

Automatic Segmentation of Retinal Layers in Multiple Neurodegenerative Disorder Scenarios

Mateo Gende , Víctor Mallen , Joaquim de Moura , Beatriz Cordón , Elena García-Martin , Clara I. Sánchez , Jorge Novo , and Marcos Ortega

Abstract—Retinal Optical Coherence Tomography (OCT) allows the non-invasive direct observation of the central nervous system, enabling the measurement and extraction of biomarkers from neural tissue that can be helpful in the assessment of ocular, systemic and Neurological Disorders (ND). Deep learning models can be trained to segment the retinal layers for biomarker extraction. However, the onset of ND can have an impact on the neural tissue, which can lead to the degraded performance of models not exposed to images displaying signs of disease during training. We present a fully automatic approach for the retinal layer segmentation in multiple neurodegenerative disorder scenarios, using an annotated dataset of patients of the most prevalent NDs: Alzheimer’s disease, Parkinson’s disease,

multiple sclerosis and essential tremor, along with healthy control patients. Furthermore, we present a two-part, comprehensive study on the effects of ND on the performance of these models. The results show that images of healthy patients may not be sufficient for the robust training of automated segmentation models intended for the analysis of ND patients, and that using images representative of different NDs can increase the model performance. These results indicate that the presence or absence of patients of ND in datasets should be taken into account when training deep learning models for retinal layer segmentation, and that the proposed approach can provide a valuable tool for the robust and reliable diagnosis in multiple scenarios of ND.

Index Terms—Deep learning, medical image segmentation, neurological disease, optical coherence tomography, retina.

Manuscript received 13 March 2023; revised 14 June 2023 and 31 July 2023; accepted 31 August 2023. Date of publication 8 September 2023; date of current version 7 November 2023. This work was supported in part by Instituto de Salud Carlos III, Government of Spain, under Grants PI17/01726 and PI20/00437, in part by Inflammatory Disease Network (RICORS) under Grant RD21/0002/0050, in part by the Ministerio de Ciencia e Innovación, Government of Spain under Grants PDC2022-133132-I00, RTI2018-095894-B-I00, PID2019-108435RB-I00, and TED2021-131201B-I00, in part by the Consellería de Cultura, Educación e Universidade, Xunta de Galicia through Grupos de Referencia Competitiva under Grant ED431C 2020/24 and predoctoral under Grant ED481A 2021/161, in part by CITIC, Centro de Investigación de Galicia under Grant ED431G 2019/01, and in part by Consellería de Educación, Universidade e Formación Profesional, Xunta de Galicia through the ERDF (80%) and Secretaría Xeral de Universidades (20%). (*Corresponding author: Joaquim de Moura.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Ethics Committee of the Miguel Servet Hospital (CEICA: Comité Ético de Investigaciones Científicas de Aragón) under Application No. C.I. PI21/113, and performed in line with the Declaration of Helsinki.

Mateo Gende, Joaquim de Moura, Jorge Novo, and Marcos Ortega are with the Grupo VARPA, Instituto de Investigación Biomédica de A Coruña, Universidad da Coruña, 15001 A Coruña, Spain, and also with the Centro de Investigación CITIC, Universidad da Coruña, 15001 A Coruña, Spain (e-mail: m.gende@udc.es; joaquim.demoura@udc.es; jnovo@udc.es; mortega@udc.es).

Víctor Mallen, Beatriz Cordón, and Elena García-Martin are with the Department of Ophthalmology, Miguel Servet University Hospital, 50009 Zaragoza, Spain, and also with the Aragon Institute for Health Research (IIS Aragón), Miguel Servet Ophthalmology Innovation and Research Group (GIMSO), University of Zaragoza, 50009 Zaragoza, Spain (e-mail: victormallen6@hotmail.com; beatrizcordonc@gmail.com; egmviav@yahoo.com).

Clara I. Sánchez is with the Quantitative Healthcare Analysis (qurAI) Group, Informatics Institute, Universiteit van Amsterdam, 1000 GG Amsterdam, The Netherlands, and also with the Biomedical Engineering and Physics, Amsterdam UMC Locatie AMC Department of Biomedical Engineering and Physics, 1000 GG Amsterdam, The Netherlands (e-mail: c.i.sanchezgutierrez@uva.nl).

