Design and implementation of an algorithm for the automated analysis of Age-Related Macular Degeneration biomarkers on Optical Coherence Tomography

Álvaro Caballero Sastre[1], Alberto J. Beltrán Carrero[1], Javier Torresano Rodríguez[2], Esther Santos Vicente, María J. Aparicio Hernández-Lastras, María J. Ledesma-Carbayo[1,3], Juan-J. Gómez-Valverde[1,3]

[1] Biomedical Image Technologies (BIT), Universidad Politécnica de Madrid, España, alvaro.caballero.sastre@upm.es

[2] Ophthalmology Service of the Provincial Ophthalmic Institute, Hospital Universitario Gregorio Marañón, Madrid, España

[3] Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, España.

Abstract

Age-related macular degeneration (AMD) is the leading cause of irreversible vision loss in the elderly in developed countries. This study evaluates advanced deep learning models, including nnU-Net, U-Mamba, and MedNeXt, for segmenting key AMD biomarkers such as intraretinal fluid (IRF), subretinal fluid (SRF), and pigment epithelial detachment (PED) in OCT images from a real-world clinical dataset with 1,332 OCT B-scans. We compared 2D, 2.5D, and 3D architectures and assessed the impact of incorporating additional biomarkers such as tubular structures and epiretinal membranes. Our results show that 2D models consistently outperform 3D models, likely due to the nature of the radial scan of our dataset. Specifically, MedNeXt 2D achieved the highest performance with Dice scores of 0.789 for IRF, 0.691 for SRF, 0.802 for PED. These results suggest that leveraging advanced 2D deep learning models can significantly enhance the accuracy of AMD monitoring and treatment, offering substantial improvements in clinical practice.

1. Introduction

Age-related macular degeneration (AMD) is a chronic, progressive retinal disorder affecting the macula, the region of the retina with maximal visual acuity. It affects approximately 196 million people globally, representing a significant public health concern. AMD is the leading cause of severe vision impairment in older adults and is projected to impact around 288 million people worldwide by 2040 [1]. Exudative or neovascular AMD is an advanced stage of the disease, characterized by the formation of abnormal neovascularization and fluid accumulation between retinal layers. The detection of these fluids or biomarkers, typically including intraretinal fluid (IRF), subretinal fluid (SRF), and pigment epithelial detachment (PED), is critical for monitoring AMD progression [2].

Quantifying the progression of these biomarkers through their segmentation in optical coherence tomography (OCT) images provides a precise way to monitor disease activity and guide treatment decisions, particularly in patients undergoing anti-VEGF therapy. Accurate segmentation enables the reliable measurement of fluid volumes, which is crucial for evaluating the effectiveness of treatment and detecting relapses or progression [2].

The application of advanced deep learning segmentation algorithms, including nnU-Net [3], U-Mamba [4] and MedNeXt [5]; has demonstrated exceptional performance in the field of medical image segmentation. This study evaluates the efficacy of several algorithms, including those previously mentioned, in accurately detecting and quantifying AMD-related biomarkers within a real-world clinical dataset, thereby facilitating more effective monitoring of patients undergoing anti-VEGF therapy. Additionally, the study assesses the impact of segmenting additional biomarkers, such as tubular structures and epiretinal membranes, and explore the use of 2D, 2.5D and 3D architectures. By employing state-of-the-art deep learning techniques and optimizing configurations, the study aims to enhance the precision and reproducibility of biomarker evaluation, ultimately contributing to improve treatment outcomes in clinical practice.

2. Materials and Methods

2.1. Dataset Description

The dataset for this study consisted of 111 macular radial scan studies, comprising a total of 1,332 B-scans, obtained with the Topcon 3D OCT-1 Maestro in 90 patients at the Ophthalmology Department of the Hospital General Universitario Gregorio Marañón (HGUGM) in Madrid. The images were retrospectively collected from patients undergoing anti-VEGF treatment for neovascular AMD. This study was conducted following approval from the hospital's Ethics Committee, and informed consent was obtained from all participating patients.

