



Outer retinal band segmentation in healthy subjects: comparative study between human grading and deep convolutional neural networks

João Duarte Afonso^{1^}, Pedro Camacho^{2^}, Bruno Pereira^{2,3^}, João Pedro Marques^{4^}, Daniel Simões Lopes^{1^}, Diogo Cabral^{5^}

¹ITTL/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal; ²H&TRC - Health & Technology Research Center, ESTeSL - Lisbon School of Health, Polytechnic Institute of Lisbon, Lisbon, Portugal; ³IRL - Instituto de Retina de Lisboa, Lisbon, Portugal; ⁴Department of Ophthalmology, Hospitais da Universidade de Coimbra, Local Health Unit of Coimbra, Coimbra, Portugal; ⁵Hospital Garcia de Orta, Almada-Seixal Local Health Unit, Almada, Portugal

Contributions: (I) Conception and design: DS Lopes, D Cabral; (II) Administrative support: JP Marques, D Cabral; (III) Provision of study materials or patients: P Camacho, B Pereira; (IV) Collection and assembly of data: JD Afonso; (V) Data analysis and interpretation: JD Afonso, P Camacho; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Diogo Cabral, MD, PhD. Centro de Responsabilidade Integrado de Oftalmologia, Unidade Local de Saúde Almada-Seixal, Hospital Garcia de Orta, Avenida Torrado da Silva, 2805-267 Almada, Portugal. Email: diogo.cabral@nms.unl.pt.

Background: Accurate identification of outer retinal bands (ORBs) on spectral-domain optical coherence tomography (SD-OCT) is clinically important for the assessment and monitoring of several retinal diseases. However, manual segmentation of these ORBs is time-consuming and subject to considerable intra- and inter-observer variability. Deep convolutional neural networks (dCNNs) offer a promising approach for achieving more consistent and reproducible segmentation of retinal microstructures, yet their performance for individual ORBs and their dependence on grader reliability remain insufficiently characterized. The aim of this study was to compare manual grading and different dCNN architectures for the segmentation of individual ORBs in healthy subjects, while assessing the reproducibility of human annotations and the performance of automated methods.

Methods: A cross-sectional observational study was conducted using foveal-centered SD-OCT scans obtained from healthy participants. Manual segmentation of the second hyperreflective band (Band 2), the third and fourth hyperreflective bands (Band 3+4) and photoreceptor outer segments (POS) was performed by two trained graders using onboard OCT software. Reproducibility was assessed using the intraclass and interclass correlation coefficient (ICC). Human annotations were used to train different dCNNs pretrained with Sparse masked modelling on unlabelled data. The performance of the dCNNs was evaluated using dice score and pixel difference analyses.

Results: One hundred eyes of healthy subjects were included. The intraclass correlation values were 0.737 (0.689, 0.779), 0.830 (0.795, 0.860) and 0.878 (0.852, 0.900), and the ICCs were 0.104 (-0.049, 0.253), 0.810 (0.740, 0.862) and 0.752 (0.678, 0.812) for Band 2, Band 3+4 and POS, respectively. From the dCNN, DRUNET showed the best performance: 0.916 for Band 2, 0.924 for Band 3+4 and 0.910 for POS. Pixel difference analysis between DRUNET and grader 1 showed a deviation of 0.89 ± 0.44 .

Conclusions: Manual segmentation of ORBs demonstrated moderate to high repeatability, whereas Band 2 segmentation showed low reproducibility. DRUNET outperformed other dCNN architectures, showing great potential to reduce inter-observer variability.

[^] ORCID: João Duarte Afonso, 0009-0002-2301-6542; Pedro Camacho, 0000-0002-2986-5652; Bruno Pereira, 0000-0003-4955-1770; João Pedro Marques, 0000-0002-1014-0483; Daniel Simões Lopes, 0000-0003-0917-9396; Diogo Cabral, 0000-0003-1968-3561.

Keywords: Optical coherence tomography (OCT); deep learning; automatic segmentation; intraclass correlation

Submitted Oct 01, 2025. Accepted for publication Dec 30, 2025. Published online Apr 13, 2026.

doi: 10.21037/qims-2025-aw-2093

View this article at: <https://dx.doi.org/10.21037/qims-2025-aw-2093>

Introduction

Photoreceptors (PhR) are highly organized and complex light-sensitive cells that are subdivided into defined compartments with functionally related organelles (1). On spectral-domain optical coherence tomography (SD-OCT), these structures can be seen as four distinctive hyper-reflective OCT bands. The outer segments of the PhR (POS) are opposed to the retinal pigment epithelium (RPE) and form an interdigitation zone (IZ) in which the apical processes of the microvilli of the RPE envelop the POS (2). This region separates the hyperreflective Band 3 (cone phagosomes, PhaZ) from Band 2, formed by the inner segment (IS) of the ellipsoids (EZ) (3). Separating Band 2 from the outermost hyperreflective Band 1 (external limiting membrane, ELM) is the cone myoids (3) (*Figure 1*). This microanatomy follows an intricate system that has neurophysiological, optical and metabolic features, enabling a successful visual cycle. A detailed understanding and identification of normal and pathologic morphologic patterns of this structure has important implications for the diagnosis, treatment and progression of various retinal diseases (2,4,5). Careful assessment of these bands is clinically relevant in degenerative conditions such as age-related macular degeneration, where alterations in the ellipsoid zone (EZ) and IZ may reflect disease progression and treatment response (6,7). Similarly, disruptions in the outer retinal bands (ORBs) are key imaging features in inherited retinal dystrophies (8) and diabetic macular edema, and have increasingly been used as structural endpoints in clinical trials (9). In clinical practice, this assessment is based on OCT analysis. However, detailed measurement of these layers requires a high level of clinical expertise and can be very time consuming (4), showing variability when measured, especially in bands with thickness similar to the resolution of commercial OCT devices (10).

Deep learning networks (DLNs) have been used to support the identification, extraction and analysis of features difficult for human physicians to interpret (5,11,12). Recent reviews have further highlighted the rapid progress of DLN methods in ophthalmic imaging and OCT-based analysis

(13,14). With promising results, these techniques can also provide consistent and reproducible measurements less susceptible to subjective differences among evaluators, as shown in different medical areas (15,16). The application of these techniques to OCT images enabled the automatic identification of specific retinal lesions, such as macular edema (17,18) and choroidal neovascularization (19), as well as the differentiation between ORBs and other anatomical structures, including the choroid (20) and cornea (21). Approaches using image filtering (22), pixel-profiling and edge enhancing (23), and OCT-specific deep convolutional neural network (dCNN) (24,25) achieved impressive results in the segmentation of different structures in the retina. Among the CNNs specifically modified for retinal layer segmentations, Two-Dimensional U-Net are used to directly segment retinal layers from individual B-scans, using a pixel-wise classification (26,27). Others use shape-based regressions to return more smooth segmentations (28), attention mechanisms that focus on relevant features (29) and also Cascaded U-Nets in cases of more severe pathologies (30). However, most studies prevent the individual characterization of ORBs given the wider retinal regions evaluated and integrated in each deep learning dataset (23,31-33), a feature that already showed clinical relevance for disease differentiation (34). Our study utilizes individual retinal labelling of each ORB, and the ability of semantic segmentation in dCNN, given their capability to classify each pixel individually (35), for the accurate characterization of relevant retinal layers.

