningful cues—PED, intra-/subretinal fluid, drusen, and surface traction—supporting the clinical plausibility of its predictions. Still, Grad-CAM is a coarse, last-layer view and saliency methods have known limits; they should complement, not substitute, quantitative evaluation.

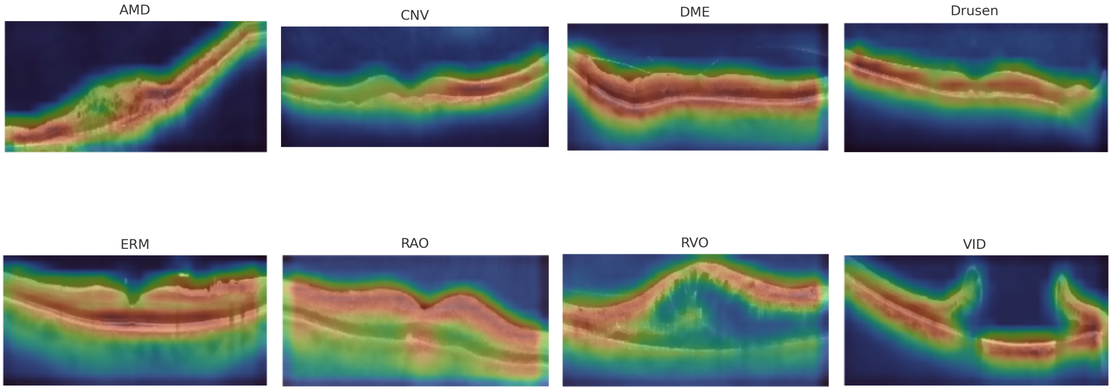


Figure 7. Grad-CAM overlays for each class (AMD, CNV, DME, Drusen, ERM, RAO, RVO, VID)

6. Conclusion and Future Work

This study, our efforts have been dedicated to refining the R50-CapsNet classification model, leveraging Deep Learning principles. Its standout performance, surpassing conventional CNNs, lies in its adeptness at discerning spatial relationships within object features. This unique trait enhances its ability to handle variations in object pose, scale, and orientation, significantly boosting overall performance.

Moreover, the anatomy of the eye has a specific shape, and the presence of a disease can deform this specific shape. Therefore, it is crucial to measure the thickness of the eye layers, the rotation of the layers, the position, and the shape of the macula. Looking ahead, our focus shifts to extending the model's application into ophthalmology, particularly in classifying stages of eye diseases. By expanding its capabilities and integrating clinical data, such as patient history and diagnostic reports, we aim to provide more comprehensive insights into disease progression. This holistic approach not only enhances predictive capabilities but also empowers healthcare professionals to make informed decisions about patient management and treatment strategies.

Through rigorous testing across diverse datasets and a stratified 5-fold cross-validation protocol (with identical preprocessing, data augmentation, class-weighted loss, and early stopping), we demonstrate the efficacy and versatility of R50-CapsNet for ophthalmic disease classification. To support clinical deployment, we also report computational efficiency (per-image inference time at batch = 1 and model size/parameters). We explicitly acknowledge limitations due to minority classes (RAO, VID) and potential dataset bias and ethical considerations, and we report macro-averaged metrics to reduce imbalance sensitivity. These additions strengthen the study's clinical relevance, transparency, and reliability, and future work will expand RAO/VID cohorts, conduct external clinician-led validation, and explore model compression (e.g., pruning/quantization) to further ensure clinical applicability.

REFERENCES

1. Vision Loss Expert Group (VLEG). Data from vleg/gbd 2020 model — global magnitude and projections. https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/global/, 2020.
2. World Health Organization. Blindness and visual impairment – fact sheet. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, 2023.
3. H. L. Cook, Praveen J. Patel, and Adnan Tufail. Age-related macular degeneration: diagnosis and management. *British medical bulletin*, 85:127–49, 2008.
4. Chandni Duphare, Koosh Desai, Priyadarshi Gupta, and Bhupendra C. Patel. *Diabetic Macular Edema*. StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL), 2023 may 23 edition, 2024. PMID: 32119271, Bookshelf ID: NBK554384.
5. Venkata M. Kanukollu and Prateek Agarwal. *Epiretinal Membrane*. StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL), 2023 jul 24 edition, 2024. PMID: 32809538, Bookshelf ID: NBK560703.
6. Evan J. Kaufman, Navid Mahabadi, Sunil Munakomi, and Bhupendra C. Patel. *Hollenhorst Plaque*. StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL), 2024 jan 11 edition, 2024. PMID: 29261979, Bookshelf ID: NBK470445.