All images in the study were annotated by two experts, labelling IRF, SRF, and PED. Additionally, a separate category labelled TERM—encompassing Tubular structures and EpiRetinal Membranes, which have distinct clinical features compared to the other biomarkers—was included to evaluate whether this additional label improved the segmentation of the three main biomarkers. An example of the manual labelling is presented in *Figure 1*.

ISBN: 978-84-09-67332-2

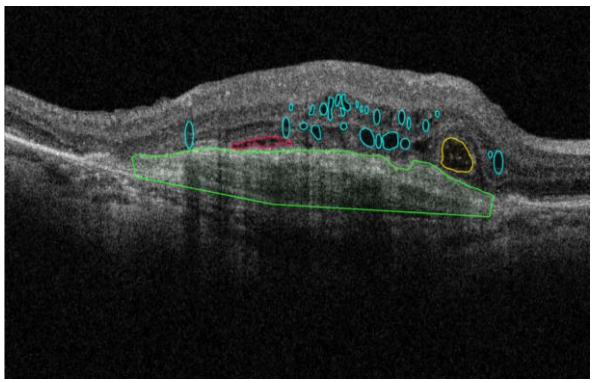

Figure 1: *OCT image illustrating expert annotations: intraretinal fluid (IRF) in blue, subretinal fluid (SRF) in red, pigment epithelial detachment (PED) in green, and TERM (tubular structures and epiretinal membranes) in orange.*

2.2. Data Preprocessing and Dataset Splitting

All B-scans were resized to 512x512 pixels and converted to NIfTI format. After these transformations, each NIfTI file containing the 12 B-scans was randomly divided into a training set (90 cases, 80%), a validation set (5 cases, 5%), and an independent test set (21 cases, 15%). The distribution of biomarkers was carefully balanced across the training, validation, and test sets to ensure an equitable representation of biomarkers in each set. Importantly, no patient was included in more than one set to ensure the integrity of the evaluation.

2.3. Performance Evaluation Metrics

For primary evaluation, we employed overlap-based metrics, including the Dice Score (DSC) and the Intersection over Union (IoU). These metrics are widely used for their effectiveness quantifying the similarity between predicted and ground truth segmentations. The Dice Score calculates the ratio of the intersection to the mean size of the two sets, providing a robust measure of segmentation accuracy. Additionally, the Intersection over Union (IoU), is particularly sensitive to small structures and thus well suited for detecting subtle retinal features in OCT images, such as IRF and subretinal fluid SRF.

Following the recommendations *Reinke et al.* [6], we also included distance-based metrics to assess the geometric similarity between predicted and true segmentations. The Hausdorff distance (HD) measures the maximum deviation between the predicted and true boundaries, thus assessing the overall shape accuracy. In addition, the Surface Mean Average (SMA) provides a more detailed evaluation by averaging the deviations across all boundary points, providing a more detailed view of the geometric accuracy of the segmentation.

We used the implementations of the metrics available in MONAI (Medical Open Network for AI), an open-source platform built on PyTorch, that provides a comprehensive suite of tools for medical image analysis [7].

2.4. Hardware and Software Requirements

Image annotation was performed using the Computer Vision Annotation Tool (CVAT) (https://www.cvat.ai/), and models were developed in Python using the PyTorch framework (version 2.4.1). The authors' implementations of nnU-Net, U-Mamba and MedNeXt were used for model training and evaluation. For the experiments we used high-performance workstations equipped with NVIDIA TITAN Xp and QUADRO P600 GPUs.

2.5. Deep Learning Architectures

2.5.1 nnU-Net

The nnU-Net, introduced by Isensee et al.[3], is an advanced, self-adapting deep learning architecture specifically designed for biomedical image segmentation. We employed nnU-Net V2 to assess the impact of incorporating the TERM biomarker on the segmentation performance of the rest of the biomarkers. We conducted two sets of training: one using only the IRF, SRF, and PED labels, and another incorporating the TERM label. Both 2D and 3D versions of nnU-Net were used. In the case of 3D, we assumed that the 12 B-scans in a radial scan together creates a 3D volume.

For these experiments, and subsequently for the rest, we configured the training with 100 epochs and employed the Dice Coefficient as the loss function.