Pretrained networks have also been used in several dCNNs, especially in medical imaging as a solution to deal with limited labelled data (28,36), improving learning of generic structures (37). Sparse masked modeling (SparK) (38), a bidirectional encoder representation from transformers (BERT)-style self-supervised learning framework, that allows unmasked patches to be treated as sparse voxels and use sparse convolution to encode them. While making it suited for any convolutional network (Convnet), it also reports superior learning of visual representations when compared to other self-supervised transformers (39).

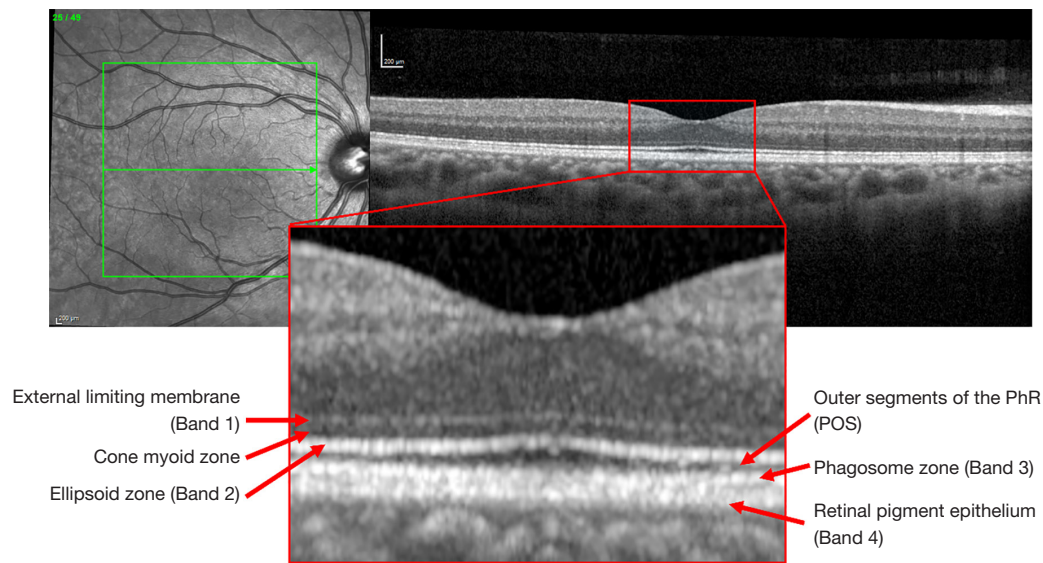


Figure 1 Visualization of the hyperreflective bands. Example of a fovea centered B-scan of a normal patient. A central patch highlighted in red shows the hyperreflective bands numbering and their locations. PhR, photoreceptors; POS, photoreceptor outer segments.

Hereby, we evaluated the performance of different approaches for the segmentation of ORBs, i.e., manual segmentation *vs.* dCNN architectures, by evaluating dCNN outputs and reproducibility of the manual annotations. We present this article in accordance with the TRIPOD+AI reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-2025-aw-2093/rc>).

Methods

The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of the Lisbon School of Health (ESTeSL) (CE-ESTeSL-No.59-2024). Informed consent was taken from all the patients. A cross-sectional, observational study based on foveal-centered SD-OCT B-scans of healthy subjects—was conducted at the retina clinic of ULS Almada-Seixal, Portugal from September to December 2024. Participation was voluntary, with no associated costs, payments, or rewards. Participants were informed that they could discontinue their participation at any time without any penalties or financial implications.

Anonymized SD-OCT scans (Spectralis; Heidelberg Engineering, Heidelberg, Germany) were obtained and retrospectively analyzed from a study dataset (IPL/IDI&CA2024/INSYDE_AMD_ESTeSL). Imaging

consisted of high-resolution macular volume scans ($20^\circ \times 20^\circ$, 49 B-scans, 7 frames per scan, each containing 1,024 A-scans per B-scan, with a depth resolution of $3.9 \mu\text{m}$). For each macular volume scan, nine horizontal B-scans were analyzed—one centered on the fovea and four located superiorly and inferiorly, each spaced $125 \mu\text{m}$ apart from the adjacent scan. Each patient was previously evaluated by an ophthalmologist.

As a recruitment strategy, following comprehensive ophthalmological evaluation, participants were selected among patients attending the clinic for cataract assessment (pre- or postoperative) or those presenting unilateral surgical conditions such as macular hole or epiretinal membrane, provided that the fellow eye showed no signs of retinal pathology. Only eyes confirmed as structurally healthy were included; additionally, the ophthalmologist confirmed that all images met the quality standards required for clinical assessment.

Exclusion criteria included any scan in which the ORBs were insufficiently distinguishable to allow a reliable definition of the hyperreflective layers, SD-OCT B-scans with an inclination greater than 10° —measured as the angle between the line connecting the outermost retinal points in each scan and the horizontal plane—and patients with myopia or hyperopia exceeding 3 dioptres.

Fifty-five subjects (50% male) contributed to this study. The mean [standard deviation (SD)] age at baseline was

