E-ISSN 2586-9027
Homepage : https://tci-thaijo.org/index.php/SciTechAsia Science & Technology Asia
Vol.30 No.3 July - September 2025 Page: [40-50]
Original research article

Automated Detection and Segmentation of Choroidal Neovascularization in OCT Using Deep Learning

Pawaris Panyasombat, Anawin Srivoranan, Chanyanud Sriyota, Teerachot Khusuwan, Pakinee Aimmanee*

School of Information Computer and Communication Technology, Sirindhorn International Institute of Technology, Thammasat University Pathum Thani 12120, Thailand

Received 14 May 2025; Received in revised form 23 July 2025
Accepted 29 July 2025; Available online 30 September 2025

ABSTRACT

Choroidal neovascularization (CNV) is a hallmark of age-related macular degeneration (AMD), marked by the growth of abnormal blood vessels beneath the retina that can severely impair vision. Optical Coherence Tomography (OCT), a non-invasive imaging technique, is widely used to detect CNV due to its ability to capture detailed cross-sectional images of the retina. However, the variability in lesion appearance and the presence of artifacts make manual interpretation challenging and time-consuming. To address these limitations, this study explored the application of deep learning models for automated CNV image detection and CNV localization in OCT images. We compared several architectures widely used for segmentation tasks: U-Net, Attention U-Net, DeepLabV3+, DeepLabV3++, Mask R-CNN, and Mask R-CNN+, tested on a balanced dataset of 500 images. Among the tested models, DeepLabV3++ achieved the highest performance with a CNV image detection accuracy of 99.4% and a CNV localization F1-score of 0.80. These findings suggest that deep learning can significantly enhance the efficiency and consistency of CNV diagnosis, paving the way for its integration into clinical workflows to support early screening and treatment of AMD.

Keywords: Choroidal neovascularization; Deep learning methods; Optical coherence tomography

***Corresponding author:** pakinee@siit.tu.ac.th **doi:** 10.14456/scitechasia.2025.45

1. Introduction

Choroidal Neovascularization (CNV) [1] refers to the abnormal growth of blood vessels beneath the retina, which can lead to vision impairment and is commonly associated with conditions like age-related macular degeneration (AMD). CNV in optical coherence tomography (OCT) [2] images is typically located beneath the retinal pigment epithelium (RPE) or between the RPE and the neurosensory retina, depending on the type of CNV. In OCT, CNV often appears as a hyperreflective, irregular, or fibrovascular lesion that disrupts the normal retinal layers. It may be associated with subretinal or intraretinal fluid, thickening of the RPE-choriocapillaris complex, and sometimes hemorrhage or exudates. The OCT image enables detailed cross-sectional imaging of the retina, allowing clinicians to detect the presence, extent, and activity of CNV. The characteristic appearance of CNV on OCT includes a serous retinal detachment, subretinal fluid accumulation, or sometimes, intraretinal fluid. In advanced cases, intraretinal or subretinal hemorrhages may be present, appearing as darker, hypo-reflective regions. The neovascular membrane may exhibit an irregularly tortuous appearance, and cystic spaces can develop in the retina or beneath the RPE. Fig. 1 depicts examples of CNV in OCT images.

Detecting and segmenting CNV in OCT images is challenging due to several factors. First, CNV lesions vary significantly in size, shape, and appearance, making them difficult to distinguish from other retinal abnormalities or artifacts. Second, the subtle and heterogeneous presentation of CNV, including small hyperreflective foci or mild RPE disruptions, can easily be overlooked or mistaken for normal anatomical variations. Third, associated features like subretinal fluid, hemorrhage, or exudates can obscure the underlying CNV, complicating accurate localization. Fourth, the quality of OCT images, affected by noise, motion artifacts, and resolution limitations, further hinders precise detection.