7. Kyle Blair and Craig N. Czyz. *Central Retinal Vein Occlusion*. StatPearls [Internet]. StatPearls Publishing, Treasure Island (FL), 2023 may 1 edition, 2024. PMID: 30252241, Bookshelf ID: NBK525985, Free Books & Documents.
8. Mikhail Kulyabin, Aleksei Zhdanov, Anastasia Nikiforova, Andrey Stepichev, Anna Kuznetsova, Mikhail Ronkin, Vasilii Borisov, Alexander Bogachev, Sergey Korotkich, Paul A. Constable, and Andreas Maier. Octdl: Optical coherence tomography dataset for image-based deep learning methods. *Scientific Data*, 11(1), 2024. Cited by: 0; All Open Access, Gold Open Access.
9. Marius Book, Martin Ziegler, Kai Rothaus, Henrik Faatz, Marie-Louise Gunnemann, Matthias Gutfleisch, Georg Spital, Albrecht Peter Lommatzsch, and Daniel Pauleikhoff. Analysis of the vascular morphology of the fibrotic choroidal neovascularization in neovascular age-related macular degeneration using optical coherence tomography angiography. *[Article in English, German]*, Aug 2020.
10. Jayakrishna Ambati, Akshay Anand, Stefan Fernandez, Eiji Sakurai, Bert C. Lynn, William A. Kuziel, Barrett J. Rollins, and Balamurali K. Ambati. An animal model of age-related macular degeneration in senescent ccl-2- or ccr-2-deficient mice. *Nature Medicine*, 9(11):1390 – 1397, 2003. Cited by: 563.
11. Md Rayhan Ahmed, Mohamed S. Shehata, and Patricia Lasserre. Integrating lightweight convolutional neural network with entropy-informed channel attention and adaptive spatial attention for oct-based retinal disease classification. *Computers in Biology and Medicine*, 190:110071, 2025.
12. Kawtar Naim and Aziz Darouichi. Retinal disease classification using ai: A novel capsnet-vgg16 architecture. 2024.
13. Lyvia J. Zhang, Elon H. C. van Dijk, Enrico Borrelli, Serena Fragiotta, and Mark Philip Breazzano. Oct and oct angiography update: Clinical application to age-related macular degeneration, central serous chorioretinopathy, macular telangiectasia, and diabetic retinopathy. *Diagnostics*, 13, 2023.
14. S. Aumann, S. Donner, J. Fischer, and F. Müller. Optical coherence tomography (oct): Principle and technical realization. In J. F. Bille, editor, *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics*, chapter 3. Springer, Cham (CH), August 2019. PMID: 32091846.
15. Mikhail Kulyabin, Aleksei Zhdanov, Anastasia Nikiforova, Andrey Stepichev, Anna Kuznetsova, Vasilii Borisov, Mikhail Ronkin, Alexander Bogachev, Sergey Korotkich, and Andreas Maier. OCTDL: Optical Coherence Tomography Dataset for Image-Based Deep Learning Methods. Mendeley Data, Version 4, 2024.
16. Will T. Nash, Tom Drummond, and Nick Birbilis. A review of deep learning in the study of materials degradation. *npj Materials Degradation*, 2:1–12, 2018.
17. Mohammad Naser Sabet Jahromi, Pau Buch-Cardona, Egils Avots, Kamal Nasrollahi, Sergio Escalera, Thomas Baltzer Moeslund, and Gholamreza Anbarjafari. Privacy-constrained biometric system for non-cooperative users. *Entropy*, 21, 2019.
18. Mohamed Berrimi. Oct images - balanced version (based on kermany dataset). https://www.kaggle.com/datasets/mohamedberrimi/oct-images-balanced-version, 2022. Balanced version derived from the Kermany 2018 OCT dataset.
19. Hrvoje Bogunović, Freerk G. Venhuizen, Sophie Klimscha, Stefanos Apostolopoulos, Alireza Bab-Hadiashar, Ulas Bagci, Mirza Faisal Beg, Loza Bekalo, Qiang Chen, Carlos Ciller, Karthik Gopinath, Amirali Khodadadian Gostar, Kiwan Jeon, Zexuan Ji, Sung Ho Kang, Dara Koozekanani, Donghuan Lu, Dustin Morley, Keshab K. Parhi, Hyoung Suk Park, Abdolreza Rashno, Marinko V. Sarunic, Saad Shaikh, Jayanthi Sivaswamy, Ruwan Tennakoon, Shivin Yadav, Sandro De Zanet, Sebastian M. Waldstein, Bianca S. Gerendas, Caroline C. W. Klaver, Clara I. Sánchez, and Ursula Margarethe Schmidt-Erfurth. Retouch: The retinal oct fluid detection and segmentation benchmark and challenge. *IEEE Transactions on Medical Imaging*, 38:1858–1874, 2019.
20. Jing Wu, Ana-Maria Philip, Dominika Podkowinski, Bianca S. Gerendas, Georg Langs, Christian Simader, Sebastian M. Waldstein, and Ursula Margarethe Schmidt-Erfurth. Multivendor spectral-domain optical coherence tomography dataset, observer annotation performance evaluation, and standardized evaluation framework for intraretinal cystoid fluid segmentation. *Journal of Ophthalmology*, 2016, 2016.
21. Shun Je Chiu, Michael J Allingham, Priyatham S Mettu, Scott W Cousins, Joseph A Izatt, and Sina Farsiu. Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema. *Biomed Opt Express*, 6(4):1172–1194, March 2015.
22. Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. *ArXiv*, abs/1710.09829, 2017.
23. Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training. In *International Conference on Machine Learning*, 2021.
24. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, and Jakob Uszkoreit. Funnel vision transformer for image classification. 2022.
25. K. Santhiya Lakshmi and B. Sargunam. Exploration of ai-powered densenet121 for effective diabetic retinopathy detection. *International Ophthalmology*, 44:1–17, 2024.
26. Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021.
27. Andrew G. Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, 2019.