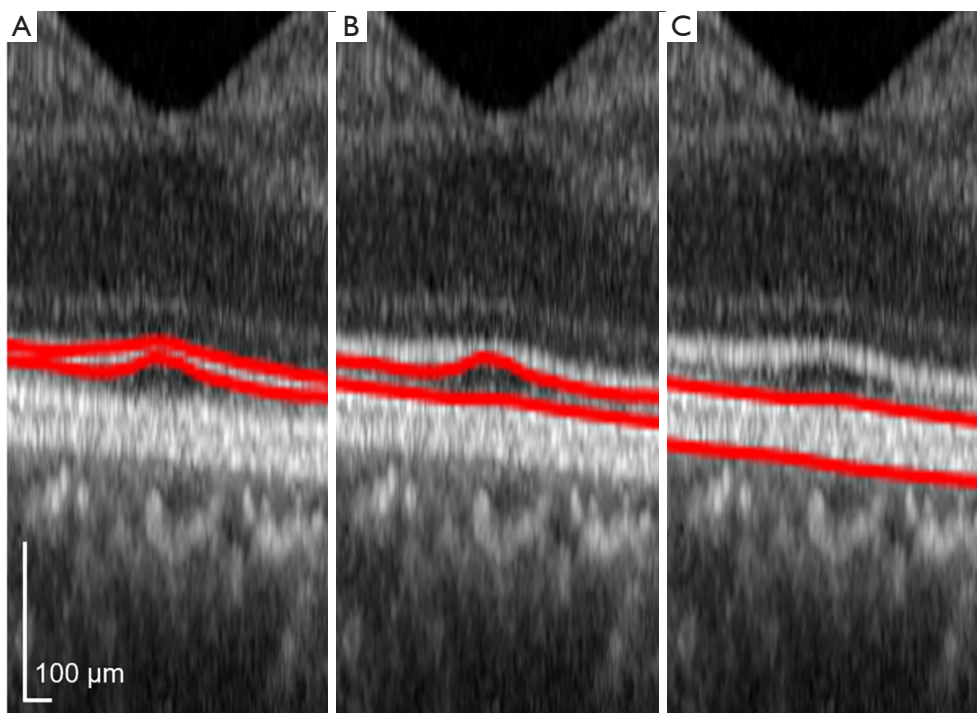


Figure 2 Manual segmentation of the three regions of interest. Example of a healthy B-OCT cropped in the center fovea with manual annotations by one of the two graders. (A) The upper and lower red lines correspond to the inner and outer boundaries of Band 2, respectively. (B) The upper and lower red lines correspond to the inner and outer boundaries of the POS region, respectively. (C) The upper and lower red lines correspond to the inner boundary of Band 3 and the outer boundary of Band 4, respectively. OCT, optical coherence tomography; POS, photoreceptor outer segments.

63.9 (12.2) years, the mean (SD) best-corrected visual acuity (BCVA) was 79.6 (10.5) letters. Ten eyes were excluded due to insufficient differentiation between the EZ and IZ. A total of 900 OCT B-scans from 100 eyes were included in the dataset.

B-scans were analysed by two experienced graders in image segmentation that, using the investigational software (SPX 1902; Heidelberg Engineering) onboard caliper and at 400% magnification, segmented the internal and external boundaries of Band 2, the photoreceptor outer segments (POS), and Band 3+4 at the whole length of the OCT B-scan, as seen in *Figure 2*. Band 2 was defined as the EZ of the PhR (3). Band 3 and Band 4 were defined as respectively the contact cylinder between the RPE apical processes and the external portion of the cone OS, and the RPE (3). POS were defined as the space between the external boundary of Band 2 and the internal boundary of Band 3.

Manual segmentation of hyperreflective bands was performed twice, where the second measurement was performed at least two days later and masked from the

first one. Each labelled image was further analysed to automatically identify the foveal center (40) and to compute the thickness value of each band at 7 specific coordinates at the x-axis: foveal center, and at 500, 750 and 1,000 μm nasal and temporal to the foveal center (*Figure 3*), as previously described (34). The thickness value of each band at the above-mentioned coordinates was exported as a .csv file for further analysis.

In this study, all analyses were performed at the image level. Each OCT B-scan was treated as an independent observation because the objective was to assess segmentation performance for each individual image, both for human graders and for the deep learning models. The performance metrics used—interclass correlation coefficient (ICC), mean absolute error (MAE), pixel differences and Dice scores—quantify agreement between segmentation methods within the same image and do not involve subject-level statistical inference. Therefore, no correction for within-subject clustering (multiple images from the same eye or two eyes from the same participant) was required.

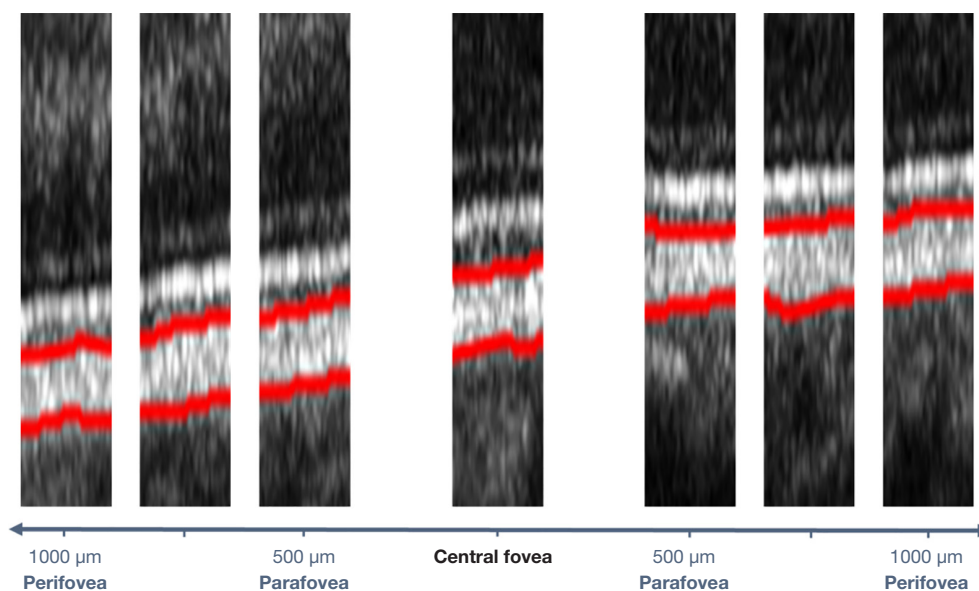


Figure 3 Sub images used to measure the repeatability in different locations of the fovea. Representation of the fovea detection algorithm in a B-OCT image, and the respective cropped images corresponding to the central fovea, parafovea, perifovea and the intermediate region between the parafovea and perifovea. OCT, optical coherence tomography.

The difference between the two measurements of the same grader evaluated the intraclass correlation, or repeatability, while the difference between each independent grader evaluated the interclass correlation, or reproducibility (41). B-scans segmented by grader 1 were used as the ‘ground truth’ to test the following dCNNs from the literature: U-Net (42), U-Net++ (43), SegResNet (44) and DRUNET (45).

Each OCT B-Scan was cropped at the center of the fovea to obtain two images of a standardized 256×256 pixels. Dataset augmentation was performed through the following techniques: contrast enhancement, gaussian noise introduction, and image flipping. Images were then randomly assigned at the patient level to the training, validation, and test sets, with 70%, 15%, and 15% of the images, respectively, ensuring that scans from the same patient were not included in more than one subset.

Previously to train each individual dCNN, a SPARK framework with unlabelled data from three different public OCT databases (46-48), approved by the corresponding ethical review board, was trained for 50 epochs. The output from trained autoencoder with 53,000 OCT B-scans was dimensioned so the weights would be introduced as the first encoder layer of each individual dCNN. After 5 initial epochs, the weights were unfrozen, and the fine tuning with the labelled images was continued.

