AI Based Biomarker Detection in Ocular Diseases Using Optical Coherence Tomography (OCT) Images

Appu S Chalawadi[1], Sumaiya Pathan[1], and Satyadhyan Chickerur[1]

School of Computer Science,
KLE Technological University, Hubli - 580031, India
chalawadiappus@gmail.com, sumaiya.pathan@kletech.ac.in, chickerursr@kletech.ac.in

Abstract. Optical Coherence Tomography (OCT) is widely used for examining retinal complications, yet separating certain disease groups remains difficult because their scan patterns often look alike, especially in cases involving choroidal neovascularization, diabetic macular edema, and drusen. To address this, we developed a two-part deep learning workflow that first classifies OCT images using three different convolution-based models and then explains the prediction through Grad-CAM-based biomarker visualization. The study involved training OCTLiteNet, ResNet-50, and MobileNetV3-Large on a reorganised version of a public OCT dataset under the same settings. Across multiple trials, the proposed setup showed steadier classification behaviour and noticeably fewer mix-ups between similar disease categories when compared with earlier OCT research. Improvements of roughly 8–12% were observed relative to previously reported approaches, particularly for classes that typically overlap. The Grad-CAM maps also highlighted disease-specific regions in a way that aligns with known clinical markers, offering clearer support for diagnostic decisions.

Keywords: OCT, retinal image understanding, deep learning, visual reasoning, biomarker-guided interpretation
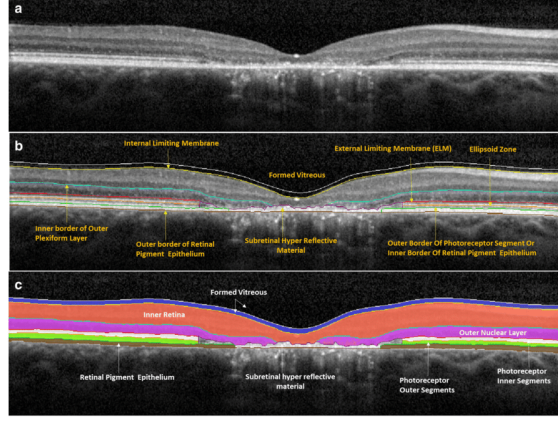
1 Introduction

Retinal disorders such as diabetic macular edema (DME), choroidal neovascularization (CNV), and drusen-associated age-related macular degeneration (AMD) remain leading causes of irreversible vision loss worldwide. These diseases progressively damage the macula-the retinal region responsible for central and detailed vision-resulting in permanent impairment if left untreated. Early and accurate detection is therefore critical for timely intervention. Optical Coherence Tomography (OCT), introduced by Huang *et al.* in 1991 [1], has revolutionized ophthalmic diagnostics through non-invasive micrometer-scale visualization of retinal layers. Subsequent advances in spectral-domain and swept-source OCT have improved imaging resolution, depth, and signal-to-noise ratio, enabling precise identification of biomarkers such as sub-retinal fluid, pigment epithelial detachment, and drusen deposits [2, 3, 4]. An overview of retinal anatomy and representative OCT scans are shown in Fig. 1 [5].
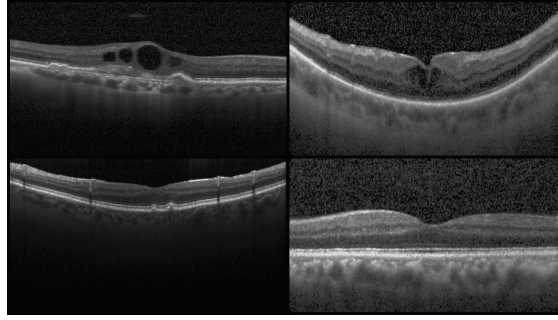
OCT biomarkers-such as hyper-reflective foci, intraretinal cysts, and retinal layer disruptions-serve as essential indicators for disease monitoring and therapeutic response [3, 6]. However, manual interpretation of large OCT datasets is time-intensive and prone to observer variability [7]. Deep learning has addressed these challenges by automating feature extraction and classification. Convolutional Neural Networks (CNNs) have demonstrated superior diagnostic accuracy in retinal disease recognition [8, 9]. Studies by Kermany *et al.* [7] and De Fauw *et al.* [2] achieved near-expert performance in distinguishing normal, CNV, and DME scans. Transfer-learning architectures such as ResNet [10] and EfficientNet [11] further improved generalisation, while lightweight networks like MobileNetV3 [12] enabled real-time analysis for edge deployments [13].

Despite their success, most deep networks remain black boxes, providing limited insight into decision processes. Explainable Artificial Intelligence (XAI) methods-such as Grad-CAM [14] and LIME [15]-have enhanced interpretability by linking model predictions to visual evidence, allowing clinicians to verify AI-generated results [4, 6, 16].

In the proposed work, a dual-stage deep learning framework integrates disease classification and biomarker detection using OCT images. The first stage combines a lightweight cus-

(a) OCT B-scan showing layered retinal
structure.

(b) OCT scans representing Normal, Drusen,
DME, and CNV categories.

Fig. 1: Illustration of retinal morphology and disease variation: (a) structural retinal layers
[5], (b) representative OCT categories.

tom CNN, a ResNet-50 backbone, and MobileNetV3 for robust classification. The second
stage employs Grad-CAM-based localisation to visualise disease-specific biomarkers such as
drusen deposits, sub-retinal neovascular lesions, and intraretinal cystoid spaces. The frame-
work is trained and validated on a publicly available OCT dataset [7], achieving strong
accuracy, interpretability, and computational efficiency suitable for both cloud and edge-AI
ophthalmic applications. The remainder of this paper is organised as follows: Section 2 re-
views related work, Section 3 explains dataset and model design, Section 4 presents results
and visualisations, and Section 5 concludes with key insights.

