Artificial Intelligence In Medicine 158 (2024) 103006

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Research paper

Fully automatic deep convolutional approaches for the screening of neurodegeneratives diseases using multi-view OCT images

Lorena Álvarez-Rodríguez [a,b], Ana Pueyo [c,d], Joaquim de Moura [a,b,*], Elisa Vilades [c,d], Elena Garcia-Martin [c,d], Clara I. Sánchez [e,f], Jorge Novo [a,b], Marcos Ortega [a,b]

[a] VARPA Group, Biomedical Research Institute of A Coruña (INIBIC), University of A Coruña, A Coruña, Spain
[b] CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain
[c] Department of Ophthalmology, Miguel Servet University Hospital, Zaragoza, Spain
[d] Aragon Institute for Health Research (IIS Aragon), Miguel Servet Ophthalmology Innovation and Research Group (GIMSO), University of Zaragoza, Zaragoza, Spain
[e] Quantitative Healthcare Analysis (qurAI) Group, Informatics Institute, Universieit van Amsterdam, Amsterdam, The Netherlands
[f] Biomedical Engineering and Physics, Amsterdam UMC Locatie AMC Department of Biomedical Engineering and Physics, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Neurodegenerative diseases
OCT
Multi-view
Retinal layers
Deep learning
Screening
Retinal layers segmentation

ABSTRACT

The prevalence of neurodegenerative diseases (NDDs) such as Alzheimer's (AD), Parkinson's (PD), Essential tremor (ET), and Multiple Sclerosis (MS) is increasing alongside the aging population. Recent studies suggest that these disorders can be identified through retinal imaging, allowing for early detection and monitoring via Optical Coherence Tomography (OCT) scans. This study is at the forefront of research, pioneering the application of multi-view OCT and 3D information to the neurological diseases domain. Our methodology consists of two main steps. In the first one, we focus on the segmentation of the retinal nerve fiber layer (RNFL) and a class layer grouping between the ganglion cell layer and Bruch's membrane (GCL-BM) in both macular and optic disc OCT scans. These are the areas where changes in thickness serve as a potential indicator of NDDs. The second phase is to select patients based on information about the retinal layers. We explore how the integration of both views (macula and optic disc) improves each screening scenario: Healthy Controls (HC) vs. NDD, AD vs. NDD, ET vs. NDD, MS vs. NDD, PD vs. NDD, and a final multi-class approach considering all four NDDs. For the segmentation task, we obtained satisfactory results for both 2D and 3D approaches in macular segmentation, in which 3D performed better due to the inclusion of depth and cross-sectional information. As for the optic disc view, transfer learning did not improve the metrics over training from scratch, but it did provide a faster training. As for screening, 3D computational biomarkers provided better results than 2D ones, and multi-view methods were usually better than the single-view ones. Regarding separability among diseases, MS and PD were the ones that provided better results in their screening approaches, being also the most represented classes. In conclusion, our methodology has been successfully validated with an extensive experimentation of configurations, techniques and OCT views, becoming the first multi-view analysis that merges data from both macula-centered and optic disc-centered perspectives. Besides, it is also the first effort to examine key retinal layers across four major NDDs within the framework of pathological screening.

1. Introduction

Neurodegenerative diseases (NDDs) are progressive, have an immediate impact on the central nervous system, and interfere with the neural networks ability to communicate with one another [1]. As life expectancy increases, the prevalence of these pathologies is also expected to rise [2], since aging is one of the main risk factors [3]. In this work, we concentrate on four key NDDs: Alzheimer's disease (AD), which is expected to affect 131 million patients by 2050 [4]; Parkinson's disease (PD), which is the most common neurological disorder with movement problems and whose prevalence has doubled,

* Correspondence to: VARPA Group, Department of Computer Science and Information Technologies, Scientific Area Building, D. 1.06, Faculty of Informatics, Campus de Elviña S/N, P.C. 15071 University of A Coruña, A Coruña, Spain.
E-mail addresses: lorena.alvarezr@udc.es (L. Álvarez-Rodríguez), ana.pueyo@gmail.com (A. Pueyo), joaquim.demoura@udc.es (J. de Moura), elisavilades@hotmail.com (E. Vilades), egmvivax@yahoo.com (E. Garcia-Martin), c.i.sanchezgutierrez@uva.nl (C.I. Sánchez), jnovo@udc.es (J. Novo), mortega@udc.es (M. Ortega).

https://doi.org/10.1016/j.artmed.2024.103006
Received 9 October 2023; Received in revised form 19 October 2024; Accepted 23 October 2024
Available online 1 November 2024

increasing faster than for any other NDD [5]; Essential tremor (ET), that affects 1% of the world's entire population [6]; and Multiple Sclerosis (MS), that is the most common non-traumatic disabling disease affecting young adults [7]. These disorders usually have a late diagnosis or even a misdiagnosis due to the similarities in symptoms and clinical findings in the tissues [8], which negatively affects the quality of life of the patient. Much effort has been put towards finding new tests and biomarkers to improve the accuracy and speed of the diagnosis, and new research has proven that not only neurological deterioration is not only found in brain images, but also in fundus of the eye images [9,10], specifically in the retinal layers. These observable characteristics make possible the diagnosis and study of the progression of the disease by analyzing images of the eye. In particular, the technique typically used is Optical Coherence Tomography (OCT) [11]. This device employs a non contact, in vivo approach based on echo time delay or frequency information of back-reflected light to capture high-resolution, micron-scale images of tissue structures. The key strength of OCT lies in its capacity to generate detailed cross-sectional and volumetric images of the desired regions, offering a comprehensive view of tissue morphology and architecture, which is ideal for studying the aforementioned retinal layers.

Numerous works have taken advantage of this important potential of OCT in relation to retinal layers and have proposed different methods for their automatic segmentation [12–15], with deep learning approaches demonstrating particularly satisfactory results. The recognition and analysis of these retinal layers seen in OCTs has been proven relevant in the diagnosis of NDDs [16], in particular the ones considered in this work like AD [17], ET [18], MS [19], and PD [20]. This lead to the developing of methods specifically to measure changes in the eye captured by OCT specially related to NDDs in order to be able to identify and quantify those changes. Most of the published works focus on MS patients, like developing a deep network for retinal layer segmentation and microcystic macular edema (MME) segmentation on MS patients [21] or analyzing their macular thickness obtained from OCT using a SVM [22] or a CNN [23]. Despite clinical findings highlighting the importance of retinal layers for diagnosing different NDDs, only one research by Gende et al. [24] takes into account multiple NDDs concurrently in a macular OCT perspective. This underlines an important gap in the current landscape of NDDs diagnostics, pointing to an unexplored field that holds substantial potential for improving early detection and prognosis of these diseases. Pushing forward the exploration of deep convolutional approaches in this emerging field, this work is the first of its kind to analyze multi-view OCT images for the screening of various neurodegenerative diseases, using both macular an optic disc scans.

Typically, studies in this area analyze each cross-sectional image, known as a B-scan, on an individual basis. However, when these B-scans are stacked together, they form 3D volumes that carry significant potential for providing additional, more complex information. These volumes not only retain the details from individual B-scans but also present spatial context across the stack of images. As such, they can reveal intricate details about tissue thickness and morphology which might not be readily apparent from examining single B-scans. Some works have pointed out how these features could be of use but are overlooked [25], possibly due to images resolution or possible motion artefacts across B-scans [26]. However, some others have shown satisfactory results using OCT volumes instead of processing each cross-sectional image. Chen et al. [27] proposed an automated method for segmenting 3D fluid-associated abnormalities in the retina from 3D OCT retinal images of subjects suffering from exudative age-related macular degeneration. Wu et al. [28] designed an automatic, 3D segmentation method to detect both neurosensory retinal detachment (NRD) and pigment epithelial detachment (PED) in spectral domain optical coherence tomography images. These same authors also proposed an automated 3D segmentation framework to detect subretinal fluid in SD-OCT volumes [29]. Maetschke et al. [30] used a deep learning technique that classified eyes as healthy or glaucomatous directly from raw, unsegmented OCT volumes of the optic nerve head using a 3D Convolutional Neural Network (CNN). However, in the NDDs domain and retinal layers segmentation, to the best of our knowledge, this work is the first to study whether using this 3D information in OCT provides interesting features for automatic segmentation of retinal layers and the consequent study of the mentioned neurological diseases in a screening scenario.

The interpretations of the data obtained from these segmentations are wide-ranging and depend on the specific tasks they are applied to. For instance, Holmberg et al. [31] applied model weights, originally trained for retinal layer segmentation, to classify diabetes grading based on fundus images. Similarly, Mohammed et al. [32] used the thickness data of retinal layers to categorize patients into different stages of diabetic retinopathy. In the realm of NDDs specifically, Garcia-Martin et al. [33] utilized the thickness of various retinal layers to facilitate early diagnosis of MS. It is important to note that these studies generally focus on a specific view of the OCT images for their classification tasks. However, medical professionals often use multiple views and modalities, suggesting that integrating data from these various sources could enhance and reinforce final diagnoses. There are works that explore these concepts: for example, He et al. [34] proposed a multi-modal methodology that incorporates OCTs and fundus images for classifying different retinopathies; DISCOVER is 2D multi-view summarization of OCT angiography for automatic diabetic retinopathy diagnosis [35]. Hence, related but different features that can be extracted from different views of OCT can be enhanced by 3D and built into strong computational biomarkers. This added to the use of a non-invasive and accessible technique such as OCT is fundamental to provide feasible early detection in real clinical practice. In this work, we implement a multi-view approach that integrates data from macula-centered and optic-disc-centered views: first by automatically segmenting 2D and 3D cubes from each view, and then by combining them to use the segmented and generated features to perform disease screening. Our pioneering approach is the first in automatically segment retinal layers in macular and optic disc OCT and then using both views information to classify patients among four major NDDs, being one of the few works to consider several at the same time. We firmly believe this methodology marks a significant advancement towards enhancing diagnostic precision and early detection of these diseases. The principal contributions of our study include:

- This research represents a new effort to segment and analyze the key retinal layers across four major NDDs in a few explored field.
- We are at the forefront of examining these four diseases within the framework of pathological screening, thus enhancing diagnostic capabilities for neurodegenerative conditions.
- We introduce the first multi-view analysis that incorporates data from both macula-centered and optic disc-centered perspectives. This novel approach effectively leverages the depth of information contained within OCT images to enhance diagnostic accuracy.

To illustrate our proposal, Fig. 1 shows an schematic of our complete pipeline, with two distinct modules: segmentation and classification. Each has two branches: 3D macular and single-scan optic-disc view.

This manuscript is organized as follows: Section 2 delineates the data and the segmentation and classification backbones utilized in this study, along with a detailed explanation of the experimental setup and the evaluation strategy. Section 3 presents the methodology employed to address each task: first the segmentation architecture used for macular and optic-disc view, and secondly the proposal for pathological screening based on segmentation and other features. Section 4 showcases the results obtained from each approach. Finally, Section 5 summarizes the key findings and conclusions of this research.

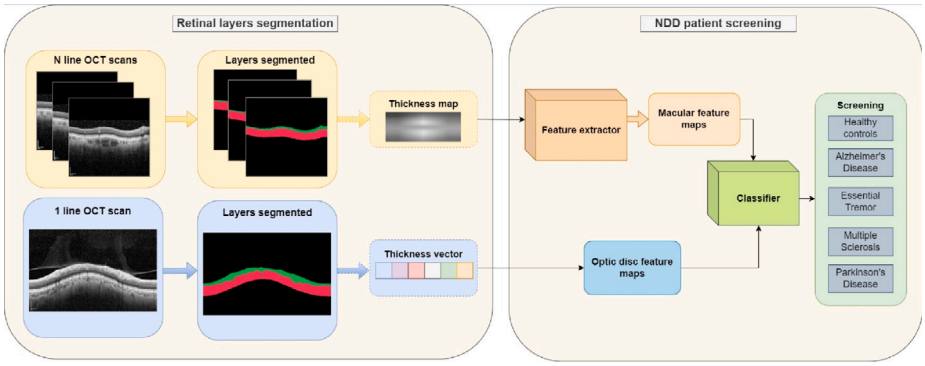

Fig. 1. General overview of our methodology: Segmentation and classification modules, with a macular and optic-disc branch each.

2. Materials and methods

In this section the materials and methods used in our pipeline are described. First, the datasets for segmentation and classification for each view (macular and optic disc). Then, the backbone used in segmentation, and the ones needed for the classification task: a macular feature extractor and a classifier.

2.1. Dataset

For this work, a main dataset was available from which subsets were made according to the task to be solved. All the samples were acquired by a Heidelberg SPECTRALIS® imaging platform and assessed by a neurophthalmologist with a focus on neurophthalmology, a neuropathologist with a focus on demyelinating illnesses, and a neuropathologist with a focus on movement disorders and dementia. The samples belonged to patients from five classes, four of them being the diseases previously discussed (AD, PD, MS, ET), and the fourth being healthy controls (HC). NDDs patients were diagnosed by experts, while HC contains patients who were referred for testing but no ocular abnormalities were found. Exclusion criteria for all participants included best-corrected visual acuity less than 0.5 (Snellen chart), refractive errors greater than 5 diopters of spherical equivalent or 3 diopters of astigmatism, intraocular pressure greater than 20 mmHg, and media opacities (nuclear color/opalescence, cortical or posterior subcapsular lens opacity > 2, according to the Lens Opacity Classification System III). In addition, patients with glaucoma, retinal diseases or systemic conditions affecting vision were excluded. All procedures followed the Declaration of Helsinki, and written informed consent was obtained from all participants. The study was approved by the Ethics Committee of Hospital Miguel Servet (CEICA: Comité Ético de Investigaciones Científicas de Aragón) with registration number C.I. PI21/113. No manual correction was applied to the OCT output.

For the segmentation subsets, we had ground truth annotations with two semantic classes: the Retinal Nerve Fiber Layer (RNFL), and a class grouping the remaining retinal layers between the Ganglion Cell Layer and Bruch's Membrane (GCL-BM). The subset used to validate the macular segmentation methodology consisted of 1250 B-scans organized into 50 OCT cubes composed of 25 equally spaced macular B-scans each taken in line-raster pattern over the macular area. There were 10 samples of size 512 × 496 per class, from 50 different patients in total. These types of OCT scans are commonly used in medical research for thickness measuring [36–39], so for the sake of replicating a real clinical scenario, we used this configuration of OCT cubes. The subset for optic disc segmentation was formed by 90 one-line scans taken in the parapapillary area, measured along a 3.45-mm-diameter around the optic disc, considering 18 1-line OCT scans per class, being each one from a different patient. Each of the segmentation subsets were randomly divided into three splits, following a 70–10%–20% proportion for the training, validation and test sets, with the same amount of patients from each class (see Tables 1–3).

Table 1
Segmentation dataset used for macular view. Each case correspond to a unique patient.

Patient type	Macular view	
	Cases	B-scan/case
HC	10	25
AD	10	25
ET	10	25
MS	10	25
PD	10	25
Total	*50*	*25*10*50=1,250*

Table 2
Segmentation dataset used for optic disc view. Each case correspond to a unique patient.

Patient type	Optic disc view	
	Cases	B-scan/case
HC	18	1
AD	18	1
ET	18	1
MS	18	1
PD	18	1
Total	*90*	*1*18*90=900*

Table 3
Classification dataset, indicating number of samples. Each case correspond to a unique patient with information from both views. The amount of B-scans per case range from 25 to 61 since different macular cube densities were used.