Each dCNN was run for 50 epochs, or until convergence. Each model was trained on a weighted loss with both Dice loss and binary cross entropy loss (39). The Adam Optimizer was utilized as an optimization technique and regularization techniques included early stopping and dropout in every hidden layer. Model’s performance was measured with four different performance metrics: Dice score, precision, recall, and pixel-accuracy (49). Deep learning algorithms and statistical analysis were performed in Python (version 3.11.5) using PyTorch framework, and training was performed in an Intel(R) Core (TM) i7-6700K CPU, NVIDIA Titan X GPU with a 16GB DIMM RAM. Training each model took between 6 and 8 hours and inference time on our test set was only a few hundred milliseconds, as compared with algorithms reported in the study by Tian *et al.* (50), ranging from 28 to 152 seconds. The framework proposed for this project is illustrated in *Figure 4*. The full code for this project, excluding the OCT data, is available in the GitHub repository (40).

Results

A total of 100 eyes were evaluated and annotated by each grader. The intraclass correlation and interclass correlation values are summarized at *Tables 1,2*, respectively. Intraclass correlation values showed good repeatability for the

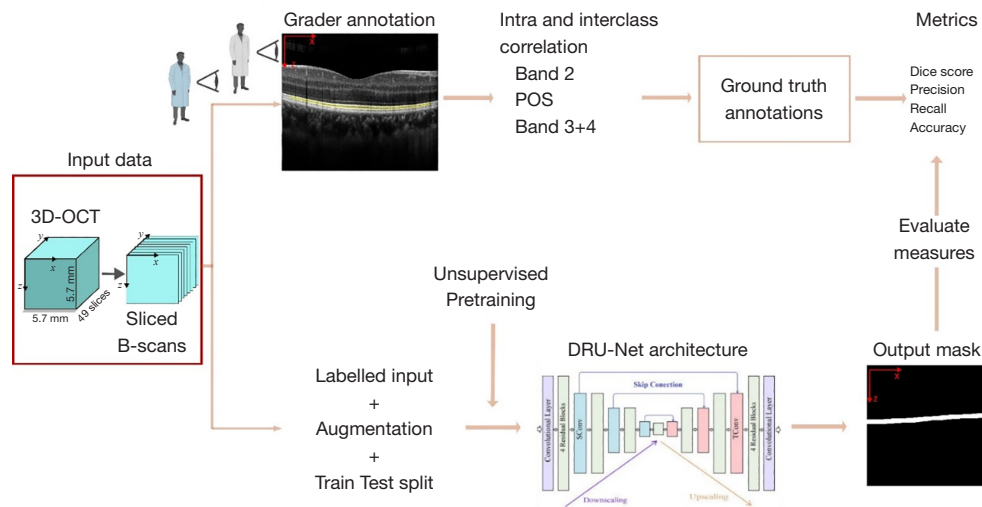


Figure 4 Pipeline of the research project. After acquiring the images from the Spectralis SD-OCT machine (input data square), two methods were tested to segment three different retinal structures, one involving manual annotations performed by two different graders (top row), and one with a DL approach, that in the end was evaluated against the ground truth annotations for various evaluation metrics. 3D-OCT, three-dimensional optical coherence tomography; DL, deep learning; POS, photoreceptor outer segments; SD-OCT, spectral-domain optical coherence tomography.

Table 1 Intraclass correlation and average pixel differences for both graders in the three regions of interest

Grader	Region	ICC	95% confidence interval		Avg Diff (pixels)
			Lower bound	Upper bound	
1	Band 2	0.737	0.689	0.779	1.384
	Band 3+4	0.830	0.795	0.860	1.812
	POS	0.878	0.852	0.900	1.444
2	Band 2	0.565	0.451	0.660	1.547
	Band 3+4	0.888	0.844	0.922	2.906
	POS	0.789	0.724	0.841	1.390
Average	Band 2	0.651	0.570	0.720	1.466
	Band 3+4	0.859	0.820	0.891	2.359
	POS	0.834	0.788	0.896	1.417

Repeatability of each grader to perform 3 different structure segmentations with its respective confidence intervals and average pixel differences between measurements. Each second measurement was performed at least two days later and masked to the first one. Avg Diff, average difference; Band 2, ellipsoid zone; Band 3+4, interdigitation zone of photoreceptors (IZP, Band 3) and retinal pigment epithelium (RPE, Band 4); ICC, intraclass correlation coefficient; POS, photoreceptor outer segments.

identification of Band 3+4 and POS, while identification of Band 2 displayed moderate values. Likewise, interclass correlation values showed good reproducibility for the identification of Band 3+4 and POS, while identification of Band 2 displayed poor reliability, according to the

evaluation metrics used by Koo and Li (51).

dCNNs performances are summarized in *Table 3*. Dice scores varied between 83.8% and 93.0%, corresponding to a mean deviation of 1.02 ± 0.49 pixels compared to grader 1 manual segmentation. DRUNET architecture achieved the

Table 2 Interclass correlation and average pixel differences between the measurements of the two graders

Grader	Region	Interclass correlation	95% confidence interval		Avg Diff (pixels)
			Lower bound	Upper bound	
Grader 1/grader 2	Band 2	0.104	-0.049	0.253	3.818
	Band 3+4	0.810	0.740	0.862	2.832
	POS	0.752	0.678	0.812	2.992

Reproducibility for both graders to perform 3 different structure segmentations with its respective confidence intervals and average pixel differences between measurements. Band 2, ellipsoid zone; Band 3+4, interdigitation zone of photoreceptors (IZP, Band 3) and retinal pigment epithelium (RPE, Band 4). Avg Diff, average difference; POS, photoreceptor outer segments.

Table 3 Segmentation results for the three different regions of interest

Model	Region	Dice	Accuracy	Precision	Recall	Avg Diff (pixels)
UNET	Band 2	0.896	0.995	0.902	0.897	0.745
	Band 3+4	0.930 [†]	0.993	0.927	0.935	1.251 [†]
	POS	0.883	0.995	0.904	0.867	0.762
UNET++	Band 2	0.902	0.996	0.900	0.913	0.660
	Band 3+4	0.873	0.990	0.960	0.834	2.082
	POS	0.838	0.994	0.934	0.798	0.970
DRUNET	Band 2	0.916 [†]	0.996	0.884	0.935	0.654 [†]
	Band 3+4	0.924	0.993	0.922	0.928	1.397
	POS	0.910 [†]	0.996	0.891	0.913	0.610 [†]
SegResNet	Band 2	0.904	0.995	0.891	0.926	0.672
	Band 3+4	0.908	0.991	0.912	0.906	1.771
	POS	0.890	0.995	0.899	0.885	0.705

Each model was trained and evaluated with the annotations from grader 1 for each region and tested with the following performance metrics. [†], best Dice score and Avg Diff for each region. Band 2, ellipsoid zone; Band 3+4, interdigitation zone of photoreceptors (IZP, Band 3) and retinal pigment epithelium (RPE, Band 4). Avg Diff, average difference; POS, photoreceptor outer segments.

best results (*Figure 5*).