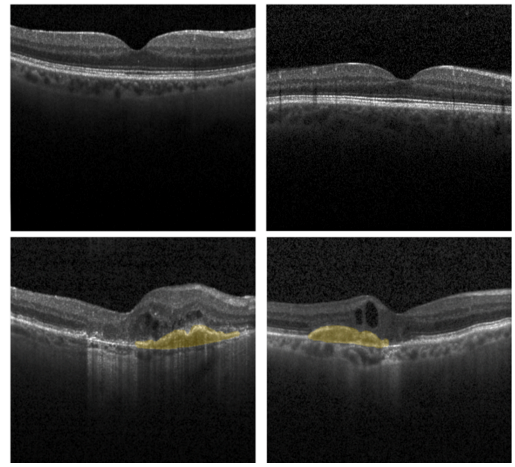

Fig. 1. Examples of healthy OCT Images without CNV (top row) and OCT images with CNV (bottom row). CNVs are in yellow highlights.

Past studies on CNV detection are limited due to the novelty of the OCT domain. The techniques used in these studies are as follows. To the best of our efforts, we found only one research study about CNV image classification. Its brief details are as follows. Ismail et al. [3] proposed an Ensemble Transfer Learning (ETL) model that integrates DenseNet169 and InceptionV3 to classify OCT images into CNV, Diabetes-Related Macular Edema (DME), drusen, and normal cases. The method compared against the sole models and ensemble learning of different models. They discovered that the proposed method achieved superior performance, with an accuracy of 99.8%, on a dataset of 4,000 OCT images (1,000 CNV, 3,000 non-CNV) from a public dataset. Key innovations include adaptive fusion of CNN/Transformer features and attention-

driven lesion localization. Still, limitations include small dataset size, lack of multicenter validation, and high computational load, potentially limiting real-world deployment.

For CNV segmentation, the following studies are found. Xi et al [4] proposed a multi-scale CNN with retinal layer information to segment CNV in OCT images. They tested their method on a set of 15 CNV OCT images. Their method achieved a Dice score of approximately 0.85 and high sensitivity of approximately 90%. It outperforms traditional techniques. Zhang et al [5] proposed MPB-CNN, a multi-scale parallel branch CNN for segmenting CNV in SD-OCT images, leveraging parallel convolutions at different scales to capture fine and coarse features. They compared their method to three baselines: (1) U-Net, (2) Fully Convolutional Network (FCN), and (3) traditional graph-cut segmentation, outperforming all with a Dice score of 0.876 and higher sensitivity (92.3%). The model was tested on a clinical dataset of which the size was unspecified with expert annotations. Meng et al. [6] used an encoder-decoder based multi-scale information fusion network MF-Net based on U-Shape architecture for CNV segmentation in retinal OCT images, combining channel attention and spatial pyramid pooling to enhance multi-scale feature learning. They obtained Dice and Jaccard of 92.9% and 86.8%, respectively, on a public dataset of 380 OCT scans annotated by experts. However, their performances were not significantly better than the comparative methods and computational cost is high. Diao et al. [7] used a Class Activation Map guided UNet (CAM-UNet) to segment CNV. The proposed method achieved the best IOU of only 66.38 for CNV segmentation on the public Duke AMD dataset (with 384 OCT volumes). The performance was moderately high and only slightly better than the comparative methods. The study of Liu et al. [8] used TSSK-Net, a weakly supervised framework for CNV segmentation in OCT images using only image-level labels. They combined token-semantic-supervised learning and adaptive key-patch mining to localize biomarkers. Compared to weakly supervised baselines like Class Activation Mapping (CAM) and Grad-CAM, while TSSK-Net achieves superior CNV segmentation performance. Still the average Dice score reported was only 53.43% on two datasets of a total of 1,950 OCT images (with image-level CNV/normal labels). The performance of CNV remained below that of fully supervised methods. Fu et al. [9] proposed a deep-learning model integrating extreme-point guidance, multi-scale feature learning, and edge-aware attention for CNV segmentation in OCT scans. However, their technique required a user to mark the four extreme points of the lesion to enhance boundary focus. Shen et al. [10] proposed an attention graph convolutional neural network, GA-U-Net, to segment both CNV and retinal layer surfaces in OCT images, utilizing graph attention U-Net as the backbone. They achieved Dice scores of 77.51% and 78.00% for their two datasets. Their method suffered from long training time as most deep learning approaches do. Mori et al. [11] used the SeResNet model with Squeeze-and-Excitation (SE) blocks added to a standard ResNet. They trained on 84,495 OCT images, with a final global average pooling layer feeding into a SoftMax classifier. Their network achieved an overall classification accuracy of 99.0%, with per-class F1-scores ranging from 0.99 to 1.00. However, as a pure classification model, it did not produce pixel-level lesion masks, this limited its utility when exact CNV localization is required. Paul et

al. [12] introduced OCTx, an ensemble model combining VGG16, DenseNet201, InceptionV3, and a custom CNN for multi-class retinal disease classification using the UCSD OCT dataset containing 84,000 images. The model achieved a high performance of nearly 99.5% for classifying CNV, DME, Drusen, and Normal cases. However, OCTx focused solely on classification and did not provide lesion localization. Additionally, the ensemble's large size and heavier inference led to slower deployment.