2.5.2 2.5D SR-Net

The Slice Recovery Net (SR-Net) proposed by *Bermejo et al.* [8] is an architecture that leverages 3D contextual information for 2D segmentation. We used this model to evaluate the contribution of adjacent B-scans in our segmentation task. As shown in *Figure 2*, the SR-Net with 3 B-scans uses a dense architecture that decodes 2D segmentation maps from a 3D encoded volume through a specialized Slice-Recovery layer. The encoding path consists of convolutional dense and Enet-style blocks, while the decoding path employs transposed convolutions and skip-connections, ensuring that only higher resolution features relevant to the 2D slice are used to generate the final segmentation map.

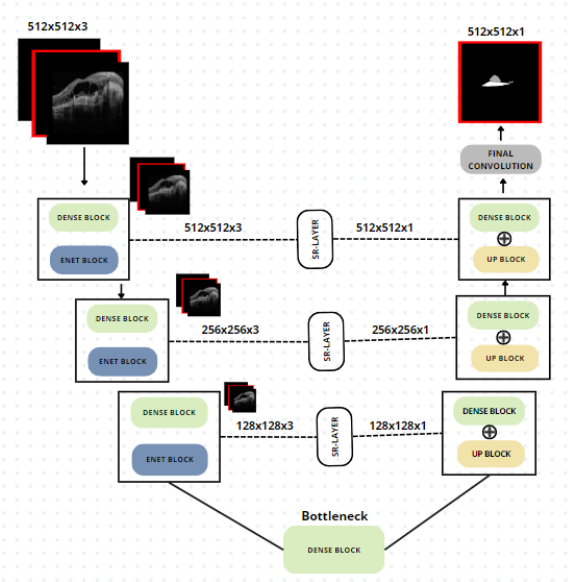

Figure 2: *SR-Net architecture for generating 2D segmentation maps from a 3D input volume.*

The model was tested using 3 and 5 adjacent B-scans around the central scan to evaluate the method's performance and assess the impact of varying the number of input images.

2.5.3 U-Mamba

U-Mamba is a state-of-the-art deep learning architecture designed to enhance segmentation tasks in biomedical imaging. A key feature of U-Mamba is its unique combination of convolutional layers, which aim to capture more intricate patterns in complex medical images [4]. Similar to nnU-Net, we tested U-Mamba using both 2D and 3D configurations, adapting the model during training.

What sets U-Mamba apart from nnU-Net is its specialized trainer, which integrates adaptive mechanisms to improve segmentation accuracy by better aligning with the characteristics of the medical imaging data. This enhanced trainer reduces artifacts and enhances performance by focusing on critical anatomical features. Consequently, U-Mamba offers a more refined and effective approach, surpassing the traditional nnU-Net trainer in handling challenging segmentation tasks.

The net was implemented in its default mode, however the number of epochs was set to a 100 and the loss function employed was the same as in the nnU-Net approach, the Dice Coefficient Loss.

2.5.4 MedNeXt

The final architecture we evaluated was MedNeXt [5], which, as noted by *Isensee et al.* [9] in their recent review of nnU-Net, achieved top performance on several public image segmentation datasets.

A key strength of MedNeXt is its specialized trainer, designed to handle complex biomedical imaging tasks. This trainer incorporates axial attention layers that efficiently capture fine details while providing both global and local context. This design enhances the model's adaptability to diverse data and reduces the occurrence of artifacts, making it particularly suitable for high-precision segmentation tasks. We prepared experiments to evaluate the architecture considering 2D and 3D trainers and employed the default parameters of the Small-MedNeXt configuration (MedNeXt-S) which includes the best balance between number of parameters and computing hours, however we modified the epoch and the loss to match the rest of the experiments (100 epochs and Dice Coefficient Loss).

3. Results and discussion

Therefore, we present in the following section the results of all experiments performed:

3.1. 2D/3D nn-Unet with/without TERM

Table 1 presents the results of the inference on the test set using the nnU-Net V2 implementation. The data indicate that the 2D architecture outperforms the 3D architecture. Additionally, the inclusion of the TERM label positively impacts the performance across all metrics for both architectures. All metrics were calculated as the mean across all labels. For subsequent experiments, we consistently included IRF, SRF, PED, and TERM in our training and evaluations.