Patient type	Cases	Macular view	Optic disc view
		B-scan/case	B-scan/case
HC	81	25–61	1
AD	29	25–61	1
ET	10	25–61	1
MS	166	25–61	1
PD	82	25–61	1
Total	*368*	*9,500*	*1*368=368*

Regarding the classification subset, for each class, we selected the most recent acquisition with both macular and optic disc information for each patient. Thus, we had 81, 29, 10, 166 and 82 patients for HC, AD, ET, MS and PD, respectively, where we had macular and optic disc information. 80% of this classification dataset was dedicated to training and 20% to test, stratifying per class in each set.

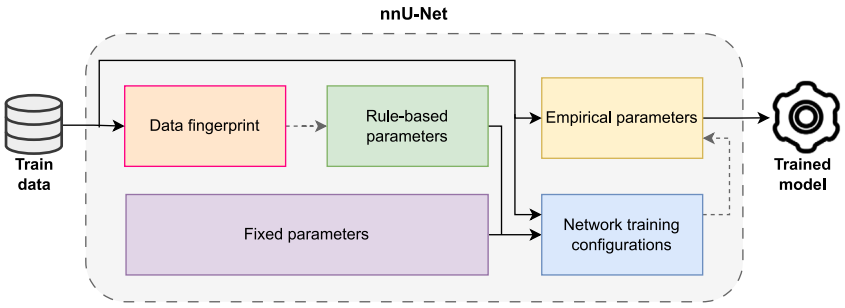

Fig. 2. Diagram of nnU-Net framework.

2.2. Segmentation backbone

In this work, we have exploited the potential provided by the nnU-Net architecture [40]. By using a set of fixed parameters, interdependent rules and empirical decisions, the authors created a self-configured deep learning segmentation method validated on 23 public datasets used in international biomedical segmentation competitions, where it surpassed most existing approaches. Its main structure is shown in Fig. 2 as an overview. Regarding parameters, there are three types: rule-based, fixed and empirical. The rule-based parameters are automatically used to configure hyperparameters of the model, like the learning rate, batch size, and number of epochs, based on the properties of the dataset. The fixed parameters are hyperparameters that are the same for all datasets and experiments, like the number of feature maps or the number of filters in the convolutional layers. Empirical parameters are those tuned empirically during training, such as the dropout rate and the weight decay coefficient. Finally, network training configurations are the settings related to the training process, such as the optimizer used for gradient descent, the loss function used for backpropagation, and the weight initialization method, that are configured based on the previous data contained in the framework. Thus, nnU-Net does not represent a new method, but a powerful systematic approach to all the steps in the training pipeline of semantic segmentation models, which have shown successful results in the medical domain, specially in MRI image [41–43]. However, at the time of writing, nnU-Net has not yet been used in OCT imaging, so this would be the first work to do so.

Most of the parameters used for this work were defined by the method itself. The architecture template is fixed, and it is based on the original U-Net and its 3D variation, with some minor variations over the original one. The initial number of feature maps was set to 32 and doubled with each downsampling operation as a compromise between performance and memory load. The final number of feature maps was limited to 320 for 3D and 512 for 2D U-Nets to control the final model size. The models trained for 1000 epochs, being one epoch an iteration over 250 mini-batches, and following a 5-fold cross-validation strategy. The weights were learned using Stochastic gradient descent with Nesterov momentum and an initial learning rate of 0.01, which decayed following the poly learning rate strategy $(1 - epoch/epoch_{max})^{0.9}$. Cross-entropy (Eq. (1)) and Dice loss (Eq. (2)) were summed as the loss function used: for each deep supervision output, a corresponding downsampled ground truth segmentation mask is used for loss computation. The training objective is the sum of the losses $\mathcal{L}$ at all resolutions, $\mathcal{L} = w_1 \times \mathcal{L}_1 + w_2 \times \mathcal{L}_2 + \cdots$. Thus, the weights ($w$) are halved with each decrease in resolution, resulting in $w_2 = \frac{1}{2} \times w_1, w_3 = \frac{1}{4} \times w_1$, etc. and are normalized to sum to 1. Also, data augmentation was performed, and it included rotations, scaling, Gaussian noise and blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring. This hyperparametrization and technique selection is the result of the domain knowledge condensation of nnUNet authors after thorough experimentation with multiple medical image datasets and tasks.

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(p_{i,c}) \tag{1}$$

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} p_{i,c}}{\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} + \sum_{i=1}^{N} \sum_{c=1}^{C} p_{i,c}} \tag{2}$$

As preprocessing, intensity is normalized by z-scoring. Then, images are resampled to the same target spacing by third-order spline interpolation. This target spacing is the median value of the spacings found in the training cases computed independently for each axis. The patch size was initialized to the median image shape after resampling and iteratively reduced while adapting the network topology accordingly until the network can be trained with a batch size of at least 2 given GPU memory constraints. As for the empirical parameters, connected component-based post-processing was applied to every class. All these parameters were either fixed or automatically selected by the framework, but a small modification was made to be able to use the models of the 3D U-Net cascade configuration. By default, the authors considered that this approach is only applicable to large images, so only if the patch size of the 3D full resolution U-Net covers less than 25% of the median resampled image shape. In this case, the images used did not reach this minimum, so the parameter controlling this limit was modified to just 100% of the median shape of the data in order to be able to test this configuration on our data. This affected the configuration for the low-resolution 3D U-Net, since with this modification it was iteratively increasing the target spacing while reconfiguring the patch size, network topology and batch size as already described until the configured patch size covers 100% of the median image shape. The batch size was automatically set to 12 for 2D, 2 for cascade and 3D, based on a strategy where the minimum was 2 and it was increased until the GPU is fully used, capping it so the total number of voxels in the mini-batch do not exceed 5% of the total number of voxels of all training cases, to prevent overfitting.

2.3. Classification backbones

For the classification problem, we use two different backbones: a feature extractor and a classifier. The first one is a pre-trained, self-supervised feature extractor that gets the representation of the different macular data. Specifically, macular data was the most complex and heterogeneous in terms of formats, so a generic feature extractor was needed to homogenize the information extraction and identification of possible computational biomarkers. To this end, we use the DINO framework [44], which has been used previously for different medical imaging tasks [45–47]. Following their strategy, we have the following main elements, simplified for the case where we have only a pair of views. First, the image x is randomly transformed into a set of different views, x_1 and x_2 in this case, which are the inputs to the student and teacher networks $g\theta_s$ and $g\theta_t$, respectively. These architectures g are composed by a projection head h and a backbone f, so $g = h \circ f$. In

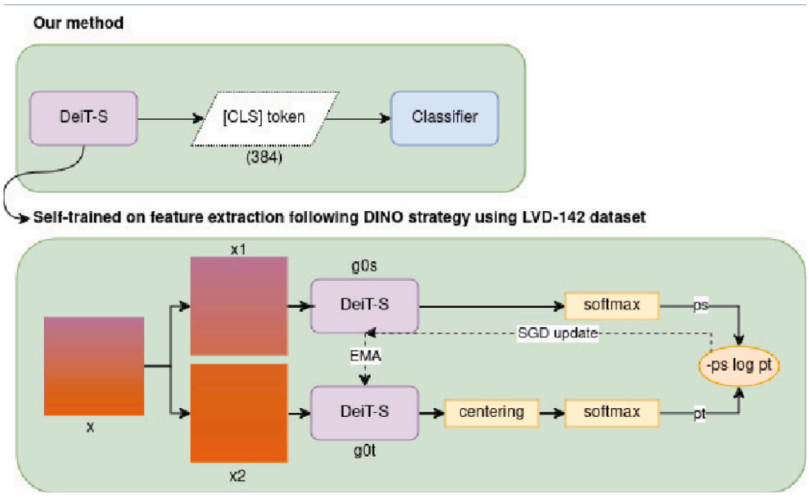

Fig. 3. Diagram illustrating how our feature extractor, DeiT-S, is used in our method, and how is previously trained following DINO strategy.

our case, h is a 3-layer MLP with hidden dimension 2048 followed by l_2 normalization and weight normalized fully connected layer with K dimensions, and f is the DeiT implementation of ViT [48], in particular the DeiT-S. DeiT has two main features. The first one is using hard-label distillation. They use the decision of the teacher model as a true label, and it is defined in Eq. (3), being Z_s the logits of the student model, y the ground truth label, ψ the softmax function, and $y_t = \arg\max_c Z_t(c)$ the hard decision of the teacher. This definition is valid for soft-labeling if we include a smoothing term ε, being $1-\varepsilon$ for the true label, and the remaining for the other classes. In our case, we used soft-label distillation, with ε being 0.1.

$$\mathcal{L}_{hard\,DistillGlobal} = \frac{1}{2}\mathcal{L}_{CE}(\psi(Z_s), y) + \frac{1}{2}\mathcal{L}_{CE}(\psi(Z_s), y_t) \tag{3}$$

The second feature of DeiT is the inclusion of a distillation token added to the initial patches and class tokens, and it acts like the latter. It interacts with the other embeddings via self-attention and the network outputs it after the last layer. It allows for the model to learn from the output of the teacher, as in regular distillation, while remaining complementary to class embedding. DeiT was created focusing on knowledge distillation, so a strong image classifier as teacher model was assumed to be available. However, in self-supervised learning as in DINO's approach, the teacher network is not provided beforehand, so we need to build it based on the iterations of the student network. The teacher parameters θ_t are updated in accordance with the student parameters θ_s using a exponential moving average. Then, the teacher network is followed by a centering and sharpening operation, which is done to avoid the collapse of the network, a common problem in self-supervised learning. Both student and teacher K dimensional feature vectors are normalized with temperature softmax over feature dimension. A stop-gradient operator is set for the teacher network so the gradients are propagated only through the student. Finally, the similarity between the feature vectors is measured using CE loss. In our case, as in other works [45,46,49], we extracted the features from the macular data with the DeiT-S model pretrained in LVD-142M, a dataset of more than 142 millions of images extracted by the DINO authors from several other datasets like ImageNet-22k, SUN397 and DTD, among others. Thus, although preliminary experiments were done on OCT fine-tuning, no benefits were obtained from this strategy and only pretrained in LVD-142M models were used. Fig. 3 shows a simplified illustration of how our method uses this feature extractor, and how it is pre-trained.

The classifier backbone is a Histogram-based Gradient Boosting Classification Tree (HGBCT) [50]. It is an ensemble learning algorithm used for classification tasks. It combines the power of Histogram-based Gradient Boosting, which uses histograms to discretize the feature space and improve efficiency, with Classification Trees, which create a tree-like structure to make decisions based on feature values. HGBCT builds a sequence of classification trees using histogram-based optimization, allowing for faster training times and efficient memory usage, making it well-suited for large datasets. It provides accurate and robust models with good generalization to new data, making it a popular choice for various machine learning tasks in the medical domain [51–53]. Given the nature of the features considered, we chose this traditional classifier due to their ability to handle engineering features designed to capture specific, predefined information from the data. They can provide interpretability, which is a key factor in medical diagnostics. CNN-based classifiers are powerful methods for high-dimensional raw data, but lack transparency and are computationally intensive, making them less suitable for scenarios where a fast and clear explanation is intended. In particular, we used the scikit-learn implementation, trained during 300 iterations with an initial learning rate of 0.01 and BCE as loss. The maximum number of leaves, depth for each tree and proportion of randomly chosen features in each and every node split were 31, 10 and 1, respectively. Also, given the class imbalance, class weights were pondered towards the classes with less samples.

2.4. Software and hardware resources

The Python (v. 3.8.10) implementation of nnU-Net provided by the authors of the architecture has been used (v. 1.7.0). The main libraries used otherwise were PyTorch (v. 1.10.1) and scikit-learn (v. 1.2.0). As for hardware, a machine was used with a NVIDIA RTX A6000 GPU, an AMD Ryzen Threadripper 3960X CPU, and 256 GB RAM.

2.5. Evaluation

The metrics used to obtain a complete analysis of the performance of the models trained were accuracy (Eq. (4)), precision (Eq. (5)), specificity (Eq. (6)), sensitivity (Eq. (7)), Dice (Eq. (8)) and F1-Score (Eq. (9)), defined as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Specificity = \frac{TN}{FP+TN} \tag{6}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{7}$$

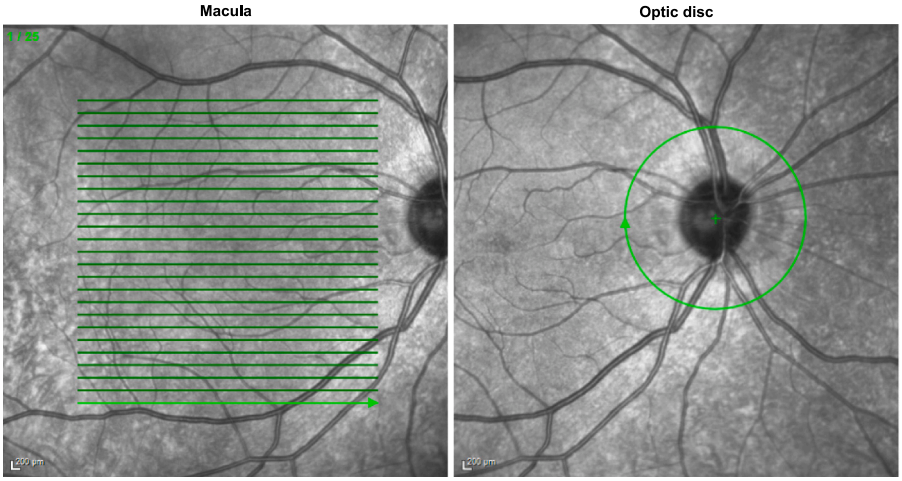

Fig. 4. Fundus image indicating from where the scans of the macular-centered volumes and optic-disc scans were obtained. The green arrows indicate the place and direction of these scans, 25 for macular volumes, while 1 one-line scan for optic-disc.

$$Dice = \frac{2 * TP}{2 * TP + FP + FN} \tag{8}$$

$$F1 - Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{9}$$

where TP, FP, TN and FN are True Positive, False Positive, True Negative, and False Negative, respectively. Additionally, to evaluate the shape of the obtained segmented layers, two additional metrics based on the Mean Absolute Error (MAE) were considered. Considering a predicted segmentation mask for a specific layer M_p of size H×W and its ground truth M_{gt}, we can compute the MAE_c as a measure of contour error by calculating the MAE of the thickness of the map per column for the prediction (V_p), and ground truth (V_{gt}), both defined in Eq. (10). Thus, MAE_c definition will be as defined in Eq. (11) and measured in micrometers by multiplying by 4 the result in pixels (scale: 1 pixels = 4 micrometers). Additionally, we can also compute the full thickness error by computing the predicted thickness T_{M_p} and the ground truth $T_{M_{gt}}$ as the sum of the corresponding V_p and V_{gt} vectors, and then compare it by MAE means, thus defining MAE_t, as in Eq. (12). By using both, we get a thorough evaluation of the thickness extracted, which is the biomarker used in the clinical domain. MAE_t focuses on this by simply comparing total thickness of the layer, while MAE_c does a exam comparing each column individually.

$$V_x = \begin{bmatrix} V_{x_i} & \dots & V_{x_W} \end{bmatrix}, V_{x_i} = \sum_{j=1}^{H} M_{x_{ij}} \tag{10}$$

$$MAE_c(M_p, M_{gt}) = MAE(V_p, V_{gt}) = \frac{1}{W} \sum_{i=1}^{W} |V_{p_i} - V_{gt_i}| \tag{11}$$

$$MAE_t(M_p, M_{gt}) = |T_{M_p} - T_{M_{gt}}|, T_{M_x} = \sum_{i=1}^{W} V_{x_i} \tag{12}$$

3. Methodology

Our methodology comprises the resolution of two tasks: segmentation of retinal layers, and classification of the information obtained among the different screening approaches. The first task is solved by an automatic retinal layer segmentation model that produces maps indicating two relevant layers from two types of views: macular and optic-disc. Then, these maps are transformed into feature vectors: the macular maps by using a self-trained feature extractor, and the optic-disc ones with a 1D projection. These feature vectors are used as input for a machine learning classifier to perform the pathological screening. In this section the details of each task and method are described.