Discussion

In this study, we addressed the challenge of OCT outer bands segmentation in healthy subjects. We observed good intra-grader repeatability for the segmentation of both the POS and Band 3+4, whereas Band 2 showed only moderate intra-grader agreement and poor agreement between different graders. Different dCNNs were able to segment all bands with high accuracy, with UNET network achieving the best results.

Some studies have shown consistent manual segmentation of 12 layers of the retina in foveal centered

OCT B-scans (10,52). Ghorbel *et al.* showed intraclass mean squared error of 2.487 ± 0.71 , on the segmentation of the outer segments + RPE, while other studies reported ICC ranging from 0.68 to 0.99, in the regions relevant to our study, reporting lower values for the segmentation of Band 2, agreeing with our observations.

The second hyperreflective band in OCT has been considered one of the most challenging structures to accurately segment (53). The normalization of pupil entry position and enhanced depth imaging (EDI) have great impact in EZ reflectivity. In the same manner, the presentation of OCT images in either logarithmic or linear scale and the lack of consensus on a standardized normalization method may impact the comparison of EZ

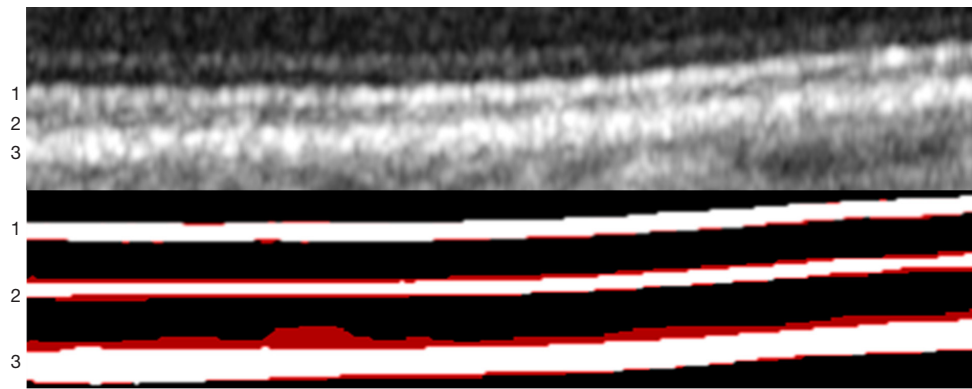


Figure 5 Comparison between a right-cropped B-OCT image (upper) and the binary annotation masks overlaid with the output of the DRUNET network for the segmentation of three retinal regions (lower), shown in red. The upper image shows a B-OCT scan from a healthy subject that was not included in the training dataset. The lower image shows the three regions of interest segmented in white, as annotated by one grader, with the corresponding DRUNET predictions overlaid in red. The numbered labels indicate the segmented layers: 1, Band 2; 2, POS; 3, Band 3+4. The segmented layers in the lower image are intentionally displayed with increased spacing for visualization purposes and do not represent their true anatomical distances in the B-OCT scan. OCT, optical coherence tomography; POS, photoreceptor outer segments.

reflectance across different studies (54). Importantly, a low sampling rate in the transverse direction—that is, a reduced number of A-scans per B-scan for a given field of view—can diminish lateral resolution and hinder the clear differentiation of outer retinal morphology (55). Beyond challenges in image acquisition, the blurred transitions between the myoid zone and POS have also been identified in the literature as factors that make the segmentation of the EZ particularly challenging.

Additionally, our analysis confirmed that the variability observed for Band 2 was largely driven by grader-specific differences. Although both graders were trained and experienced, grader 2 showed higher intra-observer variability and a consistent positional bias when delineating this ultrathin layer, which contributed significantly to the reduced inter-grader ICC. Even for grader 1 (considered ground truth), Band 2 presented the lowest ICC, highlighting that small discrepancies in marking its inner and outer boundaries have a disproportionate impact due to its minimal thickness.

Topographic analysis showed that manual segmentation of outer bands in the foveal pit achieved the best intra and interclass correlation values in comparison with the segmentation of outer bands in perifoveal locations (see Table S1). This observation aligns with previous studies evaluating outer bands in healthy subjects. Histology studies have shown different topographic distribution of light sensitive PhR cells in the macula, the central fovea

exclusively with cones, that have longer POS, and the perifoveal with a predominance of rod cells, with shorter POS (56). We conjecture that the physiologic elongation of POS at the central foveal facilitates the segmentation of this hyporeflective band in OCT at this location. As mentioned by other studies, the absence of inner retinal layers at the foveal center is associated with less light scattering and projection of retinal vessels at the outer retina (57,58). The ensemble of these observations might account for a more precise identification of POS and therefore the upper and lower boundaries of Band 2 and Band 3, with higher reproducibility values at the foveal pit compared with perifoveal measurements.

Apart from the performance metrics evaluated for each dCNN, these models were also compared with reported work on automatic segmentation of larger retinal areas, where, for example, a differentiation between the EZ and the POS was not performed (59), underscoring the importance of the individual segmentation of these regions. Other papers face the same problem, where, through U-Net-like structures, Dice scores of 95% were achieved, but failed to distinguish between the sublayers in their studies (60,61).

Pekala *et al.* (24) applied fully convolutional networks (FCNs) with Gaussian processes for post-processing and reported average pixel differences of 1.18 pixels between the outer plexiform layer (OPL) and the outer nuclear layer (ONL) compared to manual annotators. Fang *et al.* (62)

achieved below 2 pixels variations for non-exudative AMD patients, in the same regions analyzed by Roy *et al.* (31) that further reported Dice scores of 0.92 and 0.90 when segmenting between the first and second hyperreflective bands and the POS + Band 3 + Band 4 regions, respectively. Rivas Vázquez *et al.* (28) also segmented nine retinal layers and the optic disc, achieving Dice scores of 83.6%. In our study, we observed smaller pixel differences and maintained similar Dice scores, even for smaller segmentation regions.