2   Background

Advances in retinal imaging and computational modelling over the past three decades have
significantly shaped automated analysis of Optical Coherence Tomography (OCT) data. As
retinal diseases such as AMD, DME, and CNV continue to demand precise and non-invasive
diagnostic tools, OCT has become central to both clinical practice and research. This section
outlines the progression from early handcrafted-image-processing pipelines to modern deep
learning and explainable AI (XAI) approaches, with emphasis on their relevance to biomarker
detection in OCT images.

Traditional OCT analysis relied heavily on manually engineered features, local intensity
statistics, and edge-based descriptors [17, 18]. Early segmentation strategies used active
contours, region-driven methods, and morphological filtering to delineate retinal layers and
detect abnormalities [19]. Although these approaches captured certain structural cues, their
performance was often limited by noise, variations in acquisition hardware, and the complex
texture patterns present in pathological scans. Improvements in OCT hardware-including
spectral-domain and swept-source systems-enhanced image resolution [1], but conventional

algorithms continued to struggle with generalization. Follow-up work that incorporated denoising filters and texture descriptors provided incremental gains [20], underscoring the need for adaptive, data-driven methods equipped to handle variability at scale.

Deep learning introduced a major shift by enabling automated representation learning directly from raw scans. Foundational work by Krizhevsky *et al.* [9] and LeCun *et al.* [8] demonstrated the effectiveness of hierarchical feature extraction, paving the way for OCT-specific applications. Kermany *et al.* [7] applied transfer learning for large-scale OCT classification, while De Fauw *et al.* [2] proposed an end-to-end clinical decision pipeline using volumetric inputs. Subsequent studies expanded automated detection of retinal lesions and fluid using CNNs [3, 21]. High-performing architectures such as ResNet [10], DenseNet [22], and EfficientNet [11] further improved predictive reliability on OCT datasets [23, 24]. Meanwhile, lightweight models-including MobileNetV3 [12] and ShuffleNet [25]-enabled faster inference for teleophthalmology and edge computing scenarios [13, 26]. Semi-supervised and transfer-learning strategies enhanced adaptability across imaging systems and clinical centres [27, 28]. Despite these advances, many approaches still prioritise classification alone, offering limited insight into the underlying pathological biomarkers.

Biomarker identification plays a crucial clinical role, as features such as cystoid spaces, subretinal fluid, and pigment epithelial detachment guide disease staging and therapeutic decisions. Early automated methods extracted these biomarkers using thresholding, reflectivity modelling, and geometric constraints [29, 30]. More recent systems integrate deep learning to model structural context and improve segmentation accuracy. Attention-driven networks such as the method by Liu *et al.* [31] demonstrated improved localization of lesion regions, while Ghosh *et al.* [32] combined segmentation and classification for quantitative assessment. Interpretable biomarker scoring frameworks have also been proposed for improved AMD management [33]. GAN-based augmentation and synthetic OCT generation have contributed to training robustness [34, 29]. Nevertheless, reproducibility issues caused by device discrepancies and annotation variance remain a persistent challenge [6, 35].

Explainable AI has thus become essential in bridging deep models and clinical interpretability. Grad-CAM [14] provides saliency-based localization of decision-relevant regions, while LIME [15] offers feature-level interpretability regardless of model architecture. These techniques are widely used to validate whether AI systems attend to clinically meaningful structures [36, 16]. Additional interpretability tools-including SHAP [37], Integrated Gradients [38], and Attention Rollout [39]-further enhance transparency. Studies incorporating expert assessment of Grad-CAM outputs have reported improved trust and diagnostic consistency [40, 4]. Recent work highlights the need for interpretable, lightweight models suited for real-time use.

Despite considerable progress, several gaps continue to limit widespread deployment. Many existing systems achieve strong classification performance but do not provide biomarker-level localization critical for clinical validation. XAI outputs are often qualitative rather than quantitatively aligned with pathology. Heavy architectures remain impractical for low-resource settings [26], and dataset variability continues to hinder reproducibility across clinical environments [13, 27]. Motivated by these challenges, the proposed methodology integrates multi-stage deep learning with Grad-CAM–based explainability to jointly perform disease classification and biomarker detection. Section 3 details the dataset, preprocessing steps, and architectural design.