Existing studies have several limitations. For instance, CNNs are not well-suited for dense prediction tasks, and Grad-CAM is not designed for fine segmentation. MF-Net often struggles with sharp and accurate boundary localization due to its coarse multi-scale fusion strategy. Additionally, TSSK-Net has high computational complexity and training difficulties, particularly in the context of medical image segmentation or vessel extraction. While traditional U-Net is highly effective, it can be outperformed by newer models that incorporate added complexity or features. This work aims to provide a comprehensive evaluation of modern deep learning models, as well as frequently used models for precise segmentation and image classification. We tested the performance of U-Net [13], Attention U-Net [14], DeepLabV3+, DeepLabV3++ [15], Mask R-CNN [16], and Mask R-CNN+ across two tasks: CNV image detection and CNV localization. Our goal is to identify the optimal architecture for this task using a balanced dataset of OCT images. The performance of each model is assessed based on classification metrics (Accuracy, AUC, FNR, FPR) and segmentation metrics (Precision, Recall, F1-score).

2. Methodology

The proposed framework for automated CNV detection and localization in OCT images consist of three main stages: noise removal, image classification, and CNV localization. An overview of the process is illustrated in Fig. 2.

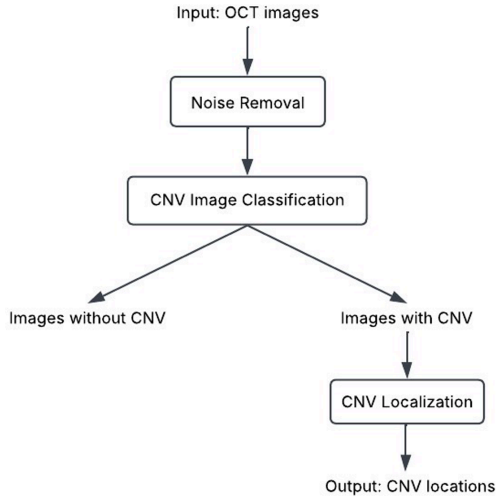

Fig. 2. A diagram depicting the procedures in the methodology.

2.1 Noise removal

OCT images often suffer from various types of noise due to motion artifacts and limitations of imaging hardware, which can negatively affect the performance of segmentation algorithms. To address this, we applied a median filtering technique to reduce noise while preserving important anatomical details. This step enhances image quality and improves the reliability of subsequent processing stages.

2.2 OCT image classification

This stage involves classifying OCT images as either CNV-positive or CNV-negative using deep learning models. We employed several widely used based architectures: U-Net, DeepLabV3+, Mask R-

CNN and their variants: Attention U-Net, Mask R-CNN+, and DeepLab V3++. Brief details about these models are as follows.

U-Net is a widely used segmentation model with an encoder-decoder structure that employs skip connections to retain important spatial information. DeepLab V3++ utilizes advanced convolution techniques to capture detailed image features, thereby improving segmentation accuracy. Mask R-CNN is a two-stage model capable of segmenting isolated areas, while Attention U-Net is an enhanced version of U-Net that incorporates attention mechanisms, allowing the model to focus on specific regions. DeepLab V3++ is a refined version of DeepLab V3+ that integrates components from Mask R-CNN for more precise results. Additionally, Mask R-CNN+ employs extra layers from DeepLab V3+ to enhance segmentation. These models were chosen for their strengths in capturing fine-grained features, handling overlapping structures, and balancing accuracy with computational efficiency.

Each model was trained on 80% of the dataset and evaluated on the remaining 20%. Several data augmentation techniques, including rotation, shifting, zooming, and flipping, were applied during training to enhance model generalization. Training utilized the Adam optimizer with an adaptive learning rate, a batch size of 16, and early stopping based on validation performance to prevent overfitting.