Digital Object Identifier 10.1109/JBHI.2023.3313392

I. INTRODUCTION

A. Context

PATIENTS of neurological disorders (ND) recurrently suffer from a late diagnosis, often when the disease has already progressed beyond a stage where there is a noticeable effect in patient quality of life. Furthermore, cognitive function testing is affected by a lack of repeatability, vague symptomatology and day-to-day fluctuations in patient conduct, which can further delay diagnosis [1], [2]. Aside from externally perceptible symptoms, these disorders can have a measurable impact on the tissue of the central nervous system. In the domain of neuro-ophthalmology, the measurement of changes to this tissue can provide a unique opportunity for an early diagnosis and treatment of these diseases, helping to mitigate the progressive degeneration and improving patient care and quality of life [3]. Changes to the human retina have been shown to correlate with the effects that these diseases present on the neurological tissue [4], [5]. Thanks to the transparency of the optic medium, this makes it possible to study the progression of ND by directly observing changes to the ocular neurological tissue.

Optical Coherence Tomography (OCT) is a medical imaging technique that uses light interferometry in order to obtain depth-wise measurements of biological tissue [6], [7]. In ophthalmology, OCT is used to obtain in-vivo cross-sectional visualizations of the ocular tissue in a non-invasive way. Its micrometre-resolution, and the fact that it can generate three-dimensional volumes of information that represent the inner ocular tissue, make OCT one of the most prevalent ophthalmic

imaging modalities [8]. OCT has been shown to be of great use for the diagnosis of several ocular diseases such as glaucoma [9], [10], uveitis [11], diabetic macular oedema [12], and age-related macular degeneration [13], among others. Furthermore, its non-invasiveness, along with the high resolution volumetric visualisation that OCT enables without the need of contrast agents, makes it an attractive alternative to brain imaging for the diagnosis and assessment of ND. Because of this, several studies have focused on finding biomarkers that allow the assessment of these diseases from OCT images [4], [5], [14], [15]. Specifically, clinical studies have found associations between changes in the retinal layers and the progression of patients of Alzheimer's Disease (AD) [16], [17], [18], [19], [20], in patients of Parkinson's Disease (PD) [18], [21], [22], and in patients of Multiple Sclerosis (MS) [23], [24], [25], among others.

In order to measure the disease-related changes in the neurological tissue of the eye, it is often necessary to identify and quantify the thickness of the retinal layers. This can be used to identify signs of disease, as well as to track patient progression and characterise the progressive thinning of the layers as a sign of degeneration. While most manufacturers provide retinal layer segmentation software suites, these may require manual correction, especially when analysing images that display signs of pathology which deviate from the norm [26], [27], as is the case of patients affected by ND. However, the manual annotation of these layers is a subjective and tiresome process, which motivates the development of fully automated retinal layer segmentation methods [28]. There are several approaches to the problem of determining the boundary and content of the retinal layers, from kernel regression [29] and graph-theory [30] based approaches to approaches that combine these methods with machine learning [31], [32], [33], [34]. Deep learning models, and more specifically, Convolutional Neural Networks (CNNs), have proven to be especially useful for retinal layer segmentation due to their ability to be trained directly from annotated data without the need manual process of feature extraction, selection or adjustment. As such, many works report success in the use of CNNs for retinal layer segmentation (for reference, [31], [35], [36], [37], [38], [39]).

B. Related Works

Recent advances in neuro-ophthalmology, as well as the current relevance of the analysis of OCT images of ND patients has lead to the development of recent works focused on the automatic segmentation of retinal layers in patients of these disorders. Among the different disorders, most works focus on the retinal segmentation in patients of MS: in [40], a set of features was extracted around each pixel of the OCT image, and a random forest classifier was used to classify them as boundary regions. A graph search algorithm was then used to further refine the boundaries in order to produce the segmentation. The method was trained and validated on control and MS patients, reporting a higher error on images corresponding to MS patients. In [41], the authors proposed a layer evolution algorithm for retinal boundary segmentation in control and MS patient OCT images. This method used a series of visual features to train a

random forest classifier, which can return a boundary probability map for the image. This probability map was then progressively refined in the layer boundary evolution process. He et al. [42] proposed the use of a segmentation network followed by a regression network to correct boundary defects. These networks could segment both the retinal layers and microcystic macular oedema in images of MS patients. Wang et al. [43] proposed a modified U-Net [44] structure with edge and boundary aware modules incorporated within the skip connections. The same authors later proposed the use of fully convolutional regression networks for layer segmentation in patients of MS and diabetic macular oedema [45]. This method was able to achieve sub-pixel segmentation accuracy in an end to end manner without the need of post-processing stages. By modelling layer boundaries as height values in the images, a boundary guarantee loss can be incorporated to ensure topological consistency. The method was evaluated on a dataset containing MS patients and controls, reporting a drop in performance on the MS patients when compared with controls. In [46], the authors proposed the use of a cascaded two-stage network, with each stage being composed of a compressed version of U-Net. A post-processing stage was incorporated in which a Laplacian-based outlier detection is applied, followed by an adaptive non-linear interpolation with the intention of filling any inconsistent holes in the segmentation. The method was trained and validated in a set of MS and control patients.