3.1. Retinal layers segmentation

The layers considered in this task were the Retinal Nerve Fiber Layer (RNFL), and those between the Ganglion Cell Layer and Bruch's Membrane (GCL-BM). For this purpose, we consider two views, as depicted in Fig. 4. For the first one centered on macula, we had obtained image volumes which provide complete information of the retinal layers disposition all over the macular area. For the second, centered on the optic disc, we considered the retinal layers disposition in the parapapillary area measured around the optic disc. We used transfer learning form the previous view to improve the results. Thus, we can obtain crucial information on how the different NDDs studied affect each specific retinal layer and zone.

3.1.1. Macular view

As shown in Fig. 5, we use nnU-Net architecture to segment two layers in 25 line OCT scans in macular view. This backbone was used following three different approaches depending on the configuration used: a 2D classical setting (2D) U-Net; a 3D multi-scale cascade configuration (3DC) which operates on downsampled images, and the second is trained to refine the segmentation maps created by the former at full resolution; and the 3D full resolution approach (3D). To validate this methodology, 50 macula-centered OCTs of 25 scans each were available. Each sample was from a different patient, with 10 patients for each NDD considered (AD, ET, MS, PD) and healthy controls (HC). Particularly, each class contributed with 175 B-scans for training (875 in total), 25 B-scans for validation (125 in total) and 50 for test (250 in total).

3.1.2. Optic disc view

For this view we decided to follow a transfer learning strategy. A specific training for this view was required despite its similarities to macular view due to the changes in morphology and arrangement of the layers between zones. Also, we had less samples for this view, which set data scarcity as a problem. Therefore, we took advantage of the relationship between the two types of images and trained a macula-centered model which weights were used as initialization in the optic-disc training. The whole methodology of this approach is summarized in diagram 6. Since transfer learning was being used, it was decided to test how the number of samples used for training affected models trained using macular weights, and models trained from scratch in the traditional way. Specifically, training sets of 4, 8, 16, 24, 32, 40 48 and 56 samples were tested, being 56 the maximum amount of samples available for training following the aforementioned division of the three sets.

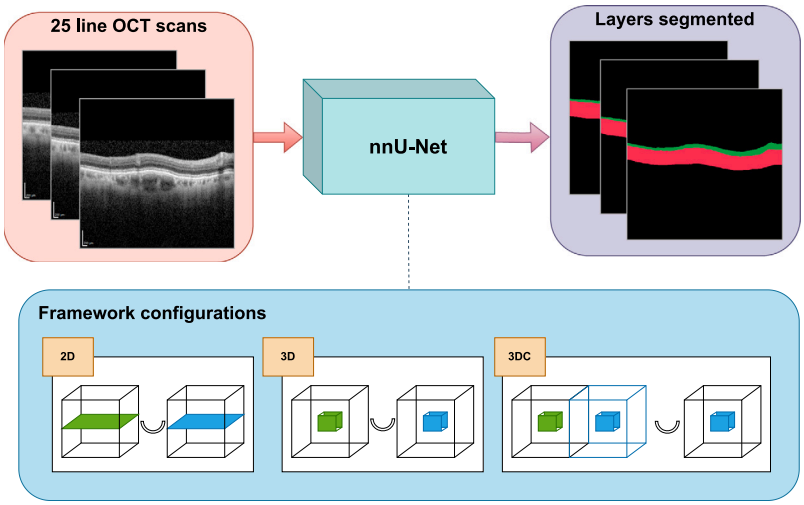

Fig. 5. Diagram of the macular-view segmentation methodology.

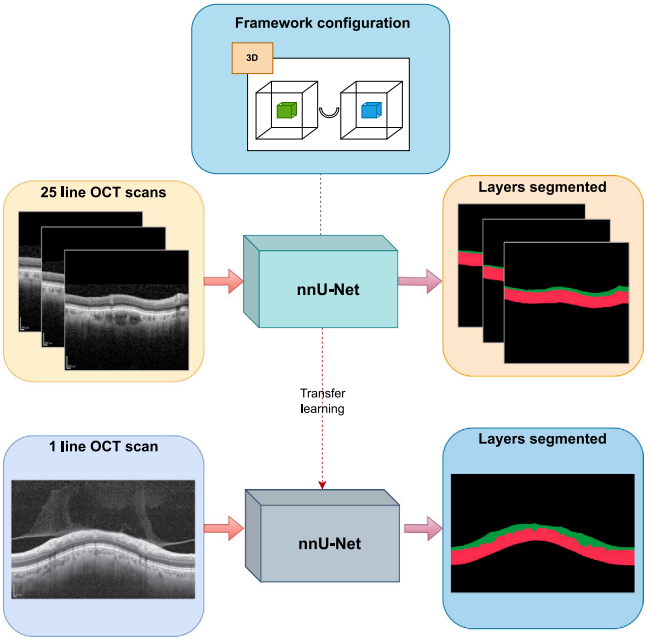

Fig. 6. Diagram of the optic-disc-view segmentation methodology.

3.2. Neurological diseases screening

For the patient screening we have followed two different strategies depending on the information used: single-view when we use either macular or optic disc information, and multi-view when we use both. Each of these views was tested with the raw image information projection, the RNFL thickness, the GCL-BM thickness, and both RNFL and GCL-BM thickness. In the case of the macular view, we also considered the 2D and 3D variants it can provide. For each strategy with different information, we have different screening approaches to study the separability of the classes considered, focusing on examining controls against NDD patients, each NDD patient against all the other diseases considered, and finally each class against each one of the others. In particular, the six approaches are: (i) HC VS NDD (AD + ET + MS + PD), (ii) AD VS NDD (ET + MS + PD), (iii) ET VS NDD (AD + MS + PD), (iv) MS VS NDD (AD + ET + PD), (v) PD VS NDD (AD + ET + MS) and (vi) All VS All (HC VS AD VS ET VS MS VS PD).

3.2.1. Single view: Macula

Complex data can be transformed into feature vectors used for classification [54,55]. Here our macular data with different representations, such as raw B-scans or segmentation maps, are used as input of our feature extractor, a pretrained DeiT model which generates the vector used for the patient screening. First, we have used the direct projection of the original volume through the Z axis maps of size $496 \times N$, where N is the number of the slices of the volume. We have also applied the same projection, but for each segmented layer and the sum of both, so we had a map of the same size but indicating the thickness of the desired element. Since these are 3D scans, we also used the raw and segmented volumes. Finally, a combination of the 2D and 3D information was used.

3.2.2. Single view: Optic disc

For the optic disc view, the information extracted is the same as in the macular view, but since we only have one scan per sample, the information maps are 1D vectors. The original scans had different sizes, so we limited the vectors size to the most common width in the original scans: 1036.

3.2.3. Multi-view

In this strategy we have used the single-view data already described: both macular and optic disc information. These samples are processed as indicated in Fig. 7, producing three main approaches to this multi-view methodology based on the type of information considered: 2D, 3D and 2D+3D.

4. Results

In this section, we present the results of the validation of our methodology. It should be noted that to assess the stability of our methods, each experiment was repeated 3 times. Hence, the metrics presented here are the average of those repetitions.

4.1. Retinal layers segmentation

In this section, the results for the first task, retinal layer segmentation, are shown for each view. The first, macular view, is given in 3D OCT scans that are processed following three different approaches: 2D, cascade and 3D. Then, the second view, optic disc is composed by single-line images. Taking advantage of the already trained macular models, a transfer learning to optic-disc view study is shown.

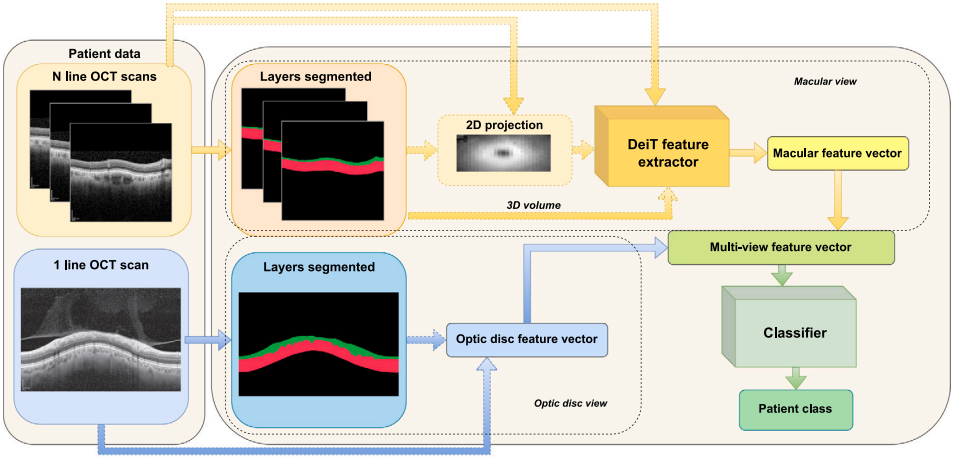

Fig. 7. Summary diagram of the classification methodology.

Table 4
Mean test and corresponding standard deviation results obtained with 2D approach.

Retinal Nerve Fibre Layer					
Class	Accuracy	Precision	Specificity	Sensitivity	Dice
HC	0.998 ± 0.0005	0.946 ± 0.0107	0.999 ± 0.0003	0.954 ± 0.0061	0.950 ± 0.0061
AD	0.997 ± 0.0005	0.937 ± 0.0240	0.999 ± 0.0006	0.933 ± 0.0119	0.935 ± 0.0119
PD	0.998 ± 0.0004	0.937 ± 0.0134	0.999 ± 0.0004	0.963 ± 0.0045	0.955 ± 0.0045
MS	0.998 ± 0.0002	0.961 ± 0.0061	0.999 ± 0.0001	0.947 ± 0.0111	0.954 ± 0.0063
ET	0.997 ± 0.0005	0.936 ± 0.0213	0.999 ± 0.0004	0.944 ± 0.0132	0.940 ± 0.0138
Ganglion Cell Layer - Bruch's Membrane					
Class	Accuracy	Precision	Specificity	Sensitivity	Dice
HC	0.997 ± 0.0003	0.988 ± 0.0016	0.998 ± 0.0003	0.990 ± 0.0018	0.989 ± 0.0010
AD	0.994 ± 0.0028	0.987 ± 0.0015	0.998 ± 0.0002	0.963 ± 0.0224	0.975 ± 0.0010
PD	0.997 ± 0.0005	0.987 ± 0.0017	0.998 ± 0.0002	0.987 ± 0.0038	0.988 ± 0.0022
MS	0.997 ± 0.0003	0.989 ± 0.0022	0.998 ± 0.0003	0.990 ± 0.0022	0.989 ± 0.013
ET	0.996 ± 0.0003	0.986 ± 0.0052	0.998 ± 0.0008	0.987 ± 0.0011	0.986 ± 0.0022

4.1.1. Macular view

2D approach. The test results for the approach using a classical 2D setting U-Net are shown in Table 4. Particularly, MAE analysis is depicted in Table 5. As for the mean values of the RNFL class, the metric with the best values was specificity, reaching almost perfect value for all patient types. In the other metrics, we have that the patients with segmentations with the best values are PD, except in precision which are those with MS. Regarding the analysis with MAE, we find a different situation, since it is MS and ET that obtain better metrics for MAE_c and MAE_t, respectively. With the GCL-BM class we have a similar result in specificity, and in the other metrics the best results are achieved for HC and MS patients. In the MAE analysis, ET again obtains the lowest errors, this time in both MAE_c and MAE_t. The standard deviation is quite low for all these mean values, showing how robust the models were. Between the two classes, the only notable peak in this regard is in the precision metrics, which has a slightly higher deviation for RNFL than GCL-BM. For each class individually, we can see the highest deviation across patient types in sensitivity and MAE_t, whose values for RNFL range from 3.732 to 3.852 micrometers and from 3.116 to 4.056 micrometers, respectively, and for GCL-BM go from 3.852 to 3.960 micrometers and from 3.472 to 9.904 micrometers, respectively. In any case, the variations between these metrics depending on the type of patient are not too large, given their definitions.

3D multi-scale cascade approach. The test results for this approach using a classical cascade setting U-Net are shown in Table 6. Particularly, MAE analysis is depicted in Table 7. As for the RNFL class, in the first analysis, the patient group with better mean values is PD for every metric except precision, in which case MS obtains the best value. In the second analysis with MAE, we see that these two classes obtain the lowest MAE: lowest MAE_c for MS, and lowest MAE_t for PD. Regarding the GCL-BM class, in the first analysis, MS obtains the best values in all metrics except for sensitivity and Dice, where HC gets the best mean values. In the MAE metrics, MS obtains the lowest values for both MAE types. In all these analysis, the mean values have a really low standard deviation, showing how robust and stable these models were. Comparing RNFL and GCL-BM classes in terms of standard deviation, we can see notable peaks in this value for RNFL precision, sensitivity and Dice, and also slightly above in the MAE analysis. Among different patient types, we do not have high variation between the RNFL and GCL-BM metrics, neither.

Table 5
Mean MAE and corresponding standard deviation results obtained with 2D approach. Metrics are depicted in micrometers.

Retinal Nerve Fibre Layer		
Class	MAE_c	MAE_t
HC	2.300 ± 0.6532	3.536 ± 0.6616
AD	2.652 ± 0.5176	4.056 ± 0.9796
PD	2.792 ± 0.9164	3.900 ± 0.7296
MS	2.108 ± 0.2004	3.332 ± 0.3616
ET	2.140 ± 1.8732	3.116 ± 2.6180
Ganglion Cell Layer - Bruch's Membrane		
Class	MAE_c	MAE_t
HC	2.984 ± 0.2020	4.600 ± 0.4636
AD	6.176 ± 2.8628	9.904 ± 4.6592
PD	3.172 ± 0.5552	4.552 ± 0.7848
MS	2.764 ± 0.3380	4.204 ± 0.4228
ET	2.404 ± 0.2020	3.472 ± 2.6924

Table 6
Mean test and corresponding standard deviation results obtained with 3D multi-scale cascade approach.

Retinal Nerve Fibre Layer					
Class	Accuracy	Precision	Specificity	Sensitivity	Dice
HC	0.997 ± 0.0009	0.932 ± 0.0133	0.998 ± 0.0005	0.929 ± 0.0223	0.930 ± 0.0151
AD	0.997 ± 0.0009	0.920 ± 0.0307	0.998 ± 0.0006	0.931 ± 0.0267	0.925 ± 0.0231
PD	0.997 ± 0.0005	0.920 ± 0.0140	0.999 ± 0.0003	0.934 ± 0.0158	0.941 ± 0.0106
MS	0.997 ± 0.0007	0.934 ± 0.0292	0.999 ± 0.0004	0.925 ± 0.0258	0.929 ± 0.0167
ET	0.997 ± 0.0008	0.928 ± 0.0308	0.998 ± 0.0006	0.923 ± 0.0196	0.925 ± 0.0213
Ganglion Cell Layer - Bruch's Membrane					
Class	Accuracy	Precision	Specificity	Sensitivity	Dice
HC	0.996 ± 0.0004	0.984 ± 0.0019	0.998 ± 0.0003	0.988 ± 0.0024	0.986 ± 0.0015
AD	0.996 ± 0.0006	0.984 ± 0.0021	0.998 ± 0.0003	0.984 ± 0.0032	0.984 ± 0.0015
PD	0.996 ± 0.0006	0.984 ± 0.003	0.998 ± 0.0005	0.986 ± 0.0015	0.985 ± 0.0025
MS	0.996 ± 0.0008	0.986 ± 0.0053	0.998 ± 0.0008	0.987 ± 0.0039	0.986 ± 0.0029
ET	0.996 ± 0.0008	0.982 ± 0.0033	0.997 ± 0.0005	0.986 ± 0.0051	0.984 ± 0.0039

Table 7
Mean MAE and corresponding standard deviation results obtained with 3D multi-scale cascade approach. Metrics are depicted in micrometers.