2.3 CNV segmentation

For images classified as CNV-positive, we further localized the CNV regions using the same set of deep learning models. Segmentation probability maps were generated, and a threshold of 0.5 was applied to create binary masks. To reduce false positives, we filtered out small noisy regions using a minimum area constraint. After thresholding, small noisy regions are filtered out by applying a minimum area constraint to eliminate insignificant connected components, thereby reducing false positives in the mask predictions. This ensures that only meaningful regions are retained during evaluation.

We employed the six deep-learning models used earlier in the OCT classification step to segment the CNV regions. The training protocol for segmentation mirrored that of classification. For U-Net and DeepLab variants, we used a combination of binary cross-entropy (BCE) [17] and Dice loss [18]. For Mask R-CNN models, the total loss combined classification loss, bounding box regression loss, and mask loss. Same data augmentation techniques were used for this task. All models were trained with the Adam optimizer, a learning rate of 0.0001, and a batch size of 4. Training was halted if validation performance did not improve for ten consecutive epochs.

3. Experiments

To evaluate the performance of the proposed deep learning models, we conducted experiments using retinal OCT images labeled with CNV obtained from Kaggle.com [19]. All images were initially in JPEG format and were resized to 512×512 pixels for consistency. This dataset was utilized to train the segmentation models in a fully supervised setting, with manually curated annotations ensuring accurate lesion delineation. For evaluation, we manually selected 500 OCT images from the dataset, including 250 CNV-positive and 250 CNV-negative (normal) scans. Training was performed using Google Colab on a personal computing setup equipped with an NVIDIA T4 GPU (16 GB VRAM). All experiments adhered to a consistent preprocessing and

resizing pipeline to standardize image inputs across models.

For evaluation schemes, CNV image classification results were assessed using Accuracy, False Negative Rate (FNR), False Positive Rate (FPR) defined in Eqs. (3.1)-(3.3). Area Under the ROC Curve (AUC) [22] was also measured to evaluate how well each model can distinguish between the two classes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3.1}$$

$$FNR = \frac{FN}{FN + TP}, \tag{3.2}$$

$$FPR = \frac{FP}{FP + TN}, \tag{3.3}$$

where TP is the number of CNV-positive images that are correctly classified as CNV, TN is the number of CNV-negative images that are correctly classified, is the number of CNV-negative images incorrectly classified as CNV-positive, and FN is the number of CNV-positive images incorrectly classified as CNV-negative. It is worth noting that unlike metrics such as Accuracy, FNR, and FPR which are calculated directly from confusion matrix counts, AUC is derived from the overall ranking performance of the classifier and computed using numerical integration (e.g., trapezoidal rule) over the ROC curve. As such, it does not have a simple closed-form formula based on TP, FP, FN, and TN, and is usually computed algorithmically. Thus the formula for AUC is omitted.

To evaluate CNV segmentation performance, we used Precision, Recall, F1-score (also known as the Dice score), and Intersection over Union (IOU), which are mathematically defined in Eqs (3.4)-(3.7).

$$Precision = \frac{tp}{tp + fp}, \tag{3.4}$$

$$Recall = \frac{tp}{tp + fn}, \tag{3.5}$$

$$F1 - score = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall}, \tag{3.6}$$

$$IOU = \frac{tp}{tp + fn + fp}, \tag{3.7}$$

where tp is the number of CNV pixels that are correctly classified as CNV, tn is the number of non-CNV pixels that are correctly classified, fp is the number of non-CNV pixels incorrectly classified as CNV-positive, and fn is the number of CNV-pixels incorrectly classified as non-CNV.

4. Result

The following subsections show results of CNV image classification and CNV segmentation

4.1 CNV image classification results

The raw classification results: TP, TN, FP, FN are presented in Table 1.

Table 1. Raw performance comparison of all models concluded from a confusion matrix.