Thus, the automatic segmentation of retinal layers in ND patients has garnered recent interest, due to its relevance for neuro-ophthalmic diagnosis [14], [15]. Nevertheless, the existing body of work is currently limited in its reach and representation of ND. Most of the existing works are focused on the automatic segmentation of MS patients, while several other NDs remain unaddressed. Furthermore, the majority of these works are based on deep learning-based models. These models rely on the visual patterns that are apparent in their training data in order to segment the images during inference. This means that the trained models can present problems in generalising to images that display visual patterns that differ from those the model was trained in [47]. In this sense, models that were trained on images of healthy patients, or of patients of a single ND such as MS, may experience a loss of performance when used on images of patients of other NDs. All the previous works that break down the results between ND patients and controls report comparatively worse results for images of the former class [40], [43]. Currently, to the extent of our knowledge, no study has addressed or compared the performance of automatic OCT layer segmentation methods for different NDs so far. From the examples in the literature it can be inferred that these diseases affect the tissue in a way that can influence the retinal layer segmentation models. Furthermore, these diseases are all caused by different physiological mechanisms and can have a different effect on neural tissue [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [48], [49]. In this regard, the impact that the lack of training set representation of the most prevalent NDs can have in the automatic retinal layer segmentation models, as well as an in-depth study on the behaviour of these models in different ND scenarios, remain to be addressed.

C. Contributions

Given this important gap in the literature, in this work, we present an approach for the fully automatic segmentation of the retinal layers in multiple ND-specific scenarios representative of the most common disorders. Furthermore, we conducted an exhaustive study of the performance and robustness of this deep learning-based approach in different scenarios of image availability. A dataset containing images of the four most prevalent NDs, namely AD, PD, MS and Essential Tremor (ET), along with Healthy Control (HC) patients was annotated with the location of the retinal layers of interest and used to train and validate state-of-the-art deep learning models. This dataset constitutes the most complete sample of retinal OCT imaging of ND patients extracted under identical circumstances. Moreover, two complementary analyses were carried out in order to conduct a comprehensive study on the performance of these automatic retinal layer segmentation models trained with the images of these representative NDs. In a first experiment, a series of models were trained on images displaying symptoms of every ND but one. Then, these models were evaluated in their ability to segment the retinal layers of patients of the ND that was excluded from the training. This allows us to compare the ability of the models to generalise to each specific ND by emulating scenarios in which these have never been exposed to the characteristic visual features that may be apparent in images of that disease. In the second experiment, the models were trained on images belonging to patients of a specific NDs (or control patients), and then evaluated on their ability to generalise to each of the other specific diseases. This allowed us to analyse how the visual patterns apparent in each specific ND can contribute to the segmentation, and which image sets may be lacking in order to address the problem of inter-disease variability when training retinal layer segmentation models. To the extent of our knowledge, this work is the first to address the automatic segmentation of retinal layers in OCT images of patients of these representative NDs, also providing an exhaustive analysis of the effects of training deep learning models for this task in the presence or absence of images that are characteristic of these disorders. The primary contributions and findings of this work can be summarised as follows:

- This work presents the first study into the impact of NDs on the performance of retinal layer segmentation models, using an annotated OCT imaging dataset comprising the four most prevalent NDs acquired under uniform conditions, and pioneering a comprehensive exploration in this domain.
- The study reveals that models trained using HC images demonstrate sub-optimal performance when confronted with images from patients suffering from NDs, leading models to over-segment the RNFL. This can mask the degenerative effects of these diseases and hinder accurate patient assessment.
- In order to address the representation challenge posed by individual NDs, the proposed approach uses images of other NDs during training. Remarkably, the models trained using this approach effectively compensate for the lack

of representation of any single ND of those considered. Furthermore, models trained only on images of AD or ET patients exhibit, on average, superior performance when segmenting the layers of patients of the other NDs.

- This study sets the stage for future research endeavours focused on identifying distinctive visual features that can characterise images representative of different NDs. These features hold the potential to significantly improve early differential diagnosis, leading to enhanced patient care and overall quality of life for individuals affected by these diseases.

D. Outline

This manuscript is structured as follows: Section II explains the data and the deep learning models that were employed in this study, along with an explanation of the experimental setup; Section III presents the results that were obtained for each of the experiments, as well as a discussion of the obtained results. Section IV outlines the main conclusions.

II. MATERIALS AND METHODS

In this section, we introduce the materials and resources that were employed for the training and validation of the models, along with an explanation of the experimental setup that was followed.

A. Dataset

In order to train and evaluate the retinal layer segmentation models, a set of macular OCT images was specifically collected and annotated for this work. This set was formed by five independent samples of the four most prevalent NDs: AD, PD, MS and ET, along with a sample of HC patients. These were prospectively recruited by three specialist clinicians (one ophthalmologist specialised in neuro-ophthalmology, one neurologist specialised in demyelinating diseases and one neurologist specialist in movement disorders and dementia).