Retinal Nerve Fibre Layer		
Class	MAE_c	MAE_t
HC	3.436 ± 1.0220	4.496 ± 1.1464
AD	3.664 ± 1.1232	4.376 ± 0.6352
PD	3.264 ± 0.8180	4.200 ± 0.4152
MS	3.212 ± 0.7944	4.248 ± 0.6688
ET	3.720 ± 1.3048	4.464 ± 1.1808
Ganglion Cell Layer - Bruch's Membrane		
Class	MAE_c	MAE_t
HC	3.768 ± 0.4324	5.852 ± 0.7088
AD	4.156 ± 0.5560	5.932 ± 0.7080
PD	3.772 ± 0.4824	5.816 ± 0.7808
MS	3.524 ± 0.7464	5.276 ± 0.7068
ET	3.820 ± 0.4324	5.604 ± 1.3040

3D full resolution approach. The results for the 3D setting U-Net are shown in Table 8 and, in particular, MAE analysis is depicted in Table 9. For the RNFL class, the best mean values were obtained for specificity. The remaining metrics provided the best results mainly for PD, except for precision where ET had the best results. In the MAE analysis, ET had also the best results. Regarding the GCL-BM class, the situation is similar for specificity, although there is not a clear class favored by the segmentations, since all are pretty close and there is not a clear winner in each metric. In the MAE metrics we have ET having the best MAE_c, and PD with the best MAE_t. In these cases, the standard deviation is quite low again, with a peak in precision between RNFL and GCL-BM. Particularly, RNFL had the highest peak of deviation at precision, with values ranging from 0.942 to 0.960, and GCL-BM at sensitivity with values between 0.979 and 0.989. In both cases in the MAE analysis, MAE_t had the highest standard deviation with values going from 3.292 to 4.000 micrometers for RNFL, and from 4.380 to 8.984 micrometers for GCL-BM. Again, the differences among these metrics were not quite significant.

Discussion. The cascade approach produced good results, with a mean Dice 0.930 ± 0.0124 for RNFL and 0.985 ± 0.0008 for GCL-BM. These metrics indicate successful model performance for both classes and robust models across all repetitions. However, these results do not outperform the remaining approaches, probably because cascade approach is thought to be used with higher-dimensional 3D images. In our case, our images are not that big, so when downsizing them, not all potential is extracted from this approach due to image sizes. This might happen because, when downsizing the images, too much information is lost due to the small initial size. Particularly, the bigger layer GCL-BM is better represented because it has more pixels, so when downsizing, not as much information as with the RNFL is lost. This aligns well with the nnU-Net methodology, which inherently considers the image size used as not suitable for this type of analysis.

In general, the 2D approach provided quite satisfactory results that could be summarized with a mean Dice 0.947 ± 0.0003 for RNFL and 0.985 ± 7.392e−5 for GCL-BM. Not only is this a significant metric of successful model performance, but also the low standard deviation indicates the robustness of the models. Regarding the metrics themselves, the GCL-BM class performs better probably because it is the major class, ignoring the background. In contrast, RNFL is much smaller and thinner, which makes it more variable and therefore its shape is more affected at each cut.

In the 3D scenario, the trend in these results is not only maintained, but improved compared to the 2D approach, as Dice scores with values 0.950 ± 0.0008 and 0.987 ± 0.0024 were obtained for RNFL and GLC-BM, respectively. The improvement 3D provides over 2D might be given by the addition of the 3D neighboring information to the training, so consecutive slices information improve the segmentation. However, the metrics increase being small can happen due to the large distance between slices. If the slices were closer, the neighboring information could be more related and provide a even more accurate segmentation. Related to this may also be the increase in the standard deviation for the GCL-BM class. In addition, the density needed for each pathology and layer to be segmented could be different, as it can be observed that for AD the best approach is usually 3D, while for PD and MS all are more or less equal. This could lead one to think that the approach of the acquisition itself could be relevant when trying to segment the layers of a patient's particular pathology.

With regard to where these results are placed in relation to other work in the same domain, at the time of writing, there is only one published work on this topic [24]. Overall, our results from the 2D and 3D approaches outperform the published ones, and to this end, Table 10 shows which approaches in this work produced the best metrics for each segmented class and each patient type. In general, 3D tends to provide the best results, except for specific cases such as HC, where 2D has produced the best metrics in all cases. In the MAE analysis it can also be seen that 2D and 3D tie in the number of cases where they produce the best results. As mentioned above, this may be because the near pixel information incorporated in the 3D approach has the potential to outperform the 2D approaches, but in our case the slices are not sufficiently dense for the improvement to be constant in all aspects. With respect to the paper taken as a reference, the U-Net used and configured by nnU-Net has generally produced better results than the MGU-Net used in that paper. Although it is a two-phase architecture specialized in layer segmentation, in this case our automatically configured U-Net has generally produced better results, which could be due to the fact that the automatic configuration of nnU-Net works in a satisfactory way, managing to find the best parameters. In the case of the 3D approach, the improvement is again probably due to the incorporation of the neighborhood information by treating the OCT slices as a single cube. This behavior has been seen not only in

Table 8
Mean test and corresponding standard deviation results obtained with 3D approach.

Retinal Nerve Fibre Layer					
Class	Accuracy	Precision	Specificity	Sensitivity	Dice
HC	0.997 ± 0.0008	0.942 ± 0.0125	0.999 ± 0.0004	0.951 ± 0.0183	0.946 ± 0.0089
AD	0.998 ± 0.0004	0.946 ± 0.0271	0.999 ± 0.0005	0.945 ± 0.0133	0.945 ± 0.0143
PD	0.998 ± 0.0004	0.946 ± 0.0035	0.999 ± 0.0002	0.958 ± 0.0059	0.957 ± 0.0037
MS	0.998 ± 0.0003	0.944 ± 0.0309	0.999 ± 0.0004	0.954 ± 0.0043	0.948 ± 0.0155
ET	0.998 ± 0.0003	0.960 ± 0.0118	0.999 ± 0.0003	0.947 ± 0.0177	0.953 ± 0.0071
Ganglion Cell Layer - Bruch's Membrane					
Class	Accuracy	Precision	Specificity	Sensitivity	Dice
HC	0.997 ± 0.0003	0.987 ± 0.0022	0.998 ± 0.0003	0.988 ± 0.0017	0.988 ± 0.0008
AD	0.996 ± 0.0019	0.987 ± 0.0028	0.998 ± 0.0004	0.979 ± 0.0149	0.983 ± 0.0008
PD	0.997 ± 0.0004	0.987 ± 0.0024	0.998 ± 0.0004	0.989 ± 0.0037	0.989 ± 0.0018
MS	0.997 ± 0.0005	0.990 ± 0.0021	0.998 ± 0.0003	0.988 ± 0.0036	0.989 ± 0.0020
ET	0.997 ± 0.0005	0.987 ± 0.0030	0.998 ± 0.0005	0.989 ± 0.0026	0.988 ± 0.0019

Table 9
Mean MAE and corresponding standard deviation results obtained with 3D approach. Metrics are depicted in micrometers.

Retinal Nerve Fibre Layer		
Class	MAE_c	MAE_t
HC	2.488 ± 1.9780	3.960 ± 1.0784
AD	2.472 ± 0.4000	4.000 ± 0.5360
PD	2.140 ± 0.3504	3.452 ± 0.4956
MS	2.288 ± 0.3100	3.620 ± 0.5852
ET	2.020 ± 0.4488	3.292 ± 0.6552
Ganglion Cell Layer - Bruch's Membrane		
Class	MAE_c	MAE_t
HC	3.136 ± 0.2780	4.984 ± 0.4428
AD	4.172 ± 1.9320	6.428 ± 3.6192
PD	2.948 ± 0.4072	4.380 ± 0.4220
MS	2.944 ± 0.5112	4.404 ± 0.7312
ET	2.912 ± 0.2780	4.408 ± 0.9596

relation to the previously published paper, but also to our own 2D configuration, which proves that this kind of approach has enough potential to produce even better results than those presented, which are already satisfactory.

4.1.2. Optic disc view

At the beginning of this approach, we trained a model with 3D architecture with macula-centered images, for which it obtained the 0.925 and 0.964 Dice for classes RNFL and GCL-BM, respectively in test. However, this metric was reduced to 0.765 and 0.840 for those same classes when using optic disc-centered images, thus reinforcing the need for specialized training on them to obtain performance comparable to that obtained on macula-centered images. To this end, we tested two approaches: training from scratch during 1000 epochs, and training initializing with the weights from the macula-centered model mentioned during 500 epochs, using for each approach different training sizes. The results obtained are depicted in Fig. 8 for RNFL and GCL-BM class. In both cases, the metric improves as the number of samples is increased for both types of training. For RNFL, we have a similar evolution for both training types, being always below transfer learning but very close up to the largest set, except for some small peaks in the 32 and 40 sample sets. Regarding the standard deviation, transfer learning is generally slightly more unstable, but both maintain a constant deviation at all sizes. For GCL-BM we have a similar increasing evolution, but with more pronounced peaks in medium sets and with transfer learning training being better than scratch more often than the previous case. Regarding the standard deviation, we have a more unstable situation, with pronounced peaks. Also, unlike the RNFL class, here we have that the training from scratch is slightly more unstable than the one with transfer learning.

Table 10
Metrics obtained for each patient class and retinal layer by the reference paper [24] and our 2D and 3D macular approaches. MAE metrics are shown in pixel units for comparison.

Retinal Nerve Fibre Layer				
Class	Dice	MAE_c	MAE_t	Paper
HC	0.943 ± 0.001	1.290 ± 0.0300	2.070 ± 0.0600	Ref
	0.950 ± 0.0061	**0.575 ± 0.1633**	**0.884 ± 0.1654**	Ours (2D)
	0.946 ± 0.0089	0.622 ± 0.4945	0.990 ± 0.2696	Ours (3D)
AD	0.926 ± 0.005	1.940 ± 0.2000	2.010 ± 0.1700	Ref
	0.935 ± 0.0119	0.663 ± 0.1294	1.014 ± 0.2449	Ours (2D)
	0.945 ± 0.0143	**0.618 ± 0.1000**	**1.000 ± 0.1340**	Ours (3D)
ET	0.943 ± 0.0020	1.310 ± 0.0500	1.980 ± 0.0700	Ref
	0.940 ± 0.0138	0.535 ± 0.4683	**0.779 ± 0.6545**	Ours (2D)
	0.953 ± 0.0071	**0.505 ± 0.1122**	0.823 ± 0.1638	Ours (3D)
MS	0.947 ± 0.0020	1.180 ± 0.0300	1.810 ± 0.0400	Ref
	0.954 ± 0.0063	**0.527 ± 0.0501**	**0.833 ± 0.0904**	Ours (2D)
	0.948 ± 0.0155	0.572 ± 0.0775	0.905 ± 0.1463	Ours (3D)
PD	0.949 ± 0.0020	2.730 ± 0.6000	2.040 ± 0.0800	Ref
	0.955 ± 0.0045	0.698 ± 0.2291	0.975 ± 0.1824	Ours (2D)
	0.957 ± 0.0037	**0.535 ± 0.0876**	**0.863 ± 0.1239**	Ours (3D)
Ganglion Cell Layer - Bruch's Membrane				
Class	Dice	MAE_c	MAE_t	Paper
HC	**0.988 ± 0.0000**	1.590 ± 0.0300	2.430 ± 0.0600	Ref
	0.989 ± 0.0010	**0.746 ± 0.0505**	**1.150 ± 0.1159**	Ours (2D)
	0.988 ± 0.0008	0.784 ± 0.0695	1.246 ± 0.1107	Ours (3D)
AD	0.978 ± 0.0010	2.690 ± 0.1400	4.200 ± 0.2500	Ref
	0.975 ± 0.0010	1.544 ± 0.7157	2.476 ± 1.1648	Ours (2D)
	0.983 ± 0.0008	**1.043 ± 0.4830**	**1.607 ± 0.9048**	Ours (3D)
ET	0.986 ± 0.0000	1.810 ± 0.0600	2.760 ± 0.1000	Ref
	0.986 ± 0.0022	**0.601 ± 0.0505**	**0.868 ± 0.6731**	Ours (2D)
	0.988 ± 0.0019	0.728 ± 0.0695	1.102 ± 0.2399	Ours (3D)
MS	**0.989 ± 0.0000**	1.520 ± 0.0400	2.300 ± 0.0500	Ref
	0.989 ± 0.0130	**0.691 ± 0.0845**	**1.051 ± 0.1057**	Ours (2D)
	0.989 ± 0.0020	0.736 ± 0.1278	1.101 ± 0.1828	Ours (3D)
PD	0.986 ± 0.0000	1.810 ± 0.0600	2.760 ± 0.1000	Ref
	0.988 ± 0.0022	0.793 ± 0.1388	1.138 ± 0.1962	Ours (2D)
	0.989 ± 0.0018	**0.737 ± 0.1018**	**1.095 ± 0.1055**	Ours (3D)

For the sake of simplicity, the results by class are presented in a series of graphs. Taking as reference the Dice and MAE metrics, Figs. 9 and 10 show the changes for each class considered and for each patient type in the training from scratch and with transfer learning, respectively. Regarding the training from scratch and comparing RNFL to GCL-BM, we have better and more stable metric. In MAE_c, the differences are not that clearly seen, but for Dice and MAE_t, we have flatter and more stable evolutions. For each of this metrics, we have a sense of how each of the patient types perform. We have AD as the typically worst performing NDD class. While in the smaller sets is not

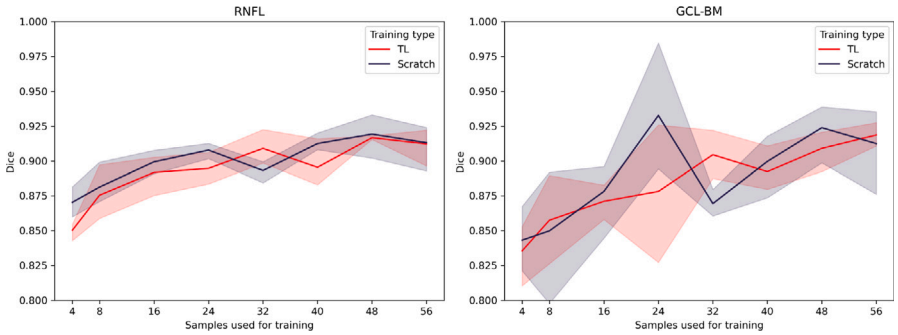

Fig. 8. Mean Dice and STD evolution for the RNFL and GCL-BM class using training from scratch and transfer learning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