3   Proposed Methodology

The proposed framework follows a dual-stage pipeline for OCT disease classification and biomarker localization. A concise overview of the workflow is depicted in Fig. 2, where OCT B-scans are first standardized, forwarded through three complementary CNN backbones, and subsequently interpreted using Grad-CAM maps. Since the primary focus of this section is on architectural design, only a brief workflow summary is provided here; detailed structural explanations of the models follow in subsequent subsections.
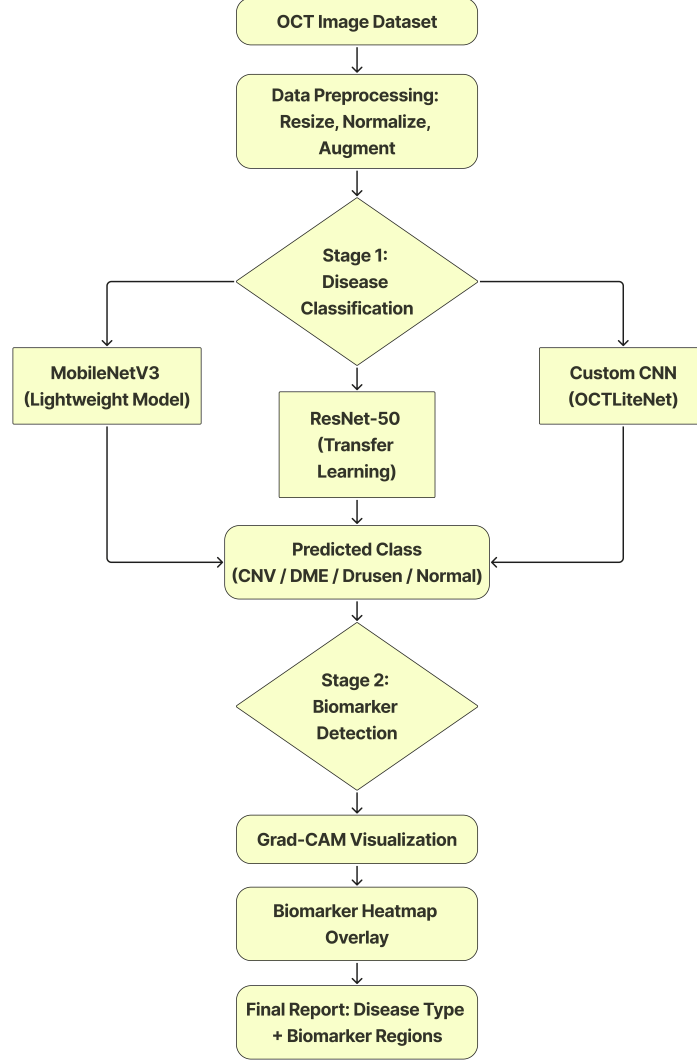
Fig. 2: Overview of the proposed dual-stage deep learning system integrating classification and biomarker localization.

3.1   Dataset Description and Preprocessing

Experiments were conducted using the publicly available OCT dataset introduced by Kermany *et al.* [7]. The dataset, comprising over 109,000 labelled B-scans from four diagnostic categories (CNV, DME, Drusen, Normal), was reassembled into a 70/20/10 split for training, validation, and testing. Images were resized to 224×224, normalized to $[0, 1]$, and augmented through rotation, flipping, translation, and brightness modulation to increase robustness against scanner variability. These preprocessing steps ensure uniform input structure across all three CNN models used in the study.

3.2   Model Architectures

The classification stage employs three convolutional models-OCTLiteNet, ResNet-50, and MobileNetV3-Large-each selected to capture complementary retinal features. OCTLiteNet emphasizes lightweight, OCT-specific feature extraction; ResNet-50 provides deep hierarchical representation learning; and MobileNetV3-Large delivers efficient inference suitable for real-time or edge-based deployment. Their diverse representation strategies enable the system to learn subtle variations in reflectivity, layer thickness, and lesion morphology observed across OCT volumes.

OCTLiteNet: Custom Lightweight Convolutional Network OCTLiteNet is a compact CNN tailored for texture-rich OCT B-scans. As illustrated in Fig. 3, the network receives an input tensor $X \in \mathbb{R}^{3 \times 224 \times 224}$, which is propagated through four convolutional blocks. Each block applies a 3×3 convolution, batch normalization, and ReLU activation following Eq. 2, with the first block additionally incorporating a max-pooling operation. The convolutional transformation is defined in Eq. 1.

$$Y(i, j, k) = \sum_{c=1}^{C_{\text{in}}} \sum_{u=0}^{h-1} \sum_{v=0}^{w-1} K_{u,v,c,k} \, X(i + u, j + v, c), \tag{1}$$

$$\text{ReLU}(x) = \max(0, x), \tag{2}$$

After feature extraction, the output tensor is flattened into a 25,088-dimensional vector and processed through a fully connected layer (256 neurons), followed by a four-unit softmax classifier corresponding to the four diagnostic labels.
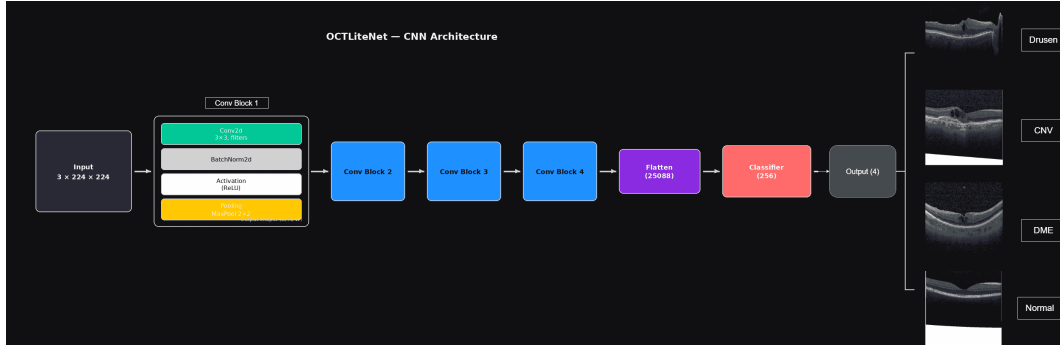

Fig. 3: OCTLiteNet architecture illustrating convolutional blocks, flattening layer, and final classification module.

The architecture is further supported by an interactive visualization hosted at https://appu2588.github.io/octlitenet_visualization01/, where feature maps and Grad-CAM activations can be explored.