The AD group includes patients with an AD diagnosis as indicated by expert neurologists in accordance with the Stroke-Alzheimer's Disease and Related Disorders Association, the National institute of Neurologic and Communicative Disorders and the Diagnostic and Statistical Manual of mental Disorders (DSM IV) [50] criteria. For inclusion, only patients with less than 5 years since diagnosis and low to moderate changes (Mini Mental State Examination ≥ 20) were considered, aiming to focus in patients representative of an early diagnosis. Standard clinical and neuroimaging criteria were used for the diagnosis of MS [51]. Only relapsing-remitting MS phenotype patients were included in order to obtain a homogeneous population. A trained neurologist diagnosed PD based on standard clinical and neuroimaging criteria [52], as well as information about disease severity using the Hoehn Yahr scale [53], according to the United Kingdom Brain Bank Criteria [54]. The diagnosis of ET was based on the Consensus statement of the Movement Disorder Society on Tremor [55], excluding cases of “probable ET” and

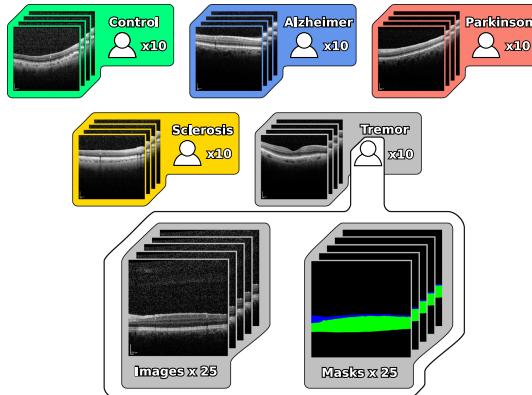


Fig. 1. Summary of the dataset distribution. 1250 images were extracted from 50 patients, grouped in 5 different ND sets.

“possible ET”. HC subjects include patients referred for refraction that underwent routine examination without abnormal ocular findings. The exclusion criteria employed were best corrected visual acuity lower than 0.5 using Snellen charts, refractive errors higher than 5 dioptres of spherical equivalent refraction or 3 dioptres of astigmatism, intraocular pressure higher than 20 mmHg, media opacifications (nuclear colour/opalescence, cortical or posterior subcapsular lens opacity > 2, according to the Lens Opacities Classification System III [56]), concomitant ocular diseases such as glaucoma or retinal pathology, and other systemic conditions that could affect the visual system. Furthermore, related medical records were carefully reviewed, including disease duration, the Expanded Disability Status Scale (EDSS) score in MS patients, the treatments, and the presence of prior episodes of optic neuritis. All procedures performed in this work are in accordance with the Declaration of Helsinki. Written informed consent was collected from all participants in the study, and the experimental protocol was approved by the Ethics Committee of the Miguel Servet Hospital (CEICA: Comité Ético de Investigaciones Científicas de Aragón) under registration number C.I. PI21/113. For each of the 50 patients, 25 equally spaced macular images were extracted and annotated, for a total of 1250 annotated images (Fig. 1). All patients contained in this study are within a similar distribution of age and sex in order to prevent any relevant differences that may cause a significant distortion of the results. These images were acquired using a Heidelberg SPECTRALIS imaging platform. OCT output was included without manual correction. An assessment of the quality of the scans was performed prior to the analysis, rejecting any poor-quality scans or images with less than 25/40 points of quality [57]. The original resolution of these images was 512×496 . These images were annotated with two semantic classes: the Retinal Nerve Fibre Layer (RNFL), the layer most commonly reported to be affected by ND, and a class grouping the rest of the retinal layers between the Ganglion Cell Layer and Bruch’s Membrane (GCL-BM), as a way to quantify the remaining retinal thickness.