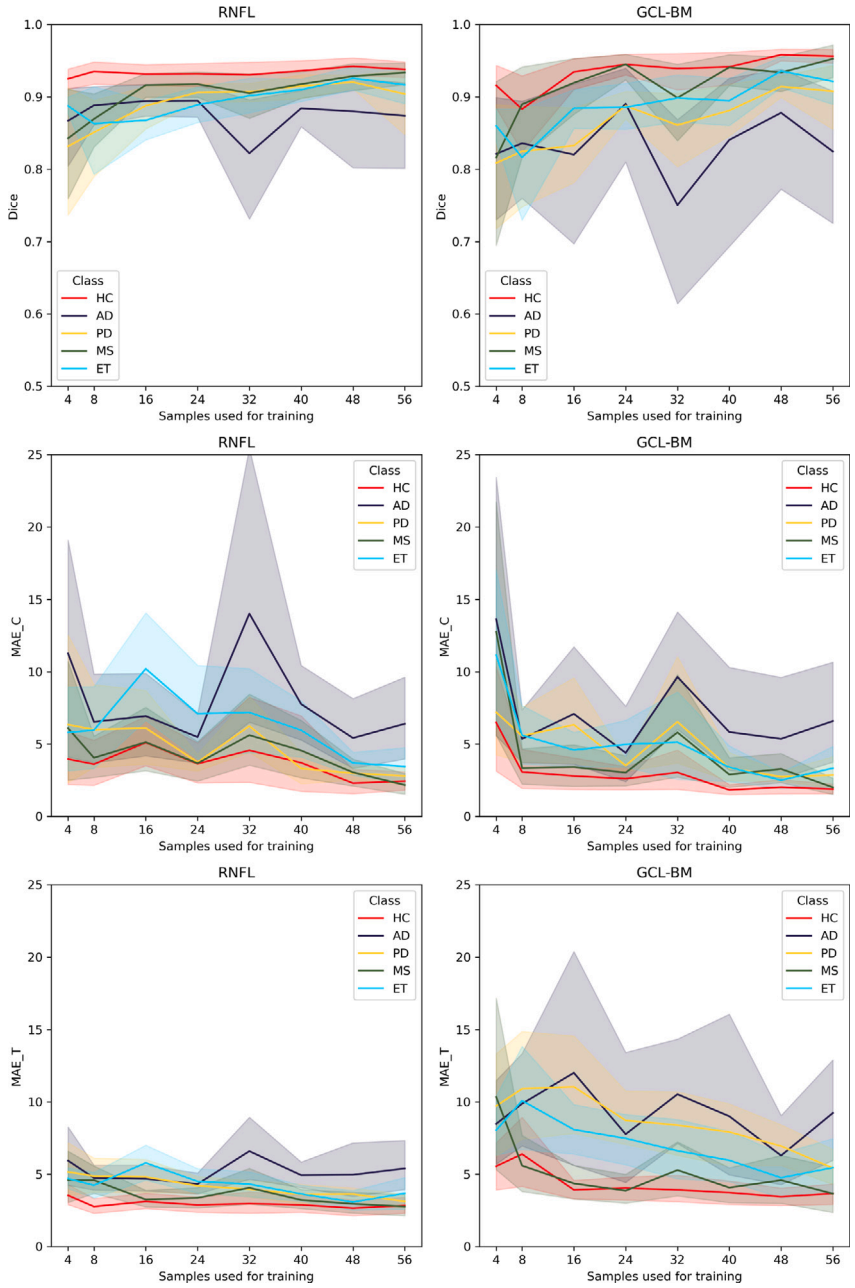

Fig. 9. Dice and MAE metrics obtained for each class and patient type using different training set sizes and training from scratch. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

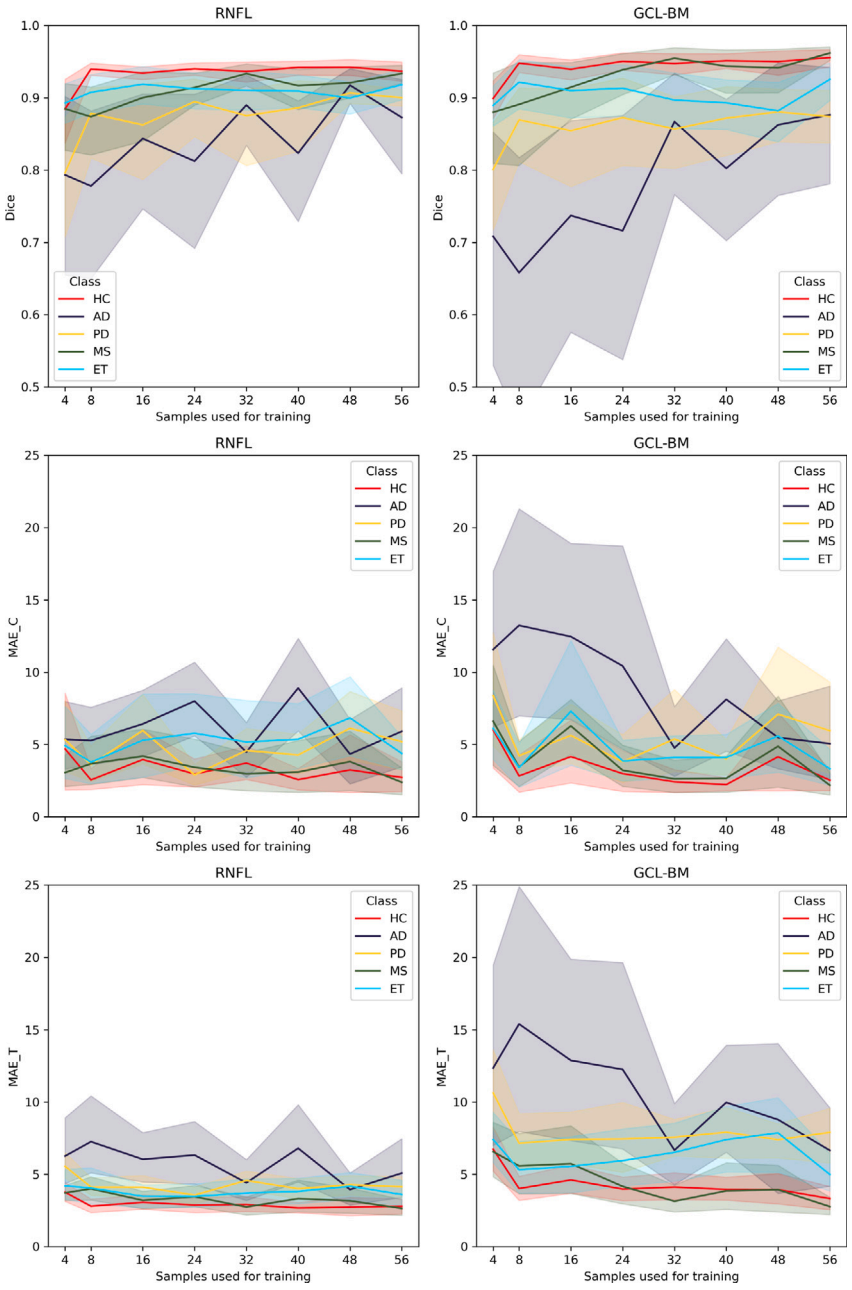

Fig. 10. Dice and MAE metrics obtained for each class and patient type using different training set sizes and training with transfer learning. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that pronounced, it rapidly gets the biggest errors and highest instability, and gets little to no improvement with bigger sets, being highly far from the other classes. In the biggest set, we can see that in most cases is far away from other NDDs and HC results. PD has sometimes a similar evolution, having the second worst results in the middle sets, but in the bigger ones reaches the performance of other classes, like ET and MS. ET does not usually have clearly the best performance, but has always an increasing upgrade the bigger the set, surpassing at some points HC and MS, which are usually the top performing classes. MS is typically the best performing NDD, with increasingly better results the bigger the sets and getting to the point HC reaches while other NDDs do not. In general, HC has the best metrics and although in some graphs we see an improvement with increasing sizes, its upgrade is way less noticeable. As for the training with transfer learning, comparing RNFL to GCL-BM, we have again that RNFL is usually more stable and does not present an increase in performance as pronounced as GCL-BM does, which typically has more noticeable and unstable evolutions. The patient types evolve in a similar fashion as in the training from scratch, but with smaller differences in the biggest sets. AD is again the worst class, with clearer improvements with training size but always reaching the best metrics in the biggest sets, as the other NDDs do. PD has a similar evolution to AD, but is not that far away from the other classes as AD. ET usually has a better performance than PD, although not as satisfactory as MS or HC, and it shows a concerning slight decreasing in performance tendency as the set size increases. Then MS is the best NDD, quite close to HC. It has a slighter improvement with increasing size. HC is not far from MS, but it shows a more clearer tendency to be top performing class in every set, as although it does not clearly change from one set to another, it gets quite satisfactory metrics at each point for each layer class and metric. In general, all these evolutions are more pronounced in the smaller sets for the transfer learning approach, until

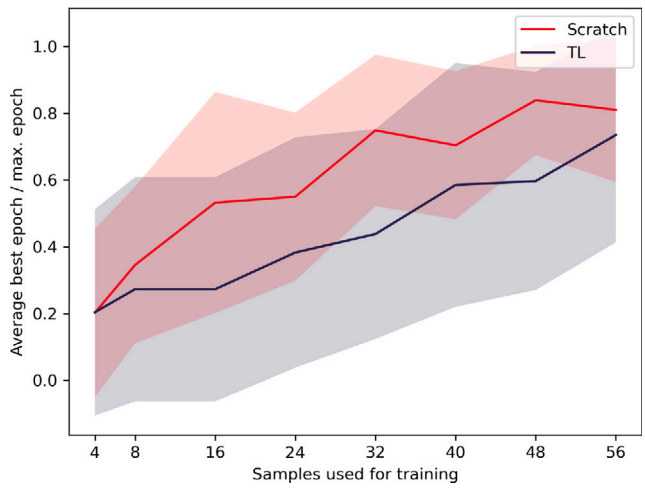

Fig. 11. Best epoch for each training type at each training size. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

it reaches some stability in the bigger sets, with similar metrics to the training from scratch.

As for the benefits transfer learning might show during traning, Fig. 11 shows the average best epoch per traning size used for each tranining approach. On the one hand, we see how the better model is achieved later the larger the training set for both cases. On the other hand, it is observed that the best model is always obtained earlier using transfer learning.

Discussion. In the transfer learning and training from scratch comparison, we saw that transfer learning did not provide any improvement metric nor stability-wise, in general. Dice metrics were quite close, but mainly under the scratch type for every training size, showing that its special initialization did not provide any benefit in the end test results. This could be due to the fact that, although related, macula-centered and optic disc-centered scans are too dissimilar, especially because of the abundant presence of blood vessels in optic disc scans. However, interesting features are transferred, since the transfer learning models converge earlier, as can be seen in Fig. 11 where it is shown in which epoch the best model is obtained for each type of training. Since these transfer learning models are obtained in half the number of epochs, this technique provides models almost as satisfactory as those trained by scratch in half the time.

Another detail that we can observe in these results are that here, unlike in the macula-centered scans, the class that obtains the best result is the RNFL in most cases with a top 3% improvement in Dice. Also, GCL-BM usually has more noticeable increments in performance as the training set increases, while RNFL shows slighter upgrades. One reason for this behavior could again be the presence of blood vessels. They manifest themselves in the images as distortions that dilute the edges of the different layers, and the bottom edge of the retinal layers, so it impacts more the GCL-BM layer. This translates into it being more difficult to determine the edges of the retinal layers both between them and with respect to the background, as the artifacts produced by the vessels are similar to some that manifest in the background.

It is also interesting to note that, despite the slight differences in performance between one type of training and the other, the same tendencies are observed in both with respect to each type of patient. AD has a markedly worse performance. It is true that it improves quite a lot as samples are included, since in small sets it has notably worse metrics, which means a much higher increase in performance than in the other cases. However, in very few occasions it manages to achieve a performance similar to the other classes, even with models trained with the complete data set. ET and PD are in an intermediate position: they are not consistently worse than AD, but neither do they consistently achieve the best metrics. In fact, PD has some tendency to follow a similar evolution as AD at certain times, although it tends to stabilize for large sets in metrics more similar to the other classes, which offers a better performance. ET is in this sense superior to PD because it reaches this stability earlier and even reaches higher metrics than the other classes in medium sets. However, the classes that do so consistently are HC and in second place MS. MS needs to increase the number of training samples before it stabilizes at metrics very close to the best obtained, which are those achieved by HC. In this sense, HC does not have an overly steep improvement curve, but tends to remain stable at the best results regardless of the number of samples in the training set. There could be a number of reasons for AD performing notably worse. Given that the samples were selected in such a way that the most recent ones were taken, we are dealing with patients with 0 to 7 years of evolution with their disease, however the distribution of this evolution length is not uniform in all classes: the mean time difference for each scan with respect to the first one is for each class: AD 0.069 ± 0.371, ET 2.100 ± 3.381, MS 4.615 ± 3.121, PD 1.976 ± 2.529. Here, we can see that AD presents the most different mean, so it could be related to its worse performance. Also, as mentioned above, these layers show some deterioration due to the disease, but it is conceivable that it could affect also the retinal blood vessels, given the cerebrovascular component that many of these NDDs can have [56]. Given the effect that these vessels have on these images and their consequent segmentation, a relationship between the patient's deterioration, the presence of vessels and the differences between types of patients could be suspected, especially when noting the difference between NDD class and HC. This could particularly concern those patients more affected by the thinning of the retinal layers due to their disease: if the layers are smaller and also the presence of vessels can be exacerbated, patients with a certain disease could have worse results when analyzing their layers, thus producing biases such as those found in these results, both when training the models from scratch and with transfer learning.

In a visual inspection, we can confirm the distortions of the vessels, which might end up producing in the final segmentations gaps and errors in the layers with respect to which layer goes on top of which. Given that in the subsequent phase we work with thickness maps, these gaps would cause problems when analyzing the thicknesses of each layer. Therefore, the results were processed using classical computer vision techniques to fill these gaps and eliminate small errors due to artifacts. An example of a post-processing can be seen in Fig. 12, and as an objective reference of the improvement of this extra step, Table 11 shows the metrics before and after post-processing the results of the models trained from scratch.

In relation with the state of the art, these results are not exactly comparable to those mentioned for macular view, as the type of scan is different. However, we can see that they are inline with them, or similar to other works that focus on retinal layer segmentations in different contexts, since Dice is typically above 0.9 [57–59].

4.2. Neurological diseases screening

In this section, the results for the second task, neurological diseases screening, are shown for each configuration. The firsts are single view, where only optic disc or macular view are considered. Then, the last one joins these two configurations into a multi-view approach.

4.2.1. Single-view: Optic disc

Taking as reference F1-Score for being better at considering the class imbalance, Fig. 13 shows we have obtained poor results for every screening approach, being the best markers RNFL and All layers in MSvsNDDS, but only with slightly above 0.6 metrics. The AllvsAll approach was the worst by far, indicating the poor ability of this type of information to separate these classes on its own. The best markers for

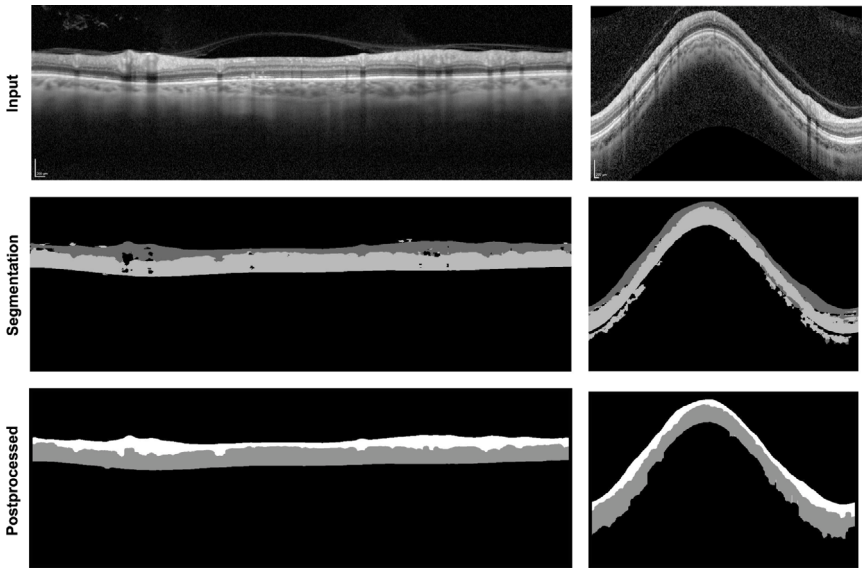

Fig. 12. Examples of optic-disc-centered segmentations before and after post-processing.