Model	TP	TN	FN	FP
U-Net	250	245	**0**	5
Attention U-Net	245	237	5	13
DeepLabV3+	**250**	**247**	**0**	**3**
DeepLabV3++	**250**	**247**	**0**	**3**
Mask RCNN	250	246	**0**	4
Mask RCNN+	250	237	1	13

The results show that DeepLabV3+, DeepLabV3++, and U-Net achieved perfect or near-perfect results with minimal false positives or false negatives, while Attention U-Net and Mask R-CNN++ exhibited higher misclassification rates. The detailed performance comparisons in AUC, Accuracy, FNR, and FPR are shown in Table 2.

DeepLabV3+ and DeepLabV3++ achieved the highest classification accuracy

Table 2. Performance comparison of deep learning models for CNV image classification (in %). The bold numbers indicate the best values in categories.

Model	AUC	Accuracy	FNR	FPR
U-Net	99.00	99.00	**0.00**	2.00
Attention U-Net	96.40	96.40	2.00	5.20
DeepLabV3+	**99.40**	**99.40**	**0.00**	**1.20**
DeepLabV3++	**99.40**	**99.40**	**0.00**	**1.20**
MaskRCNN	99.20	99.20	**0.00**	1.60
Mask RCNN+	97.20	97.20	0.40	5.20

and AUC, both with 99.40% accuracy, 0.00% FNR, and a low FPR of 1.20%. Their superior performance is attributed to their use of atrous convolutions, which effectively capture multi-scale contextual information of CNV without losing spatial resolution.

4.2 Segmentation results

The comparisons of the segmented regions from these deep learning models of a selected case are depicted in Fig. 3. All models were able to detect CNV regions with varying edge clarity. Notably, segmentation by Mask R-CNN models tended to produce blockier edges compared to the smoother outputs of DeepLabV3+ variants. Fig. 4 presents the average precision, recall, and F1-score for all models.

Here are observations. For precision, DeepLabV3+ achieved the highest precision, indicating a low rate of false positives. The DeepLabV3++ is slightly less but is still better than all others. For recall, Mask R-CNN, DeepLabV3++, and Attention U-Net performed similarly and achieved higher recall than the other models. In terms of F1-score and IOU, DeepLab V3++ also showed the highest values.

4.3 Training time comparison

As the training time may be one of the concerns in selecting the method to be used. We also provide here the comparison of the training time. Fig. 5 depicts the running time for each model.

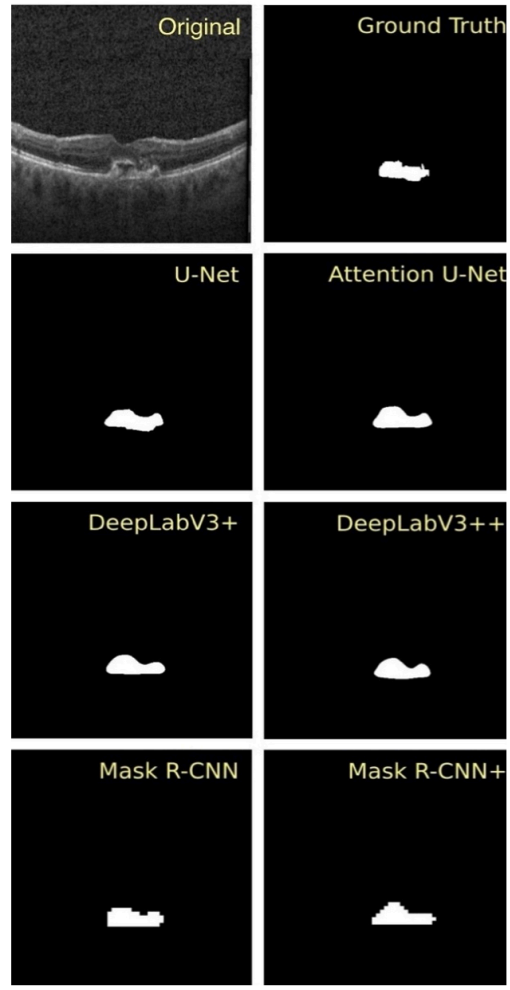

Fig. 3. Illustration of CNV regions resulting from comparative deep learning models.