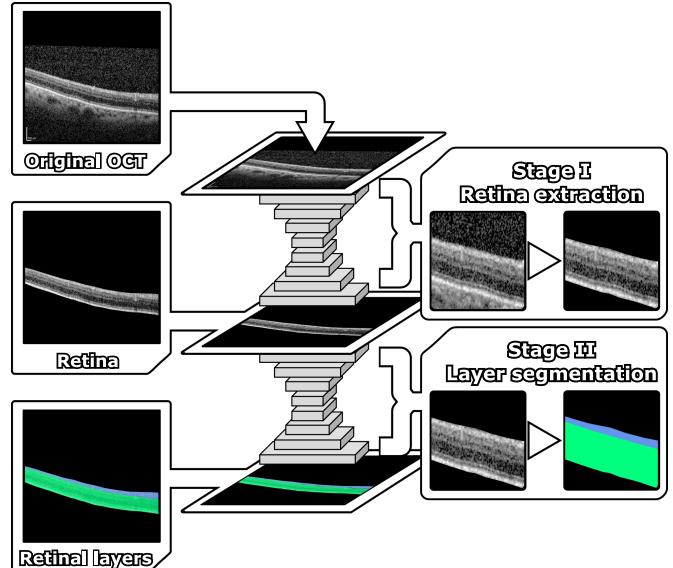


Fig. 2. Summary of the segmentation process in the MGU-Net architecture. The first stage generates a segmentation mask separating the retinal tissue from the background. This mask is applied to the image before presenting it to the second stage. This second stage separately segments each retinal layer producing the final output. The right side displays detailed views of the retina and a visualisation of the effects of each stage.

B. Software and Hardware Resources

The models were developed, trained and validated using the PyTorch [58] machine learning library (version 1.8.2) under Python (version 3.8.10). For all the image processing requirements, OpenCV [59] (version 4.5.4-dev) and NumPy (version 1.21.4) were employed. Statistical tests were performed using the SciPy library (version 1.10.1). The training and validation process of the model was performed on two NVIDIA RTX A6000 GPUs, an AMD Ryzen Threadripper 3960X CPU, and 256 GB RAM.

C. Retinal Layer Segmentation

MGU-Net [60] was used as a base network architecture for retinal layer segmentation, due to its favourable results when compared with other segmentation architectures such as U-Net [44], ReLayNet [61] and DRUNET [62]. The MGU architecture uses a two-stage process for retinal layer segmentation (Fig. 2). The first stage separates the retinal tissue from the background, separating the region of interest for further refinement. The second stage refines the area segmented by the first one into each of the individual retinal layers. This architecture was adapted for the segmentation of the RNFL and the GCL-BM by adapting its final layer.

In order to best capitalise on the amount of available data, the models were first initialised to a pre-training on a publicly available retinal layer segmentation dataset [60]. This dataset contains a total of 732 OCT scans. The pre-training process was performed for a maximum of 50 epochs under the same

conditions employed in the original publication. Afterwards, all of the layers in the models were allowed to train on the data partitions described for each experiment. For compatibility with the segmentation model, all the images were resized to 1024×992 pixels, the intensity values converted to a float value within $[0, 1]$ and normalised using the mean and standard deviation of the whole dataset. The images used for training were further subdivided at the patient level, using 90% of the images for training, and the remaining 10% for validation, ensuring that images from the same patient are not shared between the training, validation or test sets.

The models underwent training for a maximum of 50 epochs in the baseline and experiment 1, while experiment 2 extended the training to 100 epochs, which was empirically determined to ensure model convergence. Throughout training, a checkpoint of the trainable weights was saved at each epoch where the validation loss reduced, selecting the checkpoint with the lowest validation loss for model testing. For model optimisation, we employed the Adam optimiser [63]. The hyperparameters employed were a learning rate of 0.001, a momentum of 0.9, and a weight decay of 1×10^{-4} . Additionally, a batch size of 16 was used during training. To calculate the loss, each stage of the model used its separate calculation, which was then weighted (1). The segmentation loss for each stage (2) comprised the summation of Cross-Entropy Loss (3) and Dice Loss (4), defined as follows:

$$\mathcal{L} = \mathcal{L}_{stage_1} + 2 \times \mathcal{L}_{stage_2} \quad (1)$$

$$\mathcal{L}_{stage} = \mathcal{L}_{CE} + \mathcal{L}_{Dice} \quad (2)$$

$$\mathcal{L}_{CE} = -\frac{1}{C} \sum_{c=1}^C g_c \log p_i \quad (3)$$

$$\mathcal{L}_{Dice} = 1 - \frac{1}{C} \sum_{c=1}^C \frac{2 \sum_{i \in \Omega} p_c(i) \times g_c(i)}{\sum_{i \in \Omega} p_c(i) + \sum_{i \in \Omega} g_c(i)} \quad (4)$$

where g_c indicates the ground truth and p_c indicates the predicted probability for pixel i belonging to class c . In order to make the most of the amount of available data, image augmentation was applied to the training samples in the form of random rotations in the range $[-7^\circ, 7^\circ]$, random horizontal and vertical shear transformations in the range $[-7^\circ, 7^\circ]$, Gaussian noise and intensity variations in the range $[0.9, 1.1]$, as well as random horizontal flipping.

D. Impact of ND

Two complementary analyses were conducted in order to provide an exhaustive study of the performance of the retinal layer segmentation models when training in multiple scenarios of presence or absence of the different NDs. These analyses are implemented in the form of two experiments that can provide insight into how the models perform on images of different NDs when trained in the absence of images of each specific ND. To do this, the images were first subdivided in 5 sets, one for each ND, and one for the images from HC. In the first experiment, the models were trained using images from all the sets but one.