Table 11
Mean metrics before and after the postprocessing of the segmented results of the models trained from scratch. MAE metrics are given in micrometers.

	RNFL		GCL-BM	
	Before	After	Before	After
Accuracy	0.989 ± 0.0019	0.991 ± 0.0012	0.979 ± 0.0068	0.982 ± 0.0089
Precision	0.906 ± 0.0227	0.917 ± 0.0110	0.934 ± 0.0354	0.921 ± 0.0513
Specificity	0.995 ± 0.0015	0.995 ± 0.0008	0.993 ± 0.0048	0.990 ± 0.0100
Sensitivity	0.899 ± 0.0260	0.920 ± 0.0226	0.855 ± 0.0447	0.916 ± 0.0276
Dice	0.900 ± 0.0191	0.916 ± 0.0143	0.886 ± 0.0368	0.912 ± 0.0342
MAE_c	21.676 ± 9.9668	10.040 ± 3.8084	4.701 ± 3.0210	4.033 ± 2.2117
MAE_t	16.204 ± 2.7604	14.656 ± 1.8104	6.902 ± 1.6903	6.544 ± 3.7968

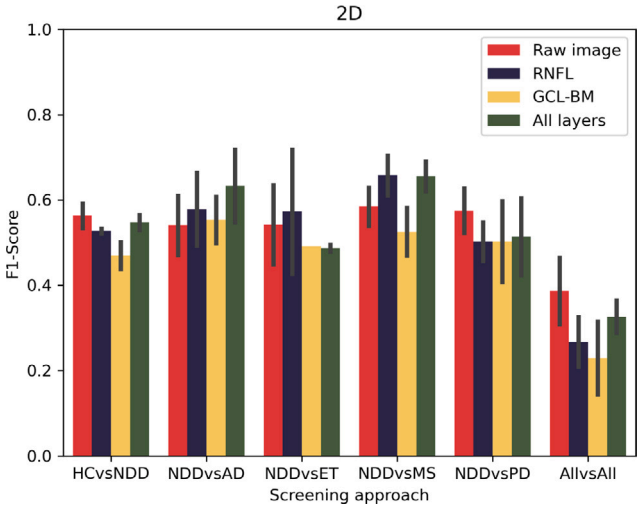

Fig. 13. F1-Score for each screening approach, considering different types of information extracted from optic-disc-centered scans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

each approach are raw image, all layers or RNFL thickness. Although in general there is not a great difference among them, in the NDDvsET that has a huge difference in the standard deviation of RNFL and GCL-BM. This last type of information, GCL-BM, did not perform noticeably well for any approach, but again the differences among different types of information are not great.

4.2.2. Single-view: Macula

Fig. 14 shows the results obtained for this method, using different types of information and considering 2D, 3D and the combination of both. For 2D, AllvsAll is the worst approach, not even reaching the 0.4 mark. Then we have NDDvsAD, which barely reaches the 0.5 metric, not having any particularly type of information perform better than the others. Then, HCvsNDD reaches the 0.6 mark using all layers thickness information. In NDDvsET we have the raw image information performing specially well among the different types reaching a 0.7 metric, but with a quite high standard deviation. In this approach, the remaining types of information performed considerably worse, barely reaching 0.5. MS and PD reached 0.6 in a more stable fashion with every type of information. Using 3D information instead, we still have AllvsAll being the worst approach. In this case, NDDvsET barely reaches 0.5 and has a high instability, specially for raw image and RNFL information, while the other approaches go from 0.6 to 0.8. NDDvsMS reached 0.8 more comfortably, specially using all layers thickness information, with low standard deviation. Second would be NDDvsPD, whose values go from 0.7 to 0.8 for all layers and GCL-BM information, respectively. Combining 2D and 2D+3D information, we have a similar scenario to 3D case. AllvsAll is the worst, while MS and PD are the best ones, reaching values above 0.7. Standard deviation is more stable, reaching peaks at NDDvsET using all layers thickness.

4.2.3. Multi-view

Fig. 15 shows the results per approach in the multi-view method. With 2D information, all approaches get at least a 0.5 metric, but the AllvsAll approach. HCvsNDD and some types of information for NDDvsET seem to be the worst cases, with barely 0.5, and having high standard deviation for ET. The remaining cases seem to be closer

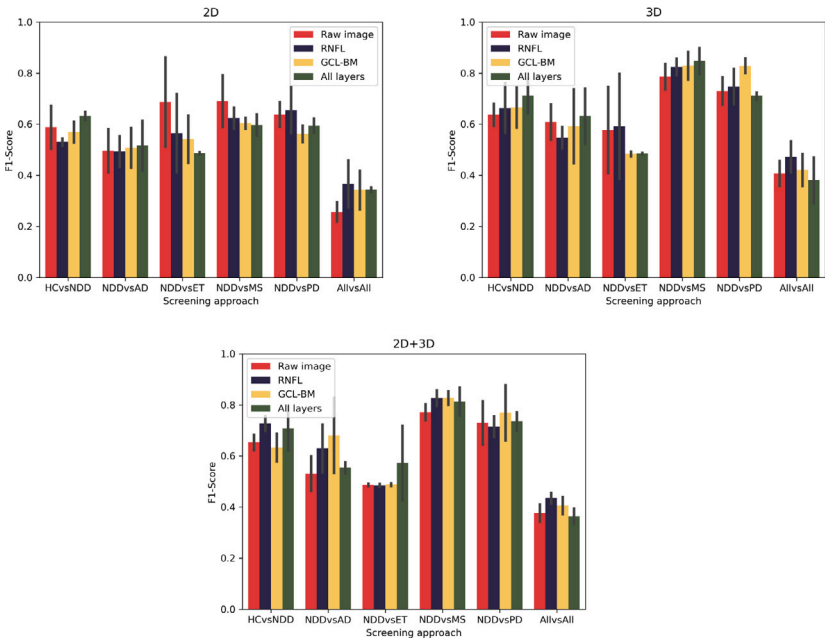

Fig. 14. F1-Score for each screening approach, considering different types of information extracted from macula-centered scans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

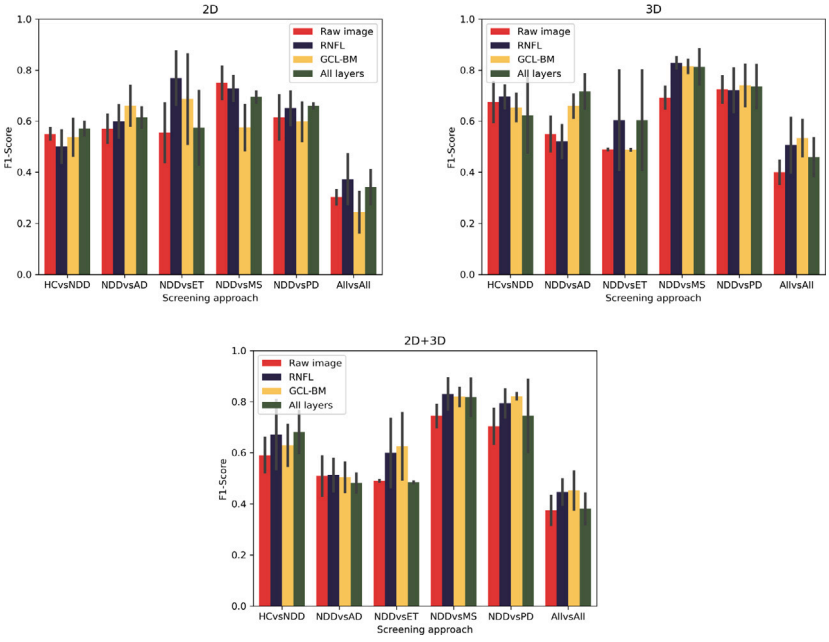

Fig. 15. F1-Score for each screening approach, considering different types of information extracted from both optic-disc-centered and macula-centered scans. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of 0.6, but for RNFL information in the NDDVSET approach, which is closer to 0.7, although with a high standard deviation too. Using 3D information we have a similar situation, but with lower standard deviation and metrics being more consistently above 0.6 for those that are not NDDvsET or AllvsAll. In the combination, the same situation happens again, but with even lower standard deviation.

4.2.4. Discussion

We can analyze the previous results though three different lenses: single and multi-view: 2D and 3D; different computational biomarkers extracted; and disease separability.

First, in Fig. 16 we have summarized the differences between the single-view and multi-view approaches for the 2D, 3D and combination scenarios. We can see a noticeable improvement of 2D multi-view with respect to optic disc and macula single-view methods, specially for the first one. We have a high standard deviation for the MS approach in both, but since this was the best approach in general, the changes in the metric still reflect a satisfactory performance, just not as improved as others when combining views, probably because it might need the information to be combined in a different way to better be exploited. In the 3D multi-view with respect to the 3D macular view, we do not have such a clear improvement. In fact, both perform similarly, considering the approaches and information types where one performs better than the other. Again, we have a high standard deviation in the MS approach. In the combination of 2D and 3D information, we have a similar situation, where we have some improvements for some

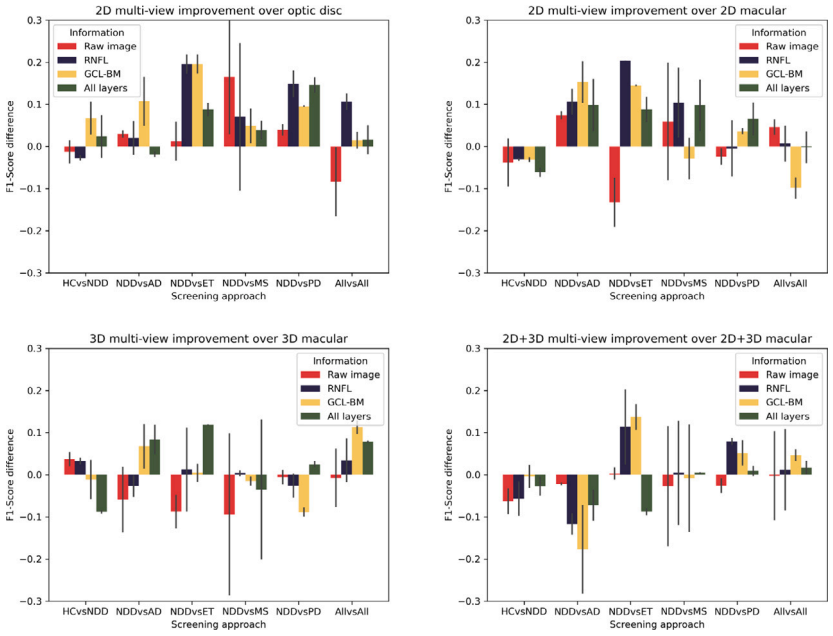

Fig. 16. F1-Score changes when comparing single-view approaches to multi-view. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

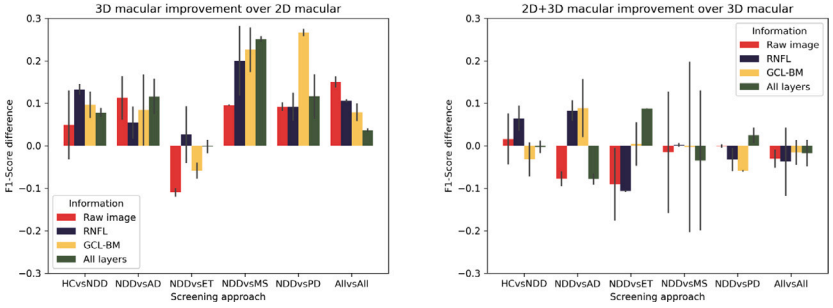

Fig. 17. F1-Score changes when comparing approaches using 2D, 3D and 2D+3D information in the macular single-view method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

approaches, and other remain the same, with high standard deviation for the MS approach. From all this, we can extract that combining multiple views does improve the performance of the system to correctly classify different NDDs. However, this improvement is not linear when adding 3D information. Including 3D information or even mixing it with 2D information might need a more specific information fusion technique than only concatenating information. In any case, this proves the potential of the multi-view method for this type of medical image.

Figs. 17 and 18 shows a different comparison, centered this time in the 2D and 3D information considered in every experiment using macular single-view and multi-view methods, respectively. In the first method, we can see clearly how using 3D information considerably improved the performance of the system, especially for the MS and PD approaches. ET one has a worse performance, as it was seen previously in every considered scenario. However, the remaining approaches experiment a notable upgrade. In the combined 2D+3D approach with respect to only 3D, the improvement is not that clear, as the improvements are of less magnitude and there is much higher instability. In the multi-view method comparison, a similar improvement is seen. As mentioned, this might show the potential this combination has, but it fails at some point, probably due to the way the information is combined. A more precise technique could be applied to merge the 2D and 3D information, instead of simple concatenation.

Regarding the different types of computational biomarkers considered in the experiments, we show in Fig. 19 the best one for each method and approach. It should be noted that the differences among the different information considered are not large, so the best one might not be outperforming the remaining ones by much. However, this summary is enough to get a first glance at the more adequate computational biomarkers for each method and approach. In general, there is no clear winner. HC uses better all layers and RNFL information. AD seems to prefer all layers or GCL-BM, while ET does for RNFL and all layers. PD prefers to center on GCL-BM in 3D approaches only, while this tendency is not clear in the 2D ones. All for all clearly prefers RNFL information in most of the cases. We can see that raw images only outperform in 2D approaches, while they are not satisfactory enough in 3D ones. Again, there is no clear predominance of one marker for each approach, which makes sense given the instability that some of them have presented and the possible problems in combining information given to the classifier. However, if we interpret these results with reservation, interesting correlations to clinical biomarkers are presented. AD patients typically show an overall thinning of the whole retinal layers [60–63], which can be represented as the biggest-area computational biomarkers in our study, which are those comprehending all layers and GCL-BM and accordingly, the best computational markers for the NDDvsAD approach. ET patients suffer from a thinning in the RNFL and GCL layers, as well as a choroid thickening [64–67]. In our results we can see that RNFL behavior reflected, as it was the best marker for many scenarios with NDDvsET approach. MS affects in a similar manner RNFL [36,68–70], but that aspect is not as clear in our best computational biomarkers

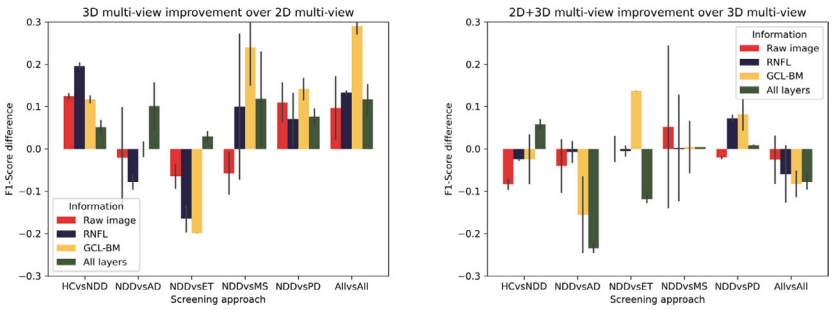

Fig. 18. F1-Score changes when comparing approaches using 2D, 3D and 2D+3D information in the multi-view method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

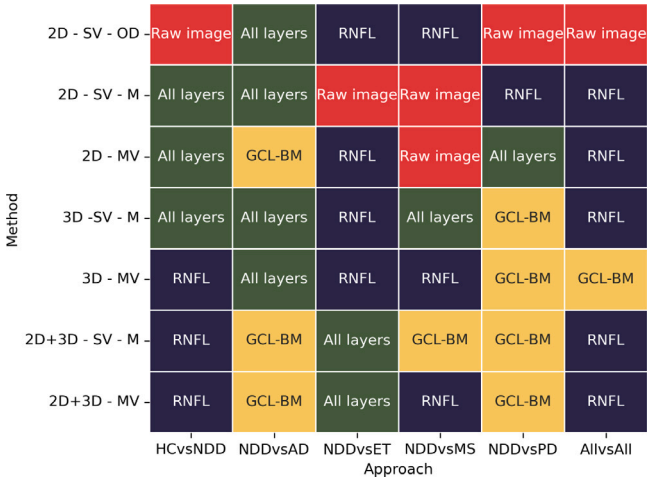

Fig. 19. Computational biomarker with the best mean F1-Score per method and screening approach. SV, MV, OD and M stand for "single-view", "multi-view", "optic disc" and "macula", respectively.

results. Maybe the variety of MS types, complications and affections might make it too heterogeneous to be easily differentiated to a NDD cohort as in the NDDvsMS approach. PD shows a thinning in the GCL and IPL layers, as well as an overall reduction in the macular thickness [71–75]. Hence, our markers accurately represent that behavior in the NDDvsPD scenario. The combination of these findings propose RNFL and overall retinal thickness as powerful indicators of the presence of these diseases in general, which at the same time is coherent with the results in HCvsNDD and AllvsAll approaches, where RNFL and retinal markers are the clear best performance measures. Therefore, the proposed computational biomarkers show a behavior consistent with the state of the art of these pathologies.