Method	DSC	IoU	HD	SMA
nnU-Net 2D (*)	**0.745±0.247**	**0.687±0.252**	**35.833±24.623**	**3.624±7.590**
nnU-Net 3D (*)	0.669± 0.227	0.607± 0.233	48.243± 18.983	6.895± 4.576
nnU-Net 2D	0.703±0.213	0.628±0.220	42.715±18.996	7.456±3.043
nnU-Net 3D	0.655±0.223	0.571±0.220	46.808±25.083	5.913±5.750

Table 1. *nnU-Net performance results for 2D and 3D inputs. (*) Includes segmentation of IRF, SRF, PED, and TERM. The best metric for each category is highlighted in bold.*

3.2. 2.3D SR-Net

Table 2 presents the results of experiments using the 2.5D SR-Net architecture with 3 and 5 adjacent B-scans. The data reveal that the 2.5D approach is less effective for our problem compared to the nnU-Net (see *Table 1*). Notably, including more adjacent B-scans led to decreased performance, with the model using 3 B-scans outperforming the model with 5 B-scans. This suggests that a greater number of B-scans may not necessarily enhance the model's accuracy in this context. As in 3.1, all metrics were calculated as the mean across all labels.

	DSC	IoU	HD	SMA
3 B-Scans	**0.508 ± 0.104**	**0.499 ± 0.111**	**66.039 ± 13. 281**	**13.911± 7.204**
5 B-Scans	0.206 ± 0. 210	0.191 ± 0.2	92.449 ± 20.971	43.070± 10.03

Table 2. *SR-Net performance results for 3 and 5 B-scans. The best metric for each category is highlighted in bold.*

3.3. U-Mamba 2D/3D and MedNeXt 2D/3D

Finally, *Table 3* presents the results of the inference from the 2D/3D training using the U-Mamba and MedNeXt methods.

Method	DSC	IoU	HD	SMA
U-Mamba 2D	0.748±0.243	0.689±0.249	39.174±33.024	3.134±6.457
U-Mamba 3D	0.680±0.265	0.617±0.268	47.482±45.571	6.905±12.952
MedNeXt 2D	**0.751±0.257**	**0.699±0.229**	**36.462±34.161**	**2.948±6.299**
MedNeXt 3D	0.729±0.219	0.661±0.260	41.023±47.006	4.574±9.111

Table 3. *Performance results for U-mamba 2D/3D and MedNeXt 2D/3D. The best metric for each category is highlighted in bold.*

The results indicate that all 2D models outperformed their 3D counterparts in segmenting the evaluated biomarkers. Specifically, MedNeXt 2D achieved the highest performance across all biomarkers, followed by U-Mamba 2D and nnU-Net 2D. This performance difference may be due to the inherent complexity of processing 3D inputs and the nature of the radial scans. Radial scans consist of a series of 2D slices rather than a cohesive 3D volume. 3D models are designed to handle and integrate full volumetric data, which adds computational complexity and may not align well with the structure of radial scans. In contrast, 2D models are better suited to exploit the high-resolution

ISBN: 978-84-09-67332-2

details and patterns in each individual slice, leading to more effective segmentation for radial scans.

To provide a more detailed analysis of model performance, *Table 4* presents the Dice score for each label. It highlights that MedNeXt achieved the best performance for IRF and SRF, nnU-Net 2D excelled in segmenting PED, and U-Mamba 2D performed best for TERM.

Method	IRF	SRF	PED	TERM
nn-UNet 2D	0.765±0.17	0.679±0.28	**0.831±0.19**	0.706±0.35
U-Mamba 2D	0.787±0.18	0.685±0.27	0.795±0.2	**0.727±0.32**
MedNeXt 2D	**0.789±0.21**	**0.691±0.29**	0.802±0.22	0.722±0.31

Table 4. *Dice Score results of developed methods for each biomarker. The best metric for each label is highlighted in bold*

Figure 3 shows the segmentation of AMD biomarkers of the best model (2D MedNeXt), stressing how accurate the model captures the fine details of these biomarkers on a representative B-scan with a wide range of structures and fluid sizes.