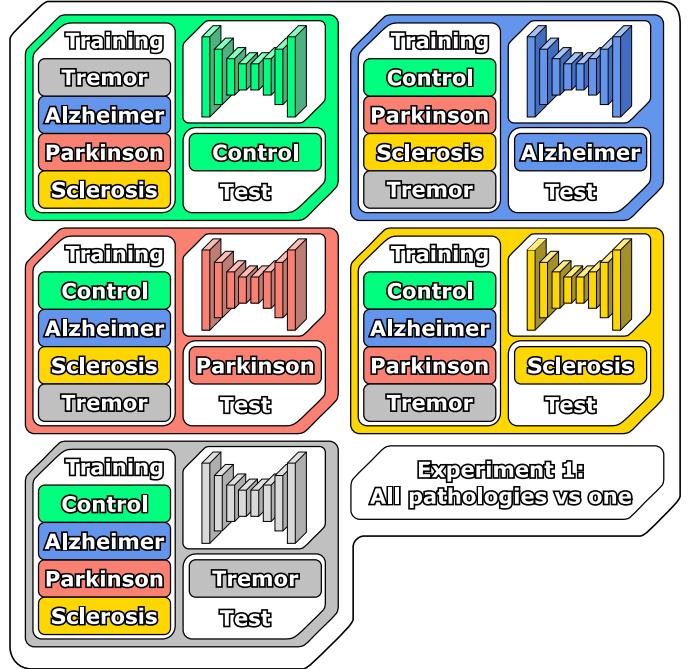


Fig. 3. First experiment: each model is trained on the images from all sets except one. The remaining set is used to evaluate each model.

Then, the remaining images were used as a test set to evaluate the model. This provides valuable insight into the particular differences in terms of visual features that may set each ND apart from the others. In the second experiment, each model was trained using images from a single set, and then these were evaluated on each of the other sets. This experiment can enable the comparison between all the available NDs, highlighting NDs that are similar in terms of visual features, and which cases may need to be supplied with additional samples in order to make up for lack of representation.

1) Analysis 1 - Multi ND Training: The main purpose of this analysis is twofold: To assess how well the retinal layer segmentation models perform on images that show visual signs of diseases that were not part of their training data, and to explore whether certain neurological disorders require the use of images that represent they characterising structural changes in the retina. To this end, the models were initially trained on all available sets of images, excluding one set. Then, the trained models were tested on the images from the excluded set. A diagram displaying this procedure and the partitioning that was employed can be found in Fig. 3.

Training the models on all available sets except one ensures that the model does not see any visual features exclusive to that particular set during training. Doing this allows the model to be exposed to shared features between multiple NDs and controls. By evaluating the performance on the excluded set, it is possible to determine which NDs can benefit from training with images from other sets (i.e., sharing features), and which NDs require representative samples from their own set to be used during training. This approach can help understand the transferability of features between different NDs and controls.

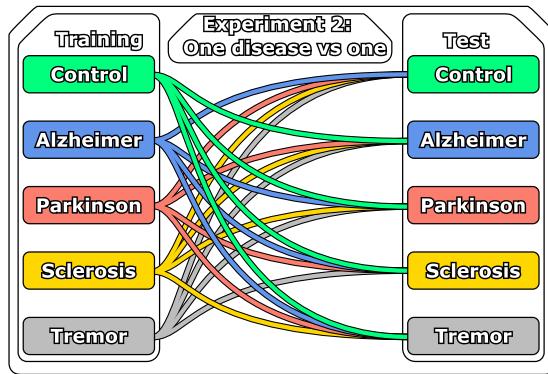


Fig. 4. Second experiment: each model is trained in a single set and then evaluated in all of the unseen ones separately.

2) Analysis 2 - Single ND Training: The main goal of this analysis is to assess the similarities between the visual features presented by each individual ND. To achieve this, an experiment was conducted where the models were trained on one of the sets and subsequently evaluated on each of the other sets. This allows the measurement of the performance of the model across various combinations of training and test sets. A summary of all of the combinations of training and test sets employed for this experiment is represented in Fig. 4.

By comparing the performance of models trained on one dataset when applied to images from another, the one-directional compatibility of NDs can be studied, providing insight into how well models trained on images of one ND perform when presented with images from another ND. A model trained on images from a disease that exhibits visual patterns similar to another ND would be expected to perform better on the latter than a model trained on images that do not share these visual patterns. By studying these mutual performance variations, it is possible to gain a deeper understanding of the shared characteristics and distinctive features of various NDs, ultimately contributing to a more comprehensive and nuanced understanding of these diseases.