In all these comparisons, we have also seen how the different screening approaches perform. There is a common scenario in all these methods where AllvsAll is the worst approach, followed by ADvsNDD and ETvsNDD. HCvsNDD one usually performs slightly better than these two, but the best approaches are those that considered how separable are MS or PD. The reason that these three approaches stand out from the others is probably because these three classes are also by far the most numerous. Although we have tried to balance the weights in the training, the lack of dimensionality is noticeable and can be seen in the observed difference in performance. Regarding the differences between these three, it seems that the selected computational biomarkers are sufficient to recognize MS or PD patients, but the differentiation between pathological and healthy is more problematic. Perhaps the differences in the thickness of the layers are very noticeable between different diseases, either because some thicken it or others thin it. Maybe HC patients are a difficult middle ground to define if we consider diseases with thicknesses at both ends of the spectrum, which means that classification with only these markers is not sufficient.

In accordance with prior studies employing OCT and retinal layer segmentation in various neurodegenerative conditions, our findings align with the consensus. Specifically, we noted that data derived from the RNFL proves highly advantageous to classifiers and emerges as a dependable tool for pathology detection, demonstrating elevated sensitivity. Nevertheless, when it comes to distinguishing between distinct pathologies, the remaining segmented retinal layers contribute significantly. This can be attributed to the unique impact of each pathology on retinal layer thickness, resulting in distinct topographical alterations in these structures, involving both thinning and thickening of specific layers.

Regarding how this work is positioned withing the state of the art, as we have previously mentioned, there is no other work that uses these kind of features together to perform neurological diseases screening. To get a sense of other metrics obtained from automatic methods for classification of neurodegenerative diseases against control patients based on OCT information, we have F1-scores of 0.70 for AD based on numerical thickness values taken from macular and optic disc views [76] and 0.68 for GCL-IPL layers measured in macular area of MS [77], while the other diseases do not have any study of this kind. These metrics seem to be inline with those obtained in our experiments, both single-view and multi-view. Additionally, currently there is no independent dataset comprehending such a variety of diseases and information similar to ours, so a fair comparison to other state of the art architectures and methods cannot be done at the time being. If we compare our metrics to others obtained solving classification tasks in OCT images, we see that there is a disparity that heavily favors those from other works [78–80]. These papers focus on abnormalities much more notorious than those seen in NDD OCT images, that manifest in many cases as subtle thickness changes. This is also the main reason of our work focusing on different techniques to process the information within our samples: to enhance the fine differences between controls and pathological cases to be able to perform a clinical screening approach.

5. Conclusions

Here, we have presented pioneering work in the field of neurological diseases in relation to OCT and automated methods, as well as the extraction of relevant features for the diagnosis and treatment of patients of this type. First, we have segmented and analyzed the key retinal layers of patients of four major NDDs using two different OCT perspectives for the first time, allowing for new ways of comprehension and diagnosis in the NDD and OCT domains. In a second stage, these four diseases and the retinal layer information extracted were also studied in a novel exhaustive pathological screening, analyzing the effect of the different computational biomarkers considered in 6 different scenarios and detecting the most relevant ones for each NDD, which has never done before for these diseases.

Regarding the segmentation of retinal layers, using macula-centered scans we were able to get quite satisfactory and stable results for the

2D configuration, which were outperformed by the 3D approach. This might indicate the importance of 3D information in this type of scan to accurately segment the RNFL and GCL-BM layers. Analyzing how this segmentation performed for each patient type, we could not see one obtaining predominately better results than the other, but we did see how some, like AD, benefited more notably from the inclusion of 3D information. Changing the view to the optic-disc-centered scans, we applied transfer learning from these successful macular models to the optic disc ones, since there is a clear link between these two types of views. We tried different training sizes and we observed the same tendency in all of them: training from scratch produced a slight improvement over the transfer learning strategy. Thus, we did not obtain a huge change in performance, but we could see how it did affect the training times: using macular weights allowed for the models to converge earlier. However, unlike the macular view, in both these methods we observed that some patient types got worse results than the others. While HC and MS typically got the highest and most stable metrics, PD and specially AD obtained the worst. In the case of AD, the evolution expected to decrease when adding more samples to the training set, was instead an unstable and flat curve with mainly the worst metric for any class. This could be related to the low dimensionality of our dataset or the influence of the blood vessels present in the optic disc view. Although our scans' quality and amount could be enough for medical experts, it was not sufficient for the models to segment the layers of the AD patients as accurately as other cohorts.

As for the classification task, our metrics were not that satisfactory in many experiments, but we could observe interesting tendencies on how the different methods and computational biomarkers used affected the results. First, applying a multi-view approach improved the single view approaches, especially when used only 2D information. In the case of the 3D and 2D+3D methods, some improvement was seen, but it might require a more precise information fusion technique to deal with the feature vector extracted. Secondly, using 3D information benefited the classifier, since notable improvements were observed when comparing 3D and 2D methods. Including 3D information to 2D did not improve the results as clearly, but it again could be related to the need of a more *ad hoc* method to fuse the 2D and 3D data instead of only concatenating. Finally, regarding the different types of computational biomarkers considered in the experiments, the differences between them were not great, but we did see some tendencies towards some types in particular in some screening approaches. The most notable one was the use of GCL-BM information in the identification of PD patients in a NDD cohort. Not only was this screening approach one of best performing approaches, but also GCL-BM thickness information provided the best separability, which is in line with the medical research that points out the thinning of this layer for this type of patient.

As future work, there are various lines to follow. For the macular view segmentation, it could be interesting to check if the addition of slices to the 3D cubes would add more information that could be used to segment the layers more accurately in the 3D approach. The optic disc view could be improved by adding more samples to ensure the absence of overfitting and its negative effects to the consequent experiments, both increasing the dataset or using GAN or stable diffusion models to synthetically generate similar images. The classification task needs to improve its metrics, for which methods could be designed to merge information more accurately than by simple concatenation. Also, more combinations of the computational biomarkers herein considered could be explored to find the most efficient way to combine views, representations and markers. The feature extractor was our first step in this part of pipeline, and we chose a generalist feature extractor. The quality of these features seems to be adequate for this domain, but further experimentation could be done in terms of the use of general features applied to medical domain. The influence of the classification model is a crucial element of this part of the methodology, and further lines of work should address other methods, like CNN-based classifier. Another aspect is the improvement of the complete pipeline by following an end-to-end training strategy, which could produce potential benefits that are not considered here. Additionally, although not possible at the time of writing for the lack of comparable datasets, it would be necessary to study the performance of our method with similar datasets.

Funding

This research was funded by Instituto de Salud Carlos III, Government of Spain, [research projects PI17/01726 and PI20/00437]; Inflammatory Disease Network (RICORS) [research project RD21/0002/0050]; Ministerio de Ciencia e Innovación, Government of Spain through the research project with grant numbers [PID2023-148913OB-I00, TED2021-131201B-I00, and PDC2022-133132-I00]; Consellería de Educación, Universidade, e Formación Profesional, Xunta de Galicia, Grupos de Referencia Competitiva, [grant number ED431C 2024/33]. Furthermore, this work was supported by the Instituto de Salud Carlos III (ISCIII) under the grant [FORT23/00010] as part of the Programa FORTALECE of Ministerio de Ciencia e Innovación. The funding organisations had no role in the design or conduct of this research.

CRediT authorship contribution statement

Lorena Álvarez-Rodríguez: Writing – original draft, Visualization, Validation, Software, Methodology. **Ana Pueyo:** Methodology, Data curation, Conceptualization. **Joaquim de Moura:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Elisa Vilades:** Methodology, Data curation, Conceptualization. **Elena Garcia-Martin:** Methodology, Funding acquisition, Data curation, Conceptualization. **Clara I. Sánchez:** Supervision, Methodology, Conceptualization. **Jorge Novo:** Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization. **Marcos Ortega:** Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] de Oliveira N, Almeida MRS, Pontes FMM, Barcelos MP, de Paula da Silva CHT, Rosa JMC, et al. Antioxidant effect of flavonoids present in euterpe oleracea martius and neurodegenerative diseases: a literature review. Cent Nerv Syst Agents Med Chem 2019;19(2). http://dx.doi.org/10.2174/1871524919666190502105855.

[2] Hansson O. Biomarkers for neurodegenerative diseases. Nat Med 2021;27(6). http://dx.doi.org/10.1038/s41591-021-01382-x.

[3] Hou Y, Dan X, Babbar M, Wei Y, Hasselbalch SG, Croteau DL, et al. Ageing as a risk factor for neurodegenerative disease. Nat Rev Neurol 2019;15(10). http://dx.doi.org/10.1038/s41582-019-0244-7.

[4] Yiannopoulou KG, Papageorgiou SG. Current and future treatments in alzheimer disease: an update. J Central Nerv Syst Dis 2020;12. http://dx.doi.org/10.1177/1179573520907397.

[5] Feigin VL, Nichols E, Alam T, Bannick MS, Beghi E, Blake N, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019;18(5). http://dx.doi.org/10.1016/S1474-4422(18)30499-X.

[6] Haubenberger D, Hallett M. Essential tremor. New Engl J Med 2018;378(19). http://dx.doi.org/10.1056/NEJMcp1707928.

[7] Dobson R, Giovannoni G. Multiple sclerosis – a review. Eur J Neurol 2018;26(1). http://dx.doi.org/10.1111/ene.13819.

[8] Bradshaw J. Fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease is qualitatively distinct. J Neurol Neurosurg Psychiatry 2004;75(3). http://dx.doi.org/10.1136/jnnp.2002.002576.

[9] Mejia-Vergara AJ, Karanjia R, Sadun AA. OCT parameters of the optic nerve head and the retina as surrogate markers of brain volume in a normal population, a pilot study. J Neurol Sci 2021;420. http://dx.doi.org/10.1016/j.jns.2020.117213.

[10] Yap TE, Balendra SI, Almonte MT, Cordeiro MF. Retinal correlates of neurological disorders. Ther Adv Chronic Dis 2019;10. http://dx.doi.org/10.1177/2040622319882205.

[11] Wolf S, Wolf-Schnurrbusch U. Spectral-domain optical coherence tomography use in macular diseases: a review. Ophthalmologica 2010;224(6). http://dx.doi.org/10.1159/000313814.

[12] Elsawy A, Abdel-Mottaleb M. PIPE-Net: A pyramidal-input-parallel-encoding network for the segmentation of corneal layer interfaces in OCT images. Comput Biol Med 2022;147. http://dx.doi.org/10.1016/j.compbiomed.2022.105595.

[13] Wang L, Shen M, Shi C, Zhou Y, Chen Y, Pu J, et al. EE-Net: An edge-enhanced deep learning network for jointly identifying corneal micro-layers from optical coherence tomography. Biomed Signal Process Control 2022;71. http://dx.doi.org/10.1016/j.bspc.2021.103213.

[14] Yan Q, Chen B, Hu Y, Cheng J, Gong Y, Yang J, et al. Speckle reduction of OCT via super resolution reconstruction and its application on retinal layer segmentation. Artif Intell Med 2020;106:101871. http://dx.doi.org/10.1016/j.artmed.2020.101871.

[15] García G, del Amor R, Colomer A, Verdú-Monedero R, Morales-Sánchez J, Naranjo V. Circumpapillary OCT-focused hybrid learning for glaucoma grading using tailored prototypical neural networks. Artif Intell Med 2021;118:102132. http://dx.doi.org/10.1016/j.artmed.2021.102132.

[16] Augustin AJ, Atorf J. The value of optical coherence tomography angiography (OCT-A) in neurological diseases. Diagnostics 2022;12(2). http://dx.doi.org/10.3390/diagnostics12020468.

[17] de Eguileta AL, Cerveró A, de Sabando AR, Sánchez-Juan P, Casado A. Ganglion cell layer thinning in alzheimer's disease. Medicina 2020;56(10). http://dx.doi.org/10.3390/medicina56100553.

[18] Castro-Roger L, Palomar EV, Ciordia BC, Rodrigo MJ, Perié MS, Campo LA, et al. OCT retinal imaging as differential diagnostic tool between Parkinson disease and essential tremor. Acta Ophthalmol 2022;100(S267). http://dx.doi.org/10.1111/j.1755-3768.2022.154.

[19] Motamedi S, Gawlik K, Ayadi N, Zimmermann HG, Asseyer S, Bereuter C, et al. Normative data and minimally detectable change for inner retinal layer thicknesses using a semi-automated OCT image segmentation pipeline. Front Neurol 2019;10. http://dx.doi.org/10.3389/fneur.2019.01117.

[20] Slotnick S, Ding Y, Glazman S, Durbin M, Miri S, Selesnick I, et al. A novel retinal biomarker for Parkinson's disease: Quantifying the foveal pit with optical coherence tomography. Mov Disorders 2015;30(12). http://dx.doi.org/10.1002/mds.26411.

[21] He Y, Carass A, Liu Y, Jedynak BM, Solomon SD, Saidha S, et al. Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT. Biomed Opt Express 2019;10(10). http://dx.doi.org/10.1364/BOE.10.005042.

[22] Cavaliere C, Vilades E, Alonso-Rodríguez M, Rodrigo M, Pablo L, Miguel J, et al. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. Sensors 2019;19(23). http://dx.doi.org/10.3390/s19235323.

[23] Ortiz M, Mallen V, Boquete L, Sánchez-Morla EM, Cordón B, Vilades E, et al. Diagnosis of multiple sclerosis using optical coherence tomography supported by artificial intelligence. Mult Scler Relat Disorders 2023;74. http://dx.doi.org/10.1016/j.msard.2023.104725.

[24] Gende M, Mallen V, de Moura J, Cordón B, Garcia-Martin E, Sánchez CI, et al. Automatic segmentation of retinal layers in multiple neurodegenerative disorder scenarios. IEEE J Biomed Health Inform 2023;1–12. http://dx.doi.org/10.1109/jbhi.2023.3313392.