The graphs show that the training time for DeepLab V3++ and Mask R-CNN is comparable and significantly shorter than that of other models. In contrast, Attention U-Net, Mask R-CNN+, and DeepLabV3+ require moderately more training time, while U-Net has the longest training time, nearly four times that of DeepLabV3++ and Mask R-CNN. Considering both performance and efficiency, DeepLabV3++ is

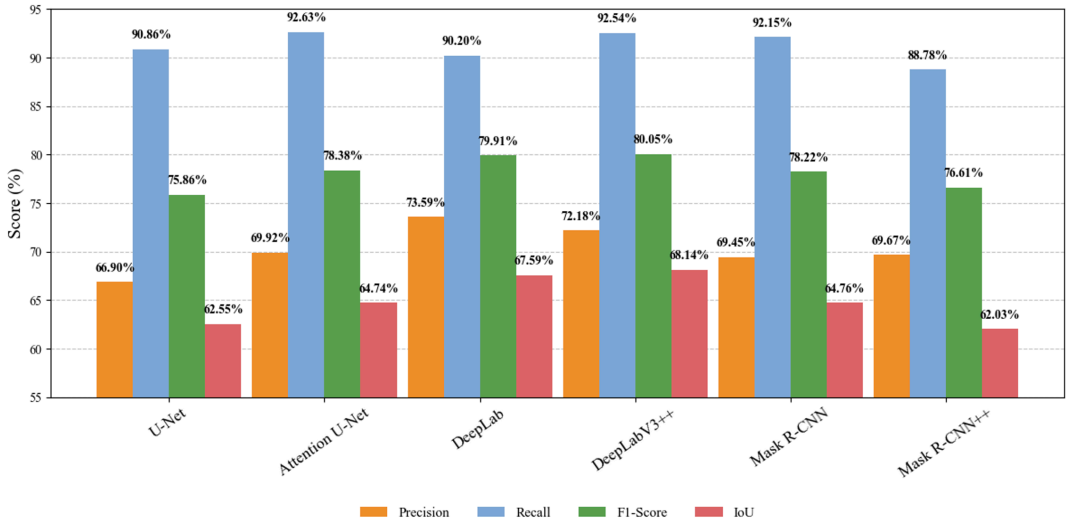

Fig. 4. Segmentation performance comparison of all models.

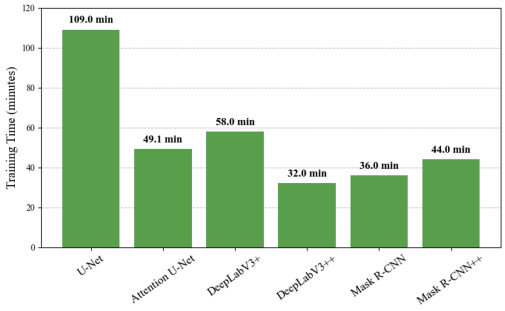

Fig. 5. Models' training time comparison.

recommended for practical use due to its combination of high precision, competitive F1-score, and short training duration.

5. Discussion and Future Work

This section presents an analysis of model performance and outlines key challenges and future directions for improving automated CNV detection and localization. Fig. 6 shows examples of errors from each model. Several recurring patterns were observed. First, models such as Mask R-CNN showed a tendency to misclassify normal anatomical textures as CNV, particularly in images with subtle retinal irregularities. This is likely due to over-sensitivity to edge-like patterns. Second, simpler models like U-Net occasionally failed to detect low-contrast or small CNV regions, underscoring the importance of feature extraction depth and spatial context. Third, most models, especially those not employing multi-scale context or attention mechanisms, struggled to delineate precise lesion borders, leading to over- or under-segmentation.

Fourth, different models sometimes produced conflicting segmentations on the same image, suggesting instability in predictions and the potential benefit of ensemble approaches. Last, inconsistencies in image formats (grayscale vs. RGB) affected performance in some models, particularly those pretrained on natural image datasets.