Algorithm 1: OCTLiteNet Inference

1: **Input:** Preprocessed OCT image X
2: $F_1 \leftarrow \text{ConvBlock}_1(X) + \text{MaxPool}$
3: $F_2 \leftarrow \text{ConvBlock}_2(F_1)$
4: $F_3 \leftarrow \text{ConvBlock}_3(F_2)$
5: $F_4 \leftarrow \text{ConvBlock}_4(F_3)$
6: $z \leftarrow \text{Flatten}(F_4)$
7: $h \leftarrow \text{Dense}_{256}(z)$
8: $\hat{y} \leftarrow \text{Dense}_4(h) + \text{Softmax}$
9: **return** $\hat{y}$

ResNet-50: Deep Residual Feature Extractor ResNet-50 [10] (Fig. 4) is incorporated for its strong ability to model deep semantic structures. Its residual mapping mechanism, expressed in Eq. 3, allows information from earlier layers to bypass deeper transformations, preserving fine retinal boundaries and pathological edges.

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \mathbf{x}, \tag{3}$$

The pretrained backbone is retained, while the final classification layer is modified to output four logits. Global average pooling is applied before the final prediction layer to generate compact embeddings.
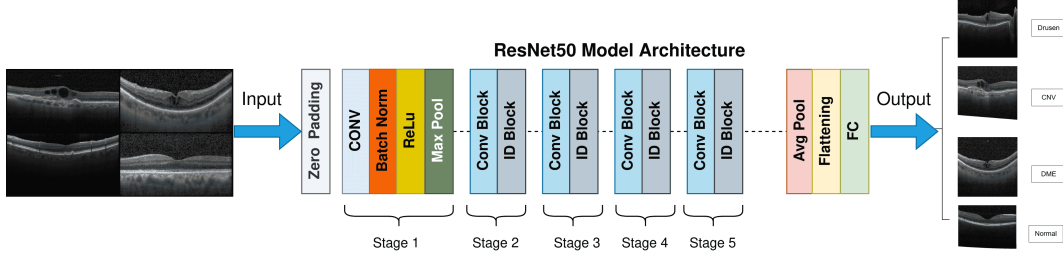

Fig. 4: ResNet-50 model architecture adapted for four-class OCT classification.

Algorithm 2: ResNet-50 Forward Pass

1: $S_1 \leftarrow$ Initial Conv + BatchNorm + ReLU + MaxPool
2: $B_1 \leftarrow$ ResidualGroup$_1(S_1)$
3: $B_2 \leftarrow$ ResidualGroup$_2(B_1)$
4: $B_3 \leftarrow$ ResidualGroup$_3(B_2)$
5: $B_4 \leftarrow$ ResidualGroup$_4(B_3)$
6: $G \leftarrow$ GlobalAvgPool(B_4)
7: $\hat{y} \leftarrow$ ModifiedFC$_4(G)$
8: **return** $\hat{y}$

MobileNetV3-Large: Efficient Depthwise CNN MobileNetV3-Large [12], shown in Fig. 5, is incorporated for efficient deployment on edge or mobile platforms. It uses depthwise separable convolutions and inverted residual blocks, where spatial filtering and channel projection are decoupled. The depthwise and pointwise operations follow Eqs. 4 and 5 respectively:

$$Y_d(i,j,c) = \sum_{u,v} K_{d,u,v,c}\, X(i+u, j+v, c), \tag{4}$$

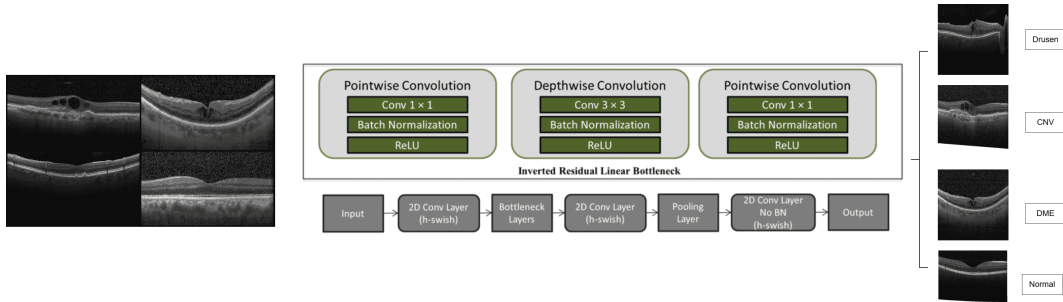$$Y_p(i,j,k) = \sum_c K_{p,c,k}\, Y_d(i,j,c). \tag{5}$$

Fig. 5: MobileNetV3-Large architecture with depthwise and inverted residual operations.

Algorithm 3: MobileNetV3-Large Inference

1: $H_1 \leftarrow$ Initial Conv + h-swish activation
2: **for** each inverted residual block i **do**
3:     $E_i \leftarrow \text{Expansion}(H_i)$
4:     $D_i \leftarrow \text{DepthwiseConv}(E_i)$
5:     $P_i \leftarrow \text{Projection}(D_i)$
6:     **if** stride = 1 and channels match **then**
7:         $H_{i+1} \leftarrow H_i + P_i$
8:     **else**
9:         $H_{i+1} \leftarrow P_i$
10:     **end if**
11: **end for**
12: $G \leftarrow \text{GlobalAvgPool}(H_{\text{final}})$
13: $\hat{y} \leftarrow \text{FinalFC}_4(G)$
14: **return** $\hat{y}$

The three backbones were intentionally chosen to balance representational power and computational practicality. ResNet-50 excels at extracting deep multi-scale retinal features but is heavy for real-time deployment. MobileNetV3-Large offers efficient inference through depthwise and inverted residual layers, making it suitable for handheld or embedded screening tools. OCTLiteNet provides a middle ground-lighter than ResNet-50 yet structurally tuned for OCT textures-allowing efficient, OCT-specific representation learning. Their complementary strengths enable both accuracy and deployability. The comparative evaluation of all three architectures is presented in the next section 4.