DRUNET is a dCNN widely used for segmentation of retinal structures. Kugelmann *et al.* (20) employed DRUNET for choroidal segmentation and showed MAE under 0.5 pixels in the delimitation of the internal limiting membrane (ILM) and the RPE. Our results showed comparable performances across all networks evaluated, suggesting that the high quality of the input data allowed all architectures to converge towards a similar performance ceiling. This highlights the importance of data quality in optimizing segmentation performance and suggests that simpler architectures, when paired with high-quality data, may suffice for segmenting specific retinal structures.

Limitations addressable in future studies include the relatively small dataset, the limited resolution of the OCT device used for the image acquisition and the observer error within each measurement. Although the study utilized over 900 images for training and validation, it lacks diversity in terms of disease types, limiting the broader applicability and generalizability of the project. For future test segmentations with clinically diagnosed patients, this limitation should be addressed, and one should consider increasing the dataset, improving the dCNNs' ability to handle unseen scenarios, reduce segmentation errors, and possibly aid disease differentiation (34).

Another limitation is the use of commercially available OCT images instead of High-Res OCT, which restricts the amount of fine structural detail available for training. Studies have reported significant differences in terms of retinal layer thickness in most of the macular sectors when comparing a commercial device to a High-Res OCT device (63). Incorporating High-Res OCT images in future work could help reduce noise and artifacts, improve annotation accuracy, and enhance the network's performance.

Finally, the potential for inaccuracies in the manual segmentation of retinal layers must be acknowledged. Despite careful annotations, small and complex structures such as the ORBs pose challenges, and observer error cannot be entirely excluded. Additionally, grader-specific

variability contributed to differences in the delineation of thinner structures, particularly Band 2. Grader 2 showed higher intra-observer variability and a consistent positional bias when marking this layer, helping explain the lower inter-grader ICC. Even with a standardised protocol and trained evaluators, manual segmentation of ORBs remains sensitive to human-dependent inconsistencies. Importantly, these inconsistencies directly affect the algorithm, as the network was trained and validated using these manual annotations. Consequently, the segmentation results for Band 2 should be interpreted with caution, and Band 2 cannot be considered a fully reliable output of the proposed model in this study. Including additional graders and adopting consensus or adjudication protocols may help reduce this source of variability in future work.

Although there are already models automatically segmenting retinal structures, these specific locations are not individually segmented and have already proven importance in disease characterization and differentiation (34).

Conclusions

In conclusion, the high repeatability observed for the POS and Band 3+4 contrasts with the reduced intra- and intergrader agreement for Band 2, emphasizing the particular difficulty of accurately segmenting this layer by hand. Our findings also highlight that grader-specific segmentation behaviour, particularly for ultrathin layers such as Band 2, contributes significantly to inter-observer variability. dCNNs particularly the DRUNET architecture, achieved high precision in segmenting these structures, highlighting their potential to reduce variability in human measurements and improve segmentation reliability. These findings support further exploration of dCNNs as a robust tool for clinical implementation in the segmentation of retinal structures.

Acknowledgments

The authors acknowledge Associação de Oftalmologistas para o Estudo de Retina (AOKER), Sociedade Portuguesa de Oftalmologia (SPO) and the Portuguese Recovery and Resilience Program (PRR) via IAPMEI/ANI/FCT (under Agenda C64502239900000057) for its financial support.

Footnote

Reporting Checklist: The authors have completed the

TRIPOD+AI reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-2025-aw-2093/rc>

Data Sharing Statement: Available at <https://qims.amegroups.com/article/view/10.21037/qims-2025-aw-2093/dss>

Funding: This work was supported by research grants from Associação de Oftalmologistas para o Estudo da Retina (Bolsa para Investigação em Retina), Sociedade Portuguesa de Oftalmologia (Bolsa de Investigação Clínica 2023) and the Portuguese Recovery and Resilience Program (PRR) via IAPMEI/ANI/FCT (under Agenda C6450223990000057).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-2025-aw-2093/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of the Lisbon School of Health (ESTeSL) (CE-ESTeSL-No.59-2024). Informed consent was taken from all the patients.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Molday RS, Moritz OL. Photoreceptors at a glance. *J Cell Sci* 2015;128:4039-45.
- Cuenca N, Ortuño-Lizarán I, Pinilla I. Cellular Characterization of OCT and Outer Retinal Bands Using Specific Immunohistochemistry Markers and Clinical Implications. *Ophthalmology* 2018;125:407-22.
- Spaide RF, Curcio CA. Anatomical correlates to the bands seen in the outer retina by optical coherence tomography: literature review and model. *Retina* 2011;31:1609-19.
- Zeppieri M, Marsili S, Enaholo ES, Shuaibu AO, Uwagboe N, Salati C, Spadea L, Musa M. Optical Coherence Tomography (OCT): A Brief Look at the Uses and Technological Evolution of Ophthalmology. *Medicina (Kaunas)* 2023;59:2114.
- Szeskin A, Yehuda R, Shmueli O, Levy J, Joskowicz L. A column-based deep learning method for the detection and quantification of atrophy associated with AMD in OCT scans. *Med Image Anal* 2021;72:102130.
- Zeng Y, Gao S, Li Y, Marangoni D, De Silva T, Wong WT, Chew EY, Sun X, Li T, Sieving PA, Qian H. OCT Intensity of the Region between Outer Retina Band 2 and Band 3 as a Biomarker for Retinal Degeneration and Therapy. *Bioengineering (Basel)* 2024;11:449.
- Frank-Publig S, Bogunovic H, Birner K, Gumpinger M, Fuchs P, Coulibaly LM, Mares V, Michel F, Schmidt FS, Schmidt-Erfurth U, Reiter GS. Quantifications of Outer Retinal Bands in Geographic Atrophy by Comparing Superior Axial Resolution and Conventional OCT. *Invest Ophthalmol Vis Sci* 2025;66:65.
- Lima LH, Sallum JM, Spaide RF. Outer retina analysis by optical coherence tomography in cone-rod dystrophy patients. *Retina* 2013;33:1877-80.
- Wu Z, De Zanet S, Blair JPM, Guymer RH. Loss of OCT Outer Retinal Bands as Potential Clinical Trial Endpoints in Intermediate Age-Related Macular Degeneration. *Ophthalmol Sci* 2025;5:100769.
- Camacho P, Dutra-Medeiros M, Salgueiro L, Sadio S, Rosa PC. Manual Segmentation of 12 Layers of the Retina and Choroid through SD-OCT in Intermediate AMD: Repeatability and Reproducibility. *J Ophthalmic Vis Res* 2021;16:384-92.
- Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. *Clin Exp Ophthalmol* 2019;47:128-39.
- ElTanboly A, Ismail M, Shalaby A, Switala A, El-Baz A, Schaal S, Gimel'farb G, El-Azab M. A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images. *Med Phys* 2017;44:914-23.
- Zhang H, Yang B, Li S, Zhang X, Li X, Liu T, Higashita R, Liu J. Retinal OCT image segmentation with deep learning: A review of advances, datasets, and evaluation metrics. *Comput Med Imaging Graph* 2025;123:102539.