[25] Danesh H, Maghooli K, Dehghani A, Kafieh R. Synthetic OCT data in challenging conditions: three-dimensional OCT and presence of abnormalities. Med Biol Eng Comput 2021;60(1). http://dx.doi.org/10.1007/s11517-021-02469-w.

[26] Bogunovic H, Venhuizen F, Klimscha S, Apostolopoulos S, Bab-Hadiashar A, Bagci U, et al. RETOUCH: the retinal OCT fluid detection and segmentation benchmark and challenge. IEEE Trans Med Imaging 2019;38(8). http://dx.doi.org/10.1109/TMI.2019.2901398.

[27] Chen X, Niemeijer M, Zhang L, Lee K, Abramoff MD, Sonka M. Three-dimensional segmentation of fluid-associated abnormalities in retinal OCT: probability constrained graph-search-graph-cut. IEEE Trans Med Imaging 2012;31(8). http://dx.doi.org/10.1109/TMI.2012.2191302.

[28] Wu M, Fan W, Chen Q, Du Z, Li X, Yuan S, et al. Three-dimensional continuous max flow optimization-based serous retinal detachment segmentation in SD-OCT for central serous chorioretinopathy. Biomed Opt Express 2017;8(9). http://dx.doi.org/10.1364/BOE.8.004257.

[29] Wu M, Chen Q, He X, Li P, Fan W, Yuan S, et al. Automatic subretinal fluid segmentation of retinal SD-OCT images with neurosensory retinal detachment guided by enface fundus imaging. IEEE Trans Biomed Eng 2018;65(1). http://dx.doi.org/10.1109/TBME.2017.2695461.

[30] Maetschke S, Antony B, Ishikawa H, Wollstein G, Schuman J, Garnavi R. A feature agnostic approach for glaucoma detection in OCT volumes. PLoS One 2019;14(7). http://dx.doi.org/10.1371/journal.pone.0219126.

[31] Holmberg OG, Köhler ND, Martins T, Siedlecki J, Herold T, Keidel L, et al. Self-supervised retinal thickness prediction enables deep learning from unlabelled data to boost classification of diabetic retinopathy. Nat Mach Intell 2020;2(11). http://dx.doi.org/10.1038/s42256-020-00247-1.

[32] Mohammed S, Li T, Chen XD, Warner E, Shankar A, Abalem MF, et al. Density-based classification in diabetic retinopathy through thickness of retinal layers from optical coherence tomography. Sci Rep 2020;10(1). http://dx.doi.org/10.1038/s41598-020-72813-x.

[33] Garcia-Martin E, Ortiz M, Boquete L, Sánchez-Morla E, Barea R, Cavaliere C, et al. Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier. Comput Biol Med 2021;129. http://dx.doi.org/10.1016/j.compbiomed.2020.104165.

[34] He X, Deng Y, Fang L, Peng Q. Multi-modal retinal image classification with modality-specific attention network. IEEE Trans Med Imaging 2021;40(6). http://dx.doi.org/10.1109/TMI.2021.3059956.

[35] El Habib Daho M, Li Y, Zeghlache R, Boité HL, Deman P, Borderie L, et al. DISCOVER: 2-D multiview summarization of Optical Coherence Tomography Angiography for automatic diabetic retinopathy diagnosis. Artif Intell Med 2024;149:102803. http://dx.doi.org/10.1016/j.artmed.2024.102803.

[36] Garcia-Martin E, Polo V, Larrosa JM, Marques ML, Herrero R, Martin J, et al. Retinal layer segmentation in patients with multiple sclerosis using spectral domain optical coherence tomography. Ophthalmology 2014;121(2):573–9. http://dx.doi.org/10.1016/j.ophtha.2013.09.035.

[37] Oberwahrenbrock T, Traber GL, Lukas S, Gabilondo I, Nolan R, Songster C, et al. Multicenter reliability of semiautomatic retinal layer segmentation using OCT. Neurol Neuroimmunol Neuroinflammation 2018;5(3). http://dx.doi.org/10.1212/nxi.0000000000000449.

[38] Aly L, Strauß E-M, Feucht N, Weiß I, Berthele A, Mitsdoerffer M, et al. Optical coherence tomography angiography indicates subclinical retinal disease in neuromyelitis optica spectrum disorders. Mult Scler J 2021;28(4):522–31. http://dx.doi.org/10.1177/13524585211028831.

[39] Rezende Filho FM, Jurkute N, de Andrade JBC, Marianelli BF, Ferraz Sallum JM, Yu-Wai-Man P, et al. Characterization of retinal architecture in spinocerebellar ataxia type 3 and correlation with disease severity. Mov Disorders 2021;37(4):758–66. http://dx.doi.org/10.1002/mds.28893.

[40] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2020;18(2). http://dx.doi.org/10.1038/s41592-020-01008-z.

[41] Dorent R, Kujawa A, Ivory M, Bakas S, Rieke N, Joutard S, et al. CrossMoDA 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation. Med Image Anal 2023;83. http://dx.doi.org/10.1016/j.media.2022.102628.

[42] Isensee F, Jäger PF, Full PM, Vollmuth P, Maier-Hein KH. nnU-Net for brain tumor segmentation. In: Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. Springer International Publishing; 2021, p. 118–32. http://dx.doi.org/10.1007/978-3-030-72087-2_11.

[43] Peng Y, Xu Y, Wang M, Zhang H, Xie J. The nnU-Net based method for automatic segmenting fetal brain tissues. Health Inf Sci Syst 2023;11(1). http://dx.doi.org/10.1007/s13755-023-00220-3.

[44] Oquab M, Darcet T, Moutakanni T, Vo HV, Szafraniec M, Khalidov V, et al. DINOv2: learning robust visual features without supervision. 2023, http://dx.doi.org/10.48550/arxiv.2304.07193.

[45] Fogarollo S, Bale R, Harders M. Towards liver segmentation in the wild via contrastive distillation. Int J Comput Assist Radiol Surg 2023;18(7). http://dx.doi.org/10.1007/s11548-023-02912-3.

[46] Kiyasseh D, Ma R, Haque TF, Miles BJ, Wagner C, Donoho DA, et al. A vision transformer for decoding surgeon activity from surgical videos. Nat Biomed Eng 2023;7(6). http://dx.doi.org/10.1038/s41551-023-01010-8.

[47] Jin X, Huang T, Wen K, Chi M, An H. HistoSSL: self-supervised representation learning for classifying histopathology images. Mathematics 2022;11(1). http://dx.doi.org/10.1038/s41551-023-01010-8.

[48] Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jégou H. Training data-efficient image transformers and distillation through attention. 2020, http://dx.doi.org/10.48550/arXiv.2012.12877.

[49] Truong T, Mohammadi S, Lenga M. How transferable are self-supervised features in medical image classification tasks? In: Proceedings of machine learning for health. Proceedings of machine learning research, vol. 158, PMLR; 2021, p. 54–74, URL: https://proceedings.mlr.press/v158/truong21a.html.

[50] Shi Y, Ke G, Chen Z, Zheng S, Liu T-Y. Quantized training of gradient boosting decision trees. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. In: Advances in neural information processing systems, vol. 35, Curran Associates, Inc.; 2022, http://dx.doi.org/10.48550/arXiv.2207.09682.

[51] de Melo VV, Ushizima DM, Baracho SF, Coelho RC. Gradient boosting decision trees for echocardiogram images. In: 2018 international joint conference on neural networks. IEEE; 2018, p. 1–8. http://dx.doi.org/10.1109/IJCNN.2018.8489523.

[52] Yang J, Yan J, Pei Z, Hu A, Zhang Y. Prediction model for in-hospital mortality of patients with heart failure based on optuna and light gradient boosting machine. J Mech Med Biol 2022;22(09). http://dx.doi.org/10.1002/jmv.27393.

[53] Gao W, Wang J, Zhou L, Luo Q, Lao Y, Lyu H, et al. Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. Comput Biol Med 2022;140. http://dx.doi.org/10.1016/j.compbiomed.2021.105097.

[54] Zhang X, Xiao Z, Higashita R, Hu Y, Chen W, Yuan J, et al. Adaptive feature squeeze network for nuclear cataract classification in AS-OCT image. J Biomed Inform 2022;128:104037. http://dx.doi.org/10.1016/j.jbi.2022.104037.

[55] Guo Z, Ao S, Ao B. Few-shot learning based oral cancer diagnosis using a dual feature extractor prototypical network. J Biomed Inform 2024;150:104584. http://dx.doi.org/10.1016/j.jbi.2024.104584.
[56] Borja AJ, Hancin EC, Zhang V, Revheim M-E, Alavi A. Potential of PET/CT in assessing dementias with emphasis on cerebrovascular disorders. Eur J Nucl Med Mol Imaging 2020;47(11). http://dx.doi.org/10.1007/s00259-020-04697-y.
[57] Wang B, Wei W, Qiu S, Wang S, Li D, He H. Boundary aware U-Net for retinal layers segmentation in optical coherence tomography images. IEEE J Biomed Health Inf 2021;25(8):3029–40. http://dx.doi.org/10.1109/jbhi.2021.3066208.
[58] Man N, Guo S, Yiu K, Leung C. Multi-layer segmentation of retina OCT images via advanced U-net architecture. Neurocomputing 2023;515:185–200. http://dx.doi.org/10.1016/j.neucom.2022.10.001.
[59] Tulsani A, Patel J, Kumar P, Mayya V, Pavithra K, Geetha M, et al. A novel convolutional neural network for identification of retinal layers using sliced optical coherence tomography images. Healthc Anal 2024;5:100289. http://dx.doi.org/10.1016/j.health.2023.100289.
[60] den Haan J, Verbraak FD, Visser PJ, Bouwman FH. Retinal thickness in Alzheimer's disease: A systematic review and meta-analysis. Alzheimer's Dement Diagn Assess Dis Monit 2017;6(1):162–70. http://dx.doi.org/10.1016/j.dadm.2016.12.014.
[61] Lu Y, Li Z, Zhang X, Ming B, Jia J, Wang R, et al. Retinal nerve fiber layer structure abnormalities in early Alzheimer's disease: Evidence in optical coherence tomography. Neurosci Lett 2010;480(1):69–72. http://dx.doi.org/10.1016/j.neulet.2010.06.006.
[62] López-Cuenca I, Marcos-Dolado A, Yus-Fuertes M, Salobrar-García E, Elvira-Hurtado L, Fernández-Albarral JA, et al. The relationship between retinal layers and brain areas in asymptomatic first-degree relatives of sporadic forms of Alzheimer's disease: an exploratory analysis. Alzheimer's Res Ther 2022;14(1). http://dx.doi.org/10.1186/s13195-022-01008-5.
[63] Gaire BP, Koronyo Y, Fuchs D-T, Shi H, Rentsendorj A, Danziger R, et al. Alzheimer's disease pathophysiology in the Retina. Prog Retin Eye Res 2024;101:101273. http://dx.doi.org/10.1016/j.preteyeres.2024.101273.
[64] Satue M, Castro L, Vilades E, Cordon B, Errea JM, Pueyo A, et al. Ability of Swept-source OCT and OCT-angiography to detect neuroretinal and vasculature changes in patients with Parkinson disease and essential tremor. Eye 2022.
[65] Tak AZA, Yıldızhan Ş, Karadağ AS. Evaluation of thickness of retinal nerve fiber layer, ganglion cell layer, and choroidal thickness in essential tremor: can eyes be a clue for neurodegeneration? Acta Neurol Belg 2017;118(2):235–41. http://dx.doi.org/10.1007/s13760-017-0852-1.
[66] Terravecchia C, Mostile G, Chisari CG, Rascunà C, Terranova R, Cicero CE, et al. Retinal thickness in essential tremor and early parkinson disease: exploring diagnostic insights. J Neuro-Ophthalmol 2023;44(1):35–40. http://dx.doi.org/10.1097/wno.0000000000001959.
[67] Fidancı H, Öksüz N, Özal Ş, Adı güzel U, Kaleağası Ş, Doğu O. Retinal nerve fiber layer thickness in patients with essential tremor and Parkinson's disease. J Surg Med 2019. http://dx.doi.org/10.28982/josam.661757.
[68] Albrecht P, Ringelstein M, Müller A, Keser N, Dietlein T, Lappas A, et al. Degeneration of retinal layers in multiple sclerosis subtypes quantified by optical coherence tomography. Mult Scler J 2012;18(10):1422–9. http://dx.doi.org/10.1177/1352458512439237.
[69] Sotirchos ES, Gonzalez Caldito N, Filippatou A, Fitzgerald KC, Murphy OC, Lambe J, et al. Progressive multiple sclerosis is associated with faster and specific retinal layer atrophy. Ann Neurol 2020;87(6):885–96. http://dx.doi.org/10.1002/ana.25738.
[70] Glasner P, Sabisz A, Chylińska M, Komendziński J, Wyszomirski A, Karaszewski B. Retinal nerve fiber and ganglion cell complex layer thicknesses mirror brain atrophy in patients with relapsing-remitting multiple sclerosis. Restor Neurol Neurosci 2022;40(1):35–42. http://dx.doi.org/10.3233/rnn-211176.
[71] Garcia-Martin E, Larrosa JM, Polo V, Satue M, Marques ML, Alarcia R, et al. Distribution of retinal layer atrophy in patients with parkinson disease and association with disease severity and duration. Am J Ophthalmol 2014;157(2). http://dx.doi.org/10.1016/j.ajo.2013.09.028.
[72] Huang L, Zhang D, Ji J, Wang Y, Zhang R. Central retina changes in Parkinson's disease: a systematic review and meta-analysis. J Neurol 2020;268(12):4646–54. http://dx.doi.org/10.1007/s00415-020-10304-9.
[73] Rascunà C, Russo A, Terravecchia C, Castellino N, Avitabile T, Bonfiglio V, et al. Retinal thickness and microvascular pattern in early parkinson's disease. Front Neurol 2020;11. http://dx.doi.org/10.3389/fneur.2020.533375.
[74] Murueta-Goyena A, Del Pino R, Galdós M, Arana B, Acera M, Carmona-Abellán M, et al. Retinal thickness predicts the risk of cognitive decline in parkinson disease. Ann Neurol 2020;89(1):165–76. http://dx.doi.org/10.1002/ana.25944.
[75] Wang X, Jiao B, Jia X, Wang Y, Liu H, Zhu X, et al. The macular inner plexiform layer thickness as an early diagnostic indicator for Parkinson's disease. npj Parkinson's Dis 2022;8(1). http://dx.doi.org/10.1038/s41531-022-00325-8.
[76] Turkan Y, Tek FB. Automated diagnosis of AD using OCT and OCTA: A systematic review. Authorea, Inc.; 2023, http://dx.doi.org/10.22541/au.168412879.91914539/v1.
[77] Khodabandeh Z, Rabbani H, Ashtari F, Zimmermann HG, Motamedi S, Brandt AU, et al. Discrimination of multiple sclerosis using OCT images from two different centers. Mult Scler Relat Disorders 2023;77:104846.
[78] Shahriari MH, Sabbaghi H, Asadi F, Hosseini A, Khorrami Z. Artificial intelligence in screening, diagnosis, and classification of diabetic macular edema: A systematic review. Surv Ophthalmol 2023;68(1):42–53. http://dx.doi.org/10.1016/j.survophthal.2022.08.004.
[79] Huang X, Ai Z, Wang H, She C, Feng J, Wei Q, et al. GABNet: global attention block for retinal OCT disease classification. Front Neurosci 2023;17. http://dx.doi.org/10.3389/fnins.2023.1143422.
[80] He J, Wang J, Han Z, Ma J, Wang C, Qi M. An interpretable transformer network for the retinal disease classification using optical coherence tomography. Sci Rep 2023;13(1). http://dx.doi.org/10.1038/s41598-023-30853-z.