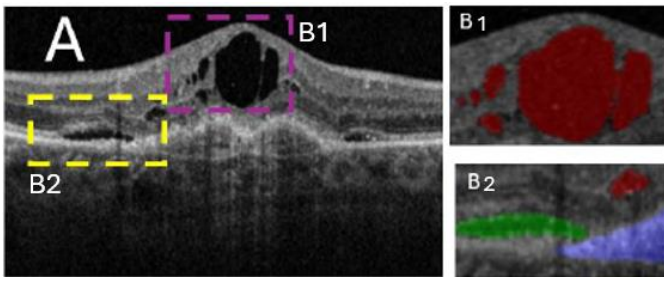

Figure 3. *Example of output segmentation using MedNeXt 2D. (A) OCT B-scan with two marked regions of interest: (B1) showing IRF in red, and (B2) highlighting IRF in red, SRF in green, and PED in blue.*

4. Conclusions

Age-related macular degeneration (AMD) is a leading cause of vision impairment. This study assesses deep learning models, including nnU-Net, U-Mamba, and MedNeXt, for segmenting AMD biomarkers—such as intraretinal fluid (IRF), subretinal fluid (SRF), and pigment epithelial detachment (PED)—in OCT images. We evaluated the impact of incorporating additional biomarkers and compared 2D, 2.5D, and 3D architectures. Results demonstrate that 2D models consistently outperform 3D models, likely due to radial scans being organized as 2D slices. Among the models, MedNeXt 2D achieved the highest performance, followed by U-Mamba 2D and nnU-Net 2D. This study's findings promise to enhance the accuracy of AMD monitoring and treatment by leveraging advanced deep learning techniques for precise biomarker segmentation.

Acknowledgments

This research is part of the grant TED2021-131951B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the the "European Union NextGenerationEU/PRTR". The authors gratefully acknowledge the Universidad Politécnica de Madrid (www.upm.es) for providing computing resources on Magerit Supercomputer.

References

[1] W. L. Wong *et al.*, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," *Lancet Glob Health*, vol. 2, no. 2, Feb. 2014, doi: 10.1016/S2214-109X(13)70145-1.

[2] M. Fleckenstein, S. Schmitz-Valckenberg, and U. Chakravarthy, "Age-Related Macular Degeneration: A Review," *JAMA*, vol. 331, no. 2, pp.147–157,Jan.2024, doi:10.1001/JAMA.2023.26074.

[3] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods 2020 18:2*, vol. 18, no. 2, pp. 203–211, Dec. 2020, doi: 10.1038/s41592-020-01008-z.

[4] J. Ma, F. Li, and B. Wang, "U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation," Jan. 2024, Accessed: Sep. 04, 2024. [Online]. Available: https://github.com/bowang-lab/U-Mamba

[5] S. Roy *et al.*, "MedNeXt: Transformer-Driven Scaling of ConvNets for Medical Image Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14223 LNCS, pp. 405–415, 2023,doi:10.1007/978-3-031-43901-39/TABLES/2.

[6] A. Reinke *et al.*, "Understanding metric-related pitfalls in image analysis validation," *Nature Methods 2024 21:2*, vol. 21, no. 2, pp. 182–194, Feb. 2024, doi: 10.1038/s41592-023-02150-0.

[7] "Project MONAI · GitHub." Accessed: Sep. 03, 2024.[Online].Available: https://github.com/Project-MONAI

[8] D. Bermejo-Peláez, Y. Okajima, G. R. Washko, M. J. Ledesma-Carbayo, R. San, and J. Estépar, "A SR-NET 3D-TO-2D architecture for parasepta emphysema segmentation." In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)(pp.303-306).IEEE.

[9] F. Isensee *et al.*, "nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation," Apr. 2024, Accessed: Sep. 05, 2024. [Online]. Available: https://arxiv.org/abs/2404.09556v2