14. Quintana-Quintana OJ, Aceves-Fernández MA, Jesús Carlos Pedraza-Ortega, Gendry Alfonso-Francia, Tovar-Arriaga S. Deep Learning Techniques for Retinal Layer Segmentation to Aid Ocular Disease Diagnosis: A Review. *Computers* 2025;14:298.
15. Fortin JP, Sweeney EM, Muschelli J, Crainiceanu CM, Shinohara RT; Alzheimer's Disease Neuroimaging Initiative. Removing inter-subject technical variability in magnetic resonance imaging studies. *Neuroimage* 2016;132:198-212.
16. Roy S, Chowdhury A, McCreadie K, Prasad G. Deep Learning Based Inter-subject Continuous Decoding of Motor Imagery for Practical Brain-Computer Interfaces. *Front Neurosci* 2020;14:918.
17. Schlegl T, Waldstein SM, Bogunovic H, Endstraßer F, Sadeghipour A, Philip AM, Podkowinski D, Gerendas BS, Langs G, Schmidt-Erfurth U. Fully Automated Detection and Quantification of Macular Fluid in OCT Using Deep Learning. *Ophthalmology* 2018;125:549-58.
18. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express* 2017;8:3440-8.
19. Wang J, Hormel TT, Gao L, Zang P, Guo Y, Wang X, Bailey ST, Jia Y. Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning. *Biomed Opt Express* 2020;11:927-44.
20. Kugelman J, Alonso-Caneiro D, Read SA, Hamwood J, Vincent SJ, Chen FK, Collins MJ. Automatic choroidal segmentation in OCT images using supervised deep learning methods. *Sci Rep* 2019;9:13298.
21. Dos Santos VA, Schmetterer L, Stegmann H, Pfister M, Messner A, Schmidinger G, Garhofer G, Werkmeister RM. CorneaNet: fast segmentation of cornea OCT scans of healthy and keratoconic eyes using deep learning. *Biomed Opt Express* 2019;10:622-41.
22. Bagci AM, Shahidi M, Ansari R, Blair M, Blair NP, Zelkha R. Thickness profiles of retinal layers by optical coherence tomography image segmentation. *Am J Ophthalmol* 2008;146:679-87.
23. Shahidi M, Wang Z, Zelkha R. Quantitative thickness measurement of retinal layers imaged by optical coherence tomography. *Am J Ophthalmol* 2005;139:1056-61.
24. Pekala M, Joshi N, Liu TYA, Bressler NM, DeBuc DC, Burlina P. Deep learning based retinal OCT segmentation. *Comput Biol Med* 2019;114:103445.
25. Li Q, Li S, He Z, Guan H, Chen R, Xu Y, Wang T, Qi S, Mei J, Wang W. DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning. *Transl Vis Sci Technol* 2020;9:61.
26. Moradi M, Chen Y, Du X, Seddon JM. Deep ensemble learning for automated non-advanced AMD classification using optimized retinal layer segmentation and SD-OCT scans. *Comput Biol Med* 2023;154:106512.
27. Zhilin Z, Yan W, Zeyu P, Yunqing Z, Rugang Y, Guogang C. Dual Attention Network for Retinal Layer and Fluid Segmentation in OCT. 2023 8th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS); 2023:179-83.
28. Rivas Vázquez E, Barreira Rodríguez N, López Varela E, González Penedo M. Deep learning for segmentation of optic disc and retinal layers in peripapillary optical coherence tomography images. Zhou J (Jessica), Osten W, Nikolaev DP, editors. Fifteenth International Conference on Machine Vision (ICMV 2022); 2023;56.
29. Cazanias-Gordon A, da Silva Cruz LA. Multiscale Attention Gated Network (MAGNet) for Retinal Layer and Macular Cystoid Edema Segmentation. *IEEE Access*. 2022;10:85905-17.
30. Chen Z, Zhang H, Linton EF, Johnson BA, Choi YJ, Kupersmith MJ, Sonka M, Garvin MK, Kardon RH, Wang JK. Hybrid deep learning and optimal graph search method for optical coherence tomography layer segmentation in diseases affecting the optic nerve. *Biomed Opt Express* 2024;15:3681-98.
31. Roy AG, Conjeti S, Karri SPK, Sheet D, Katouzian A, Wachinger C, Navab N. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed Opt Express* 2017;8:3627-42.
32. Venhuizen FG, van Ginneken B, Liefers B, van Grinsven MJJP, Fauser S, Hoyng C, Theelen T, Sánchez CI. Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. *Biomed Opt Express* 2017;8:3292-316.
33. Kugelman J, Alonso-Caneiro D, Chen Y, Arunachalam S, Huang D, Vallis N, Collins MJ, Chen FK. Retinal Boundary Segmentation in Stargardt Disease Optical Coherence Tomography Images Using Automated Deep Learning. *Transl Vis Sci Technol* 2020;9:12.
34. Heath Jeffery RC, Lo J, Thompson JA, Lamey TM, McLaren TL, De Roach JN, Ayton LN, Vincent AL, Sharma A, Chen FK. Analysis of the Outer Retinal Bands in ABCA4 and PRPH2-Associated Retinopathy using OCT. *Ophthalmol Retina* 2024;8:174-83.
35. Csurka G, Volpi R, Chidlovskii B. Semantic Image