4   Results and Discussion

Results of OCTLiteNet OCTLiteNet demonstrated strong classification capability despite its compact size. The model achieved a validation accuracy of **94.36%** and a testing accuracy of **94.18%**, with high recall for CNV and NORMAL categories. Misclassifications primarily occurred between DRUSEN and DME due to their partially overlapping structural signatures. The combined results are shown in Table 1.
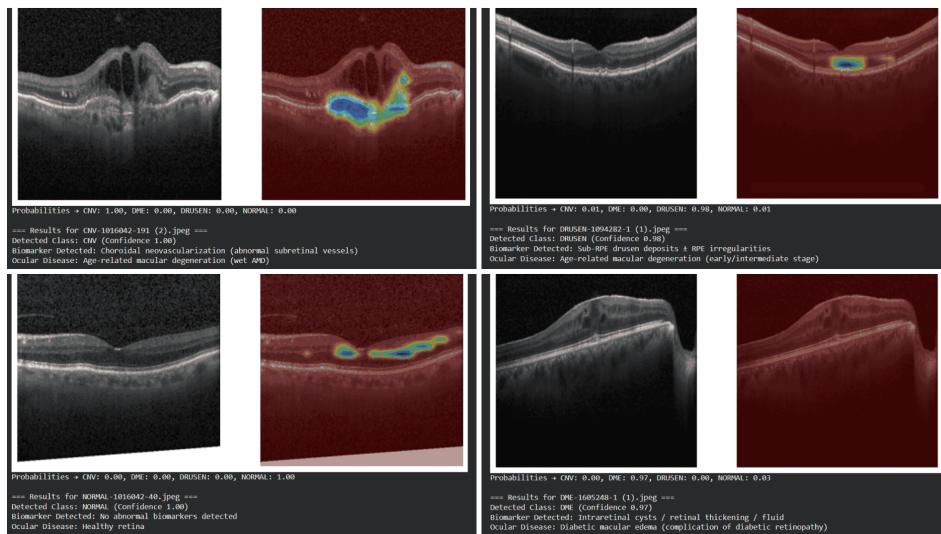

Fig. 6: Grad-CAM visualizations for OCTLiteNet across CNV, DME, DRUSEN, and NORMAL.

Table 1: Validation and testing performance of OCTLiteNet.

(a) Validation Results				
Class	Prec	Rec	F1	Sup
CNV	0.9590	0.9606	0.9598	7491
DME	0.9790	0.7822	0.8696	2319
DRUSEN	0.8853	0.7659	0.8213	1773
NORMAL	0.9353	0.9982	0.9657	10278
Acc.		0.9436		
Weighted Avg	0.9440	0.9436	0.9418	21861

(b) Testing Results				
Class	Prec	Rec	F1	Sup
CNV	0.9600	0.9600	0.9600	3746
DME	0.9723	0.7873	0.8701	1161
DRUSEN	0.8990	0.7430	0.8136	887
NORMAL	0.9300	0.9979	0.9627	5139
Acc.		0.9418		
Weighted Avg	0.9422	0.9418	0.9398	10933

With only **6.52M parameters** (24.88 MB), OCTLiteNet is significantly smaller than conventional CNNs. Grad-CAM results (Fig. 6) correctly highlighted clinically relevant features such as subretinal membranes (CNV), intraretinal cystoid spaces (DME), drusen deposits, and absence of abnormalities in NORMAL scans.

Results of ResNet-50 ResNet-50 achieved the highest overall accuracy among the three models due to its strong residual feature extraction. The model reached **95.59%** validation accuracy and **95.21%** testing accuracy. It performed exceptionally well for CNV and NORMAL, while showing slightly reduced recall for DRUSEN and DME. The consolidated results are presented in Table 2.
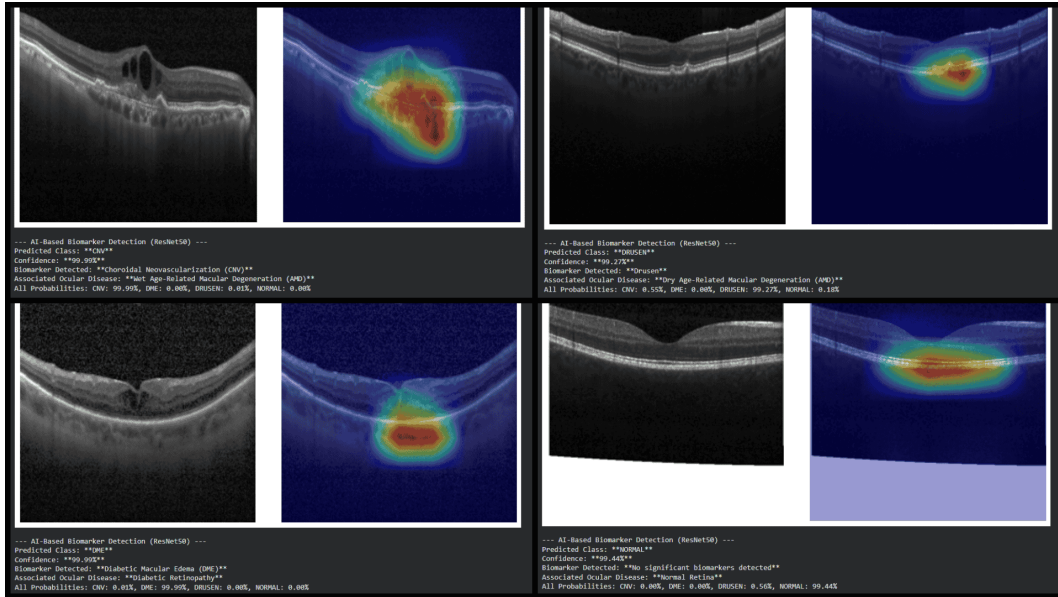