In this selected case, all deep learning models under-segmented the CNV region. The less bright CNV regions were often not detected by the models, as indicated by the large red area representing false negatives. These occurrences frequently appear near CNV boundaries. Additionally,

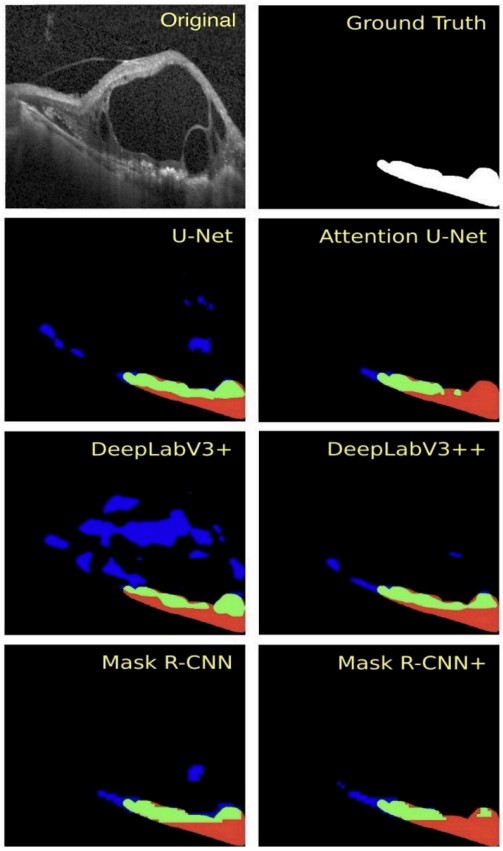

Fig. 6. Illustration of pixel-level classification outcomes for each model, highlighting true positives (green), false positives (blue), and false negatives (red).

regions sharing similar characteristics with CNVs, such as the Retinal Pigment Epithelium (RPE) layer, were also incorrectly identified by the deep learning models. Surprisingly, some areas that were clearly not CNV were incorrectly classified as CNV, as seen in the blue region. We believe these misclassifications may result from data augmentation artifacts or small, noisy background areas. These errors are particularly concerning clinical scenarios because the CNV size is one of the indicators of AMD severity.

To address the identified limitations, several remedies for improvement are proposed. Model fusion or ensemble methods that combine predictions from multiple architectures may yield more robust and stable outputs, particularly in ambiguous cases. Additional techniques can also be explored. For instance, incorporating Conditional Random Fields (CRFs), morphological operations, or shape priors may enhance the sharpness and accuracy of segmentation masks. Additionally, exploring transformer-based or hybrid CNN-transformer models may offer superior performance by improving long-range contextual understanding. In terms of explainability, tools such as Gradient-weighted Class Activation Mapping (Grad-CAM) [21] and SHapley Additive exPlanations (SHAP) [22] could be integrated to visualize model decision-making, thereby increasing clinical trust and interpretability. For example, Grad-CAM can be used to visualize regions of interest, while SHAP can be applied to the metadata to explain which other features influenced the model. These strategies collectively aim to develop a more accurate, reliable, and clinically deployable system for CNV screening in retinal OCT images.

6. Conclusion

This study presents a comparative evaluation of deep learning architectures for the automated detection and localization of choroidal neovascularization (CNV) in retinal OCT images. Given the clinical significance of early CNV identification in managing age-related macular degeneration (AMD), the ability to automate this process reliably is of growing interest in ophthalmic diagnostics. We implemented and assessed several state-of-the-art models—U-Net, Attention U-Net, DeepLabV3+, DeepLabV3++, Mask R-CNN, and Mask R-CNN+—on a balanced

dataset comprising 500 OCT images. Classification performance was evaluated using accuracy, false positive/negative rates, and AUC, while segmentation quality was assessed using precision, recall, and F1-score. Among the models, DeepLabV3++ demonstrated superior performance in both classification (99.4% accuracy, AUC = 0.994) and localization (F1-score = 0.80), attributed to its effective multi-scale feature extraction and robust spatial localization. Mask R-CNN also performed competitively, particularly in precision and training efficiency, making it a strong candidate for real-world applications. While the results are promising, challenges remain in handling subtle CNV features, reducing false positives, and ensuring generalization to varied clinical data. Future work will focus on ensemble approaches, attention-enhanced architectures, and expanded datasets to further refine the robustness and reliability of these models for clinical deployment.

Acknowledgements

The authors would like to thank the Thailand Science Research and Innovation (grant ID TUFF 48/2568) and Sirindhorn International Institute of technology, Thammasat University for their financial support for this study.

References

[1] Wu L, Dahl AA, editors. Choroidal neovascularization (CNV). Updated 2023 Jan 5 [cited 2025 Jun 14]. In: Medscape Reference. Available from: https://emedicine.medscape.com/article/-1190818-overview.