- Segmentation: Two Decades of Research. *Foundations and Trends® in Computer Graphics and Vision* 2022;14:1-162.
36. Hartanto TA, Hansun S. Comparative Analysis of Pre-trained CNN Models on Retinal Diseases Classification. *International Journal of Industrial Engineering & Production Research* 2024;35:21-32.
 37. Li H, Zhang R, Min Y, Ma D, Zhao D, Zeng J. A knowledge-guided pre-training framework for improving molecular representation learning. *Nat Commun* 2023;14:7568.
 38. keyu-tian/SparK: The first successful BERT/MAE-style pretraining on any convolutional network; Pytorch impl. of “Designing BERT for Convolutional Networks: Sparse and Hierarchical Masked Modeling”. [Accessed 2025 May 22]. Available online: <https://github.com/keyu-tian/SparK?tab=readme-ov-file>
 39. BCEWithLogitsLoss — PyTorch 2.7 documentation. [Accessed 21 Jun 2025] Available online: <https://docs.pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html>
 40. Afonso JD. Fovea Detection Algorithm. Available online: https://github.com/Joao-afonso11/OCT_segmentation
 41. Intra class vs Inter class correlation in statistics-Statistical Aid. [accessed 11 Dec 2024] Available online: <https://www.statisticalaid.com/intra-class-vs-inter-class-correlation>
 42. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*; 2015:234-41.
 43. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018)* 2018;11045:3-11.
 44. Myronenko A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. Springer, Cham; 2019:311-20.
 45. Devalla SK, Renukanand PK, Sreedhar BK, Subramanian G, Zhang L, Perera S, Mari JM, Chin KS, Tun TA, Strouthidis NG, Aung T, Thiéry AH, Girard MJA. DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images. *Biomed Opt Express* 2018;9:3244-65.
 46. Kulyabin M, Zhdanov A, Nikiforova A, Stepichev A, Kuznetsova A, Ronkin M, Borisov V, Bogachev A, Korotkich S, Constable PA, Maier A. OCTDL: Optical Coherence Tomography Dataset for Image-Based Deep Learning Methods. *Sci Data* 2024;11:365.
 47. Lakshminarayanan V, Priyanka R, Peyman G. Normal Retinal OCT images. *Optical Coherence Tomography Image Retinal Database*. 2018. doi: 10.5683/SP/WLW4ZT.
 48. Kermany D, Zhang K, Goldbaum M. Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images. 2018. doi: 10.17632/RSCBJBR9SJ.3.
 49. Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms - Review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed* 2018;158:71-91.
 50. Tian J, Varga B, Tatrai E, Fanni P, Somfai GM, Smiddy WE, Debuc DC. Performance evaluation of automated segmentation software on optical coherence tomography volume data. *J Biophotonics* 2016;9:478-89.
 51. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155-63.
 52. Ghorbel I, Rossant F, Bloch I, Tick S, Paques M. Automated segmentation of macular layers in OCT images and quantitative evaluation of performances. *Pattern Recognition* 2011;44:1590-603.
 53. Lee KE, Heitkotter H, Carroll J. Challenges Associated With Ellipsoid Zone Intensity Measurements Using Optical Coherence Tomography. *Transl Vis Sci Technol* 2021;10:27.
 54. Hu Z, Nittala MG, Sadda SR. Comparison of retinal layer intensity profiles from different OCT devices. *Ophthalmic Surg Lasers Imaging Retina* 2013;44:S5-10.
 55. Srinivasan VJ, Monson BK, Wojtkowski M, Bilonick RA, Gorczynska I, Chen R, Duker JS, Schuman JS, Fujimoto JG. Characterization of outer retinal morphology with high-speed, ultrahigh-resolution optical coherence tomography. *Invest Ophthalmol Vis Sci* 2008;49:1571-9.
 56. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. Human photoreceptor topography. *J Comp Neurol* 1990;292:497-523.
 57. Provis JM, Dubis AM, Maddess T, Carroll J. Adaptation of the central retina for high acuity vision: cones, the fovea and the avascular zone. *Prog Retin Eye Res* 2013;35:63-81.
 58. Chui TY, Zhong Z, Song H, Burns SA. Foveal avascular zone and its relationship to foveal pit shape. *Optom Vis Sci* 2012;89:602-10.
 59. Chiu SJ, Li XT, Nicholas P, Toth CA, Izatt JA, Farsiu S. Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation. *Opt Express* 2010;18:19413-28.
 60. Karn PK, Abdulla WH. Advancing Ocular Imaging: A Hybrid Attention Mechanism-Based U-Net Model for Precise Segmentation of Sub-Retinal Layers in OCT

- Images. *Bioengineering* (Basel) 2024;11:240.
61. Matovinovic IZ, Loncaric S, Lo J, Heisler M, Marinko Sarunic. Transfer Learning with U-Net type model for Automatic Segmentation of Three Retinal Layers In Optical Coherence Tomography Images. *IEEE*; 2019. doi: 10.1109/ISPA.2019.8868639
 62. Fang L, Cunefare D, Wang C, Guymer RH, Li S, Farsiu S. Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search. *Biomed Opt Express* 2017;8:2732-44.
 63. Barbieri L, Motta S, Cozzi M, Romano F, Antonioli L, Invernizzi A, Staurenghi G. Macular thickness layers comparison between new High-Res OCT and Spectralis OCT2. *Invest Ophthalmol Vis Sci* 2023;64:3378.

Cite this article as: Afonso JD, Camacho P, Pereira B, Marques JP, Lopes DS, Cabral D. Outer retinal band segmentation in healthy subjects: comparative study between human grading and deep convolutional neural networks. *Quant Imaging Med Surg* 2026;16(5):408. doi: 10.21037/qims-2025-aw-2093

Table S1 Intraclass correlation coefficient (ICC) and average pixel differences for regions around the central fovea analysed by two graders

Location	Image type	Region	ICC	95% confidence interval		Avg diff (pixels)
				Lower bound	Upper bound	
Center fovea	Grader 1	Band 2	0.764	0.629	0.855	1.227
		Band 3+4	0.848	0.745	0.912	1.965
		POS	0.864	0.778	0.918	1.602
	Grader 2	Band 2	0.698	0.417	0.857	1.854
		Band 3+4	0.776	0.549	0.897	3.243
		POS	0.749	0.503	0.883	2.130
500 μm	Grader 1	Band 2	0.730	0.630	0.806	1.361
		Band 3+4	0.715	0.601	0.800	1.992
		POS	0.746	0.652	0.817	1.380
	Grader 2	Band 2	0.598	0.379	0.753	1.544
		Band 3+4	0.722	0.523	0.841	3.405
		POS	0.093	-0.130	0.330	2.711
750 μm	Grader 1	Band 2	0.697	0.588	0.781	1.551
		Band 3+4	0.852	0.787	0.898	1.632
		POS	0.800	0.723	0.857	1.316
	Grader 2	Band 2	0.539	0.305	0.712	1.411
		Band 3+4	0.860	0.719	0.927	2.897
		POS	0.316	0.052	0.543	2.074
1000 μm	Grader 1	Band 2	0.732	0.633	0.807	1.345
		Band 3+4	0.837	0.766	0.887	1.658
		POS	0.730	0.630	0.806	1.477
	Grader 2	Band 2	0.539	0.301	0.714	1.411
		Band 3+4	0.870	0.748	0.931	2.917
		POS	0.577	0.352	0.739	1.756

Repeatability of each grader to perform 3 different structure segmentations with its respective confidence intervals and average pixel differences between measurements. Each second measurement was performed at least two days later and masked to the first one. The central fovea detection algorithm was used to compare the repeatability in the different regions around the central fovea. ICC, intraclass correlation coefficient; Avg Diff, average difference; Band 2, ellipsoid zone; Band 3+4, interdigitation zone of photoreceptors (IZP, Band 3) and retinal pigment epithelium (RPE, Band 4); POS, photoreceptor outer segments.