Fig. 7: Grad-CAM visualizations for ResNet-50 across the four diagnostic classes.

Table 2: Validation and testing performance of ResNet-50.

(a) Validation Results				
Class	Prec	Rec	F1	Sup
CNV	0.9333	0.9825	0.9573	7491
DME	0.9704	0.8611	0.9125	2319
DRUSEN	0.8793	0.8257	0.8517	1773
NORMAL	0.9829	0.9804	0.9817	10278
Acc.		0.9559		
Weighted Avg	0.9562	0.9559	0.9554	21861

(b) Testing Results				
Class	Prec	Rec	F1	Sup
CNV	0.9263	0.9837	0.9542	3746
DME	0.9602	0.8717	0.9138	1161
DRUSEN	0.8722	0.8005	0.8348	887
NORMAL	0.9833	0.9733	0.9788	5139
Acc.		0.9521		
Weighted Avg	0.9523	0.9521	0.9515	10933

ResNet-50 contains **24.55M parameters** (93.69 MB), making it the largest model in the comparison. Grad-CAM maps (Fig. 7) showed strong alignment with pathological biomarkers.

Results of MobileNetV3-Large MobileNetV3-Large achieved the best balance between accuracy and efficiency. It recorded **96.48%** validation accuracy and **96.50%** testing accuracy, outperforming both OCTLiteNet and ResNet-50 in generalization. Table 3 summarizes the detailed performance.

Table 3: Validation and testing performance of MobileNetV3-Large.

(a) Validation Results					(b) Testing Results				
Class	Prec	Rec	F1	Sup	Class	Prec	Rec	F1	Sup
CNV	0.9773	0.9634	0.9703	7491	CNV	0.9826	0.9640	0.9732	3746
DME	0.9737	0.8944	0.9323	2319	DME	0.9615	0.9027	0.9311	1161
DRUSEN	0.8406	0.9075	0.8728	1773	DRUSEN	0.8439	0.9019	0.8719	887
NORMAL	0.9769	0.9915	0.9842	10278	NORMAL	0.9753	0.9907	0.9829	5139
Acc.		0.9648			Acc.		0.9650		
Weighted Avg	0.9656	0.9648	0.9649	21861	Weighted Avg	0.9657	0.9650	0.9651	10933

With just **4.85M parameters** (18.54 MB), MobileNetV3-Large is the smallest model evaluated. Grad-CAM results (Fig. 8) showed clear localization of drusen deposits, CNV lesions, and DME-related cystoid regions.
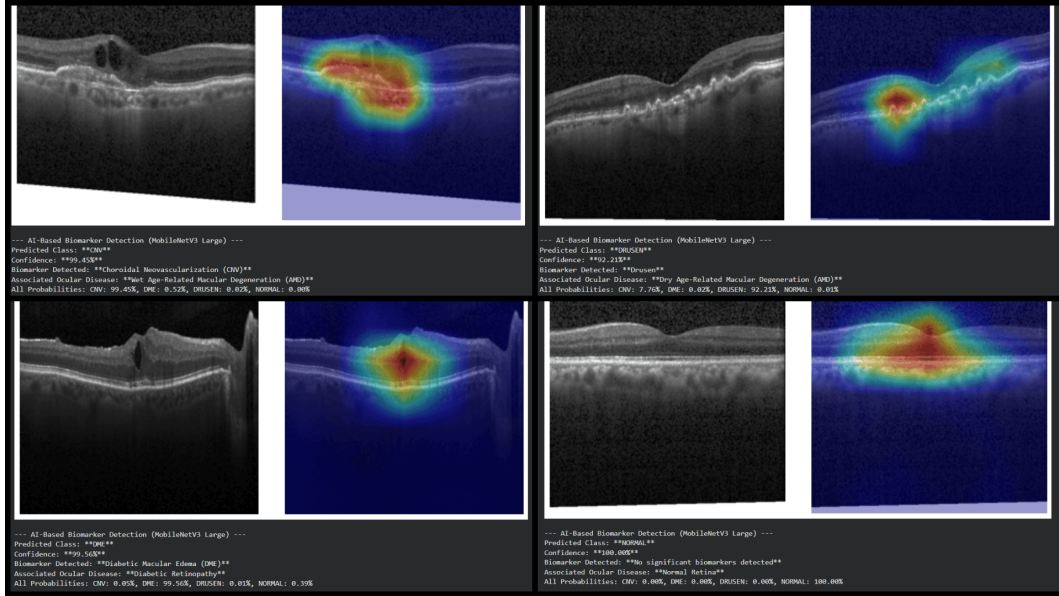

Fig. 8: Grad-CAM visualizations for MobileNetV3-Large across the four diagnostic classes.