3) Baseline: A baseline experiment was also performed in order to enable a comparison with an ideal case in which images from all NDs are simultaneously available. In it, the models are trained with images from all the sets, setting a patient from each set apart for testing. This process is repeated until all patients have been used for testing, enabling a fair comparison with the other experiments.

E. Evaluation

A comprehensive set of metrics was used in order to achieve a fair and complete evaluation of the experiments performed in this work. These metrics are, namely, Precision (5), Sensitivity (6), and Dice Score (7), as defined by:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (7)$$

Where TP, FP, TN and FN denote True Positives, False Positives, True Negatives and False Negatives, respectively. Each pixel was considered to belong to the layer with the highest score returned by the segmentation mask produced by the model. Along with these pixel-wise metrics, the segmentation error for the models was also computed. The contour surrounding every layer was extracted both from each ground truth and automatically segmented output image. Following other related retinal layer segmentation works (for reference, [40], [41], [46]), the Mean Absolute Error (MAE) was computed between the pixel heights for every image column as a measurement of contour error (MAE_C), along with the MAE for the thickness of each layer (MAE_T). In order to provide a robust measurement of these metrics, and following other similar approaches in the literature [64], [65], bootstrapping with 1000 repetitions was applied at the image level to all tests.

III. RESULTS AND DISCUSSION

This section presents the results that were obtained for each of the experiments previously described, along with a detailed discussion highlighting any conclusions that can be extracted, in order to provide an exhaustive and comprehensive analysis of the performance of the models and the effect of the different scenarios of ND presence.

A. Experiment 1

In the first experiment, the models were trained on images that displayed signs of every ND except the one used for testing. Table I displays the results of this first experiment, while Table II displays the test results achieved by the baseline models, which were allowed to train with images from all of the sets.

The results that were obtained show that the models are able to achieve a performance that is on par with the trained baseline, in which they were exposed to images of every ND, including the one used for testing. There are no remarkable differences between the results from the first experiment and the baseline. The biggest observable differences are those in the PD image set, but these may be explained by the training variability. This indicates that other NDs may be used to supply the missing visual features that characterise each specific ND.

The average layer thickness at the eye level was computed for the ground truth and the automatically measured values from this experiment (Fig. 5). This shows that the automatically measured values closely align with the annotated thicknesses. It also allows a comparison between the layer thicknesses of all NDs. Patients of MS show the highest variability in RNFL thickness, while ET patients display the lowest. For the GCL-BM layers, the healthy patients generally show a higher thickness than the ND patients, which is in line with previous reports of thinner ganglion cell layer in patients of NDs [66], [67], [68], [69]. Nevertheless, a one-way analysis of variance test found no significant differences between the thickness of the retinal layers for the eyes of patients of these different NDs ($F = 1.811$, $p = 0.143$ for the RNFL; $F = 1.247$, $p = 0.305$ for the GCL-BM layers).

TABLE I

TEST RESULTS ACHIEVED BY THE EXPERIMENT 1 MODELS FOR EACH LAYER. PRECISION, SENSITIVITY AND DICE COEFFICIENT VALUES REPRESENT A RATIO BETWEEN 0 AND 1. MAE_C AND MAE_T VALUES REPRESENT NUMBER OF PIXELS. ALL VALUES REPRESENTED AS MEAN \pm STANDARD DEVIATION

Retinal Nerve Fibre Layer					
Set	Precision	Sensitivity	Dice	MAE_C	MAE_T
HC	0.936 \pm 0.003	0.950 \pm 0.002	0.943 \pm 0.001	1.33 \pm 0.03	2.10 \pm 0.06
AD	0.926 \pm 0.003	0.931 \pm 0.006	0.928 \pm 0.004	1.90 \pm 0.17	2.69 \pm 0.14
PD	0.938 \pm 0.002	0.958 \pm 0.002	0.948 \pm 0.002	2.88 \pm 0.63	2.08 \pm 0.07
MS	0.949 \pm 0.002	0.946 \pm 0.002	0.947 \pm 0.001	1.20 \pm 0.03	1.87 \pm 0.05
ET	0.945 \pm 0.002	0.936 \pm 0.004	0.940 \pm 0.002	1.37 \pm 0.06	2.07 \pm 0.08

Ganglion Cell Layer - Bruch's Membrane					
Set	Precision	Sensitivity	Dice	MAE_C	MAE_T
HC	0.986 \pm 0.001	0.990 \pm 0.000	0.988 \pm 0.000	1.64 \pm 0.04	2.45 \pm 0.06
AD	0.985 \pm 0.001	0.973 \pm 0.002	0.979 \pm 0.001	2.70 \pm 0.15	4.14 \pm 0.23
PD	0.984 \pm 0.002	0.987 \pm 0.000	0.985 \pm 0.001	2.95 \pm 0.55	2.87 \pm 0.21
MS	0.989 \pm 0.000	0.988 \pm 0.000	0.988 \pm 0.000	1.54 \pm 0.04	2.35 \pm 0.06
ET	0.985 \pm 0.001	0.986 \pm 0.001	0.986 \pm 0.000	1.88 \pm 0.06	2.89 \pm 0.11