[2] Spirn MJ, Feldman BH, Kozak A, Shah VA, Kim LA, Tripathy K, Khadamy J, de Carlo T, Lim JI, Ichlangod AM, Bose A. Optical coherence tomography. Eyewiki; 2024 [cited 2025 Apr 29]. Available from: https://eyewiki.org/Optical_Coherence_-Tomography.

[3] Ismail AM, El-Samie FEA, Omer AO, El-Sayed HS. Ensemble transfer learning networks for disease classification from retinal optical coherence tomography images. J Opt. 2024.

[4] Xi X, Meng X, Yang L, Nie X, Yang G, Chen H, Fan X, Yin Y, Chen X. Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior. Multimed Syst. 2019;25:95–102.

[5] Zhang Y, Ji Z, Wang Y, Niu S, Fan W, Yuan S, Chen Q. MPB-CNN: a multi-scale parallel branch CNN for choroidal neovascularization segmentation in SD-OCT images. OSA Contin. 2019;2(3):1011–27.

[6] Meng Q, Wang L, Wang T, Wang M, Zhu W, Shi F, Chen Z, Chen X. MF-Net: multi-scale information fusion network for CNV segmentation in retinal OCT images. Front Neurosci. 2021;15:743769.

[7] Diao S, Su J, Yang C, Zhu W, Xiang D, Chen X, Peng Q, Shi F. Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks. Biomed Signal Process Control. 2023;84:104810.

[8] Liu X, Liu Q, Zhang Y, Wang M, Tang J. TSSK-Net: Weakly supervised biomarker localization and segmentation with image-level annotation in retinal OCT images. Comput Biol Med. 2023;153:106467.

[9] Fu G, Cui W, Yang M, Zhang Y, Zhou Y, Sun Y, Xu X, Wang F, Lin H, Shen Y. Imaging-aware extreme interaction network for choroidal neovascularization segmentation in OCT images. Opt Commun. 2025;583:131615.

[10] Shen Y, Li J, Zhu W, Yu K, Wang M, Peng Y, Zhou Y, Guan L, Chen X. Graph attention U-Net for retinal layer surface detection and choroid neovascularization segmentation in OCT images. IEEE Trans Med Imaging. 2023;42(11):3140–54.

[11] Mori Y, Modi N. Enhancing OCT image classification for retinal disease diagnosis: a novel approach using squeeze-and-excitation ResNet. In: Proceedings of the 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC); 2024 Jul 27; Rajkot, India. Piscataway (NJ): IEEE; 2024. p. 940–5.

[12] Paul D, Tewari A, Ghosh S, Santosh K. OCTx: ensembled deep learning model to detect retinal disorders. In: Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); 2020 Jul 1; Rochester, MN, USA. Piscataway (NJ): IEEE; 2020. p. 526–31.

[13] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [Preprint]. arXiv:1505.04597 [cs.CV]. 2015 [cited 2025 Jun 13]. Available from: https://arxiv.org/abs/1505.04597.

[14] Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention U-Net: learning where to look for the pancreas [Preprint]. arXiv:1804.03999 [cs.CV]. 2018. Available from: https://arxiv.org/abs/1804.03999.

[15] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation [Preprint]. arXiv:1802.02611 [cs.CV]. 2018. Available from: https://arxiv.org/abs/1802.02611.

[16] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017. p. 2980–8.

[17] Hussain M, Khan MA, Khan W. Binary cross entropy with deep learning technique for image classification. Int J Adv Trends Comput Sci Eng. 2020;9(4):4561–7.

[18] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Cham: Springer; 2017. p. 240–8.

[19] Kermany DS, Goldbaum M, Cai W, Lewis MA, Xia H, Zhang K. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172(5):1122–31.e9.

[20] Steen D. Understanding the ROC curve and AUC [Internet]. Towards Data Science. 2025 Jul 22 [cited 2025 Jul 22]. Available from: https://towardsdatascience.com/understanding-the-roc-curve-and-auc-dd4f9a192ecb.

[21] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017 [cited 2025 Jul 22]. Available from: https://doi.org/10.1109/ICCV.2017.74.

[22] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS); 2017 [cited 2025 Jul 22]. Available from: https://doi.org/10.48550/arXiv.1705.07874.