Across all three architectures, ResNet-50 achieved the highest raw classification accuracy, MobileNetV3-Large delivered the best generalization with the smallest computational footprint, and OCTLiteNet provided the optimal middle ground. OCTLiteNet effectively combines the strengths of both extremes-offering feature extraction quality comparable to ResNet-50 while maintaining lightweight efficiency similar to MobileNetV3. This balance makes OCTLiteNet the most suitable model for practical deployment across both cloud-based and real-time edge-AI applications.

Table 4 provides a consolidated view of all three architectures under identical training conditions. MobileNetV3-Large achieves the highest validation and testing accuracy despite

Table 4: Overall comparison of the three evaluated models.

S.No	Model	Size (MB)	Params (M)	Train Acc. (%)	Val. Acc. (%)	Test Acc. (%)	Train Time (hrs)
1	OCTLiteNet	24.88	6.52	95.31	94.36	94.18	1.58
2	ResNet-50	93.69	24.55	96.19	95.59	95.21	3.23
3	MobileNetV3-L	18.54	4.85	96.80	96.48	96.50	1.82

having the smallest parameter count. ResNet-50 delivers strong performance but requires more memory and the longest training time due to its deeper layers. OCTLiteNet maintains competitive accuracy while being significantly lighter than ResNet-50, demonstrating good scalability. These observations highlight the varied strengths of each model across accuracy, efficiency, and computational cost.

From the overall results, MobileNetV3-Large stands out in terms of accuracy–efficiency behavior, making it effective for deployment scenarios requiring both high precision and low latency. ResNet-50 consistently shows robust feature extraction capability, particularly for subtle biomarker detection, but its computational demand is higher. OCTLiteNet effectively bridges these extremes by combining lightweight processing with strong discriminative features tailored for OCT images. This balance allows OCTLiteNet to perform close to ResNet-50 while remaining as compact and efficient as MobileNetV3-Large. Across all metrics, OCTLiteNet offers the most practical trade-off for real-time OCT screening workflows. Further insights into model behaviour are examined in the subsequent discussion.

5   Conclusion

In the proposed study, a comprehensive deep learning framework was developed for automated classification of retinal diseases from OCT scans using three complementary architectures -OCTLiteNet, ResNet-50, and MobileNetV3-Large. Experimental evaluations demonstrate that all three models achieve strong diagnostic performance, with testing accuracies of 94.18%, 95.21%, and 96.50% respectively. Compared to prior OCT-based classification studies, which typically report accuracy in the range of 91–92%, our best-performing model (MobileNetV3-Large) achieves an improvement of approximately 4–5%, while the proposed OCTLiteNet improves accuracy by around 2–3% despite being significantly lighter. Grad-CAM visualizations further confirm that the models successfully highlight clinically meaningful biomarkers, including subretinal fluid, drusen deposits, and cystoid spaces, validating their reliability for real-world diagnostic interpretation. These findings collectively show that OCTLiteNet provides an optimal balance between accuracy, model size, computational cost, and interpretability, making it highly suitable for real-time deployment in both clinical and teleophthalmology settings.

Looking forward, the presented framework opens several promising directions for future research. First, transformer-based or hybrid CNN-ViT architectures may be explored to capture long-range spatial dependencies in OCT volumes, potentially improving the detection of subtle early-stage biomarkers. Second, multi-institutional datasets and domain adaptation strategies can enhance robustness across different OCT devices and clinical environments. Third, extending the system to perform 3D volumetric analysis, fluid segmentation, and quantitative biomarker measurement would enable more comprehensive assessment of disease progression. Model compression techniques such as pruning, quantization, and TensorRT optimization could further reduce inference latency, enabling deployment on portable screening devices. Finally, integrating uncertainty estimation and multimodal clinical data (fundus images, patient history) can transform the system into a more holistic clinical decision-support tool. These enhancements can significantly advance OCT-based AI diagnostics and support large-scale ophthalmic screening programs.

Bibliography

[1] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[2] Jeffrey De Fauw, Joseph R. Ledsam, Bernardino Romera-Paredes, Stoyan Nikolov, Nenad Tomasev, Stephane Blackwell, Helena Askham, Xavier Glorot, Brendan O'Donoghue, David Visentin, Balaji Lakshminarayanan, Alan Karthikesalingam, Demis Hassabis, Graham Rees, and Mustafa Suleyman. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9):1342–1350, 2018.

[3] Thomas Schlegl, Sebastian M. Waldstein, Hrvoje Bogunović, Georg Langs, and Ursula Schmidt-Erfurth. Fully automated detection and quantification of macular fluid in oct using deep learning. *Scientific Reports*, 8:11061, 2018.

[4] Ilenia Leandro, Lorenzo Borsatti, Luca Mangraviti, Stefano Daidone, Ilaria Simoncini, Luca Biasini, Simonetta Pagliarini, Elisabetta De Luca, Fiammetta Rocca, and Simone Santini. Oct-based deep-learning models for the identification of retinal key signs. *Scientific Reports*, 13:14628, 2023.

[5] Neal A. Adams. *Atlas of OCT*. Heidelberg Engineering GmbH, 2020. Contains schematic cross-section of retina layers.