TABLE II

TEST RESULTS ACHIEVED BY THE BASELINE MODELS FOR EACH LAYER. PRECISION, SENSITIVITY AND DICE COEFFICIENT VALUES REPRESENT A RATIO BETWEEN 0 AND 1. MAE_C AND MAE_T VALUES REPRESENT NUMBER OF PIXELS. ALL VALUES REPRESENTED AS MEAN \pm STANDARD DEVIATION

Retinal Nerve Fibre Layer					
Set	Precision	Sensitivity	Dice	MAE_C	MAE_T
HC	0.941 \pm 0.002	0.949 \pm 0.002	0.945 \pm 0.001	1.29 \pm 0.03	2.07 \pm 0.06
AD	0.928 \pm 0.004	0.926 \pm 0.007	0.926 \pm 0.005	1.94 \pm 0.20	2.01 \pm 0.17
PD	0.942 \pm 0.002	0.957 \pm 0.003	0.949 \pm 0.002	2.73 \pm 0.60	2.04 \pm 0.08
MS	0.946 \pm 0.002	0.949 \pm 0.002	0.947 \pm 0.001	1.18 \pm 0.03	1.81 \pm 0.04
ET	0.944 \pm 0.002	0.941 \pm 0.003	0.943 \pm 0.002	1.31 \pm 0.05	1.98 \pm 0.07

Ganglion Cell Layer - Bruch's Membrane					
Set	Precision	Sensitivity	Dice	MAE_C	MAE_T
HC	0.987 \pm 0.000	0.989 \pm 0.000	0.988 \pm 0.000	1.59 \pm 0.03	2.43 \pm 0.06
AD	0.984 \pm 0.001	0.973 \pm 0.002	0.978 \pm 0.001	2.69 \pm 0.14	4.20 \pm 0.25
PD	0.984 \pm 0.002	0.987 \pm 0.000	0.986 \pm 0.001	3.30 \pm 0.64	2.84 \pm 0.21
MS	0.990 \pm 0.000	0.988 \pm 0.000	0.989 \pm 0.000	1.52 \pm 0.04	2.30 \pm 0.05
ET	0.988 \pm 0.001	0.986 \pm 0.001	0.986 \pm 0.000	1.81 \pm 0.06	2.76 \pm 0.10

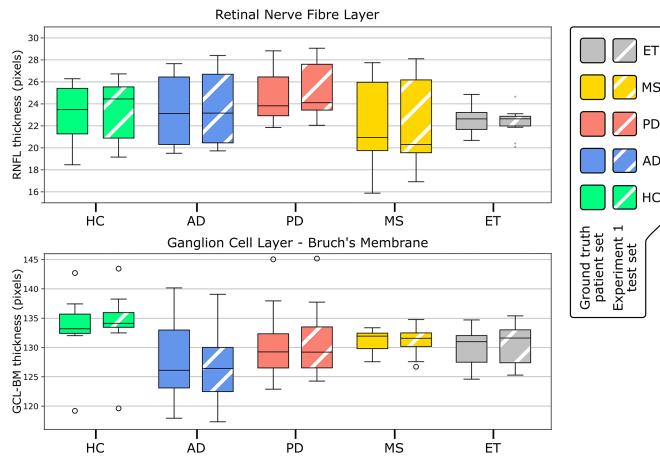


Fig. 5. Retinal thickness averaged per eye for the ground truth and the values measured by the first experiment models, grouped by image set. Solid colours show values extracted from the annotated ground truth images. Striped colours show values automatically measured by the Experiment 1 models.

B. Experiment 2

In the second experiment, the models were trained on every individual image set and then evaluated separately in every other

set. Tables III and IV display the test results for the RNFL and GCL-BM layers, respectively. Fig. 6 summarises the contour and thickness errors grouped by the training and test sets used. Furthermore, a Wilcoxon signed-rank test was performed in order to validate any statistically significant differences between the models trained with every set of images. In this series of tests, the models were grouped according to the images used for training each of them. Then, for every one-on-one comparison, the images shared between the test sets of the models to be compared were grouped together. Thus, for each comparison of the models trained on two different image sets, the corresponding results for the test images of these models applied to the images of the three remaining sets were selected. This paired test was applied to the contour and thickness errors achieved by the models, and a summary of the results can be found in Fig. 7, where