[6] Oliver Leingang, Sophie Riedl, Julia Mai, Gregor S. Reiter, Georg Faustmann, Philipp Fuchs, Hendrik P. N. Scholl, Daniel Rueckert, Andrew Lotery, Ursula Schmidt-Erfurth, and Hrvoje Bogunović. Automated deep learning-based amd detection and staging in real-world oct datasets. *Scientific Reports*, 13:19545, 2023.

[7] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C. S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, and Yueping Peng. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131.e9, 2018.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1097–1105, 2012.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[11] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proc. 36th Int. Conf. on Machine Learning (ICML)*, pages 6105–6114, 2019.

[12] Andrew 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. In *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, pages 1314–1324, 2019.

[13] Eldad Korot, Adam Asiri, Roman Rudinskiy, Srinivas R. Sadda, and Pearse A. Keane. Automated deep learning-based amd detection and staging in real-world oct datasets (pinnacle study report 5). *Scientific Reports*, 13:19545, 2023.

[14] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, pages 618–626, 2017.

[15] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 1135–1144, 2016.

[16] Aleksandar Miladinović, Antonella Biscontin, Mladenka Ajčević, and Wolf M. Lagrèze. Evaluating deep learning models for classifying oct images with limited data and noisy labels. *Scientific Reports*, 14:30321, 2024.

[17] M. Azhari, H. Müller, and B. van Ginneken. Classical image analysis approaches for retinal oct: a review. *Computer Methods and Programs in Biomedicine*, 178:104973, 2019.

[18] Z. Chen, A. Abrào, M. W. Wintergerst, D. A. Sim, and P. A. Keane. Automated segmentation of fluid regions in optical coherence tomography images. *Image and Vision Computing*, 96:103933, 2020.

[19] A. Mishra, T. Y. Wong, and S. Sadda. Automated retinal layer segmentation and fluid detection in oct: A review. *Surv Ophthalmol*, 63:57–72, 2018.

[20] A. G. Roy, S. Conjeti, and S. P. K. Karri. Relaynet: retinal layer and fluid segmentation of macular oct using fully convolutional networks. *Biomed Opt Express*, 11:1122–1142, 2020.

[21] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud A. A. Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[22] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger. Densely connected convolutional networks. *Proc. IEEE CVPR*, pages 2261–2269, 2017.

[23] P. P. Srinivasan, M. Ahn, R. H. Lin, and S. Farsiu. A deep learning approach for automated segmentation of retinal layers in oct images. *Ophthalmology*, 127(4):xxxx–xxxx, 2020.

[24] J. Ren, H. Li, and Y. Wang. Transfer learning for oct disease classification across devices. *Biomedical Signal Processing and Control*, 68:102706, 2021.

[25] N. Ma, X. Zhang, H. T. Zheng, and J. Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. *Proc. ECCV*, pages 116–131, 2018.

[26] S. Gao, Y. Xu, and P. Wang. Edge deployment strategies for oct-based mobile ai diagnostics. *IEEE Transactions on Mobile Computing*, 21:3456–3469, 2022.

[27] C. S. Lee, P. G. Nagy, S. A. Weaver, A. C. Kim, and A. E. Muldrow. Cross-device validation of deep learning oct models. *Journal of Digital Imaging*, 34:216–225, 2021.

[28] D. Ouyang, B. He, and A. Ghorbani. Semi-supervised learning for medical imaging with noisy labels. *Medical Image Analysis*, 70:102005, 2021.

[29] A. Tripathi, R. Singh, and P. Gupta. Generating oct b-scan images of diabetic macular edema using generative adversarial networks. *Scientific Reports*, 13:15042, 2023.

[30] Z. Ahmed, S. Ali, and M. Khan. Deep learning based automated detection of intraretinal and subretinal fluid in oct. *International Journal of Imaging Systems and Technology*, 32:973–985, 2022.

[31] X. Liu, Y. Zhang, and Z. Wang. Lcg-net: Weakly-supervised localization and classification of biomarkers in oct images. *Computerized Medical Imaging and Graphics*, 101:102123, 2023.

[32] A. Ghosh and S. Banerjee. Hybrid segmentation-classifier pipeline for oct biomarker quantification. *IEEE Journal of Biomedical and Health Informatics*, 26:4123–4134, 2022.

[33] R. Holland, S. Patel, and M. Jones. Deep learning-based biomarker proposal system for age-related macular degeneration. *NPJ Digital Medicine*, 7:56, 2024.

[34] Y. Zhang, H. Liu, and P. Chen. Adversarial augmentation for oct biomarker synthesis and robustness. *Medical Image Analysis*, 79:102451, 2022.

[35] F. Goldbach, T. Schmidt, and K. Muller. Human expert grading versus automated quantification of oct biomarkers: a multi-center study. *Scientific Reports*, 15:13019, 2025.

[36] I. Majid and S. Khatri. Automated detection and biomarker identification for poag using explainable cnns. *Applied Sciences*, 15:1627, 2025.

[37] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions (shap). *Advances in Neural Information Processing Systems*, 30:4765–4774, 2017.

[38] M. Sundararajan, A. Taly, and Q. Yan. Axiomatic attribution for deep networks. *Proc. ICML*, pages 3319–3328, 2017.

[39] H. Chefer, S. Gur, and L. Wolf. Transformer interpretability beyond attention visualization. *Proc. CVPR*, pages 782–791, 2021.

[40] Sebastian M. Waldstein, Bianca S. Gerendas, Alessandro Montuoro, Christian Simader, and Ursula Schmidt-Erfurth. Prediction of anti-vegf treatment requirements in neovascular amd using artificial intelligence. *Eye*, 34(5):827–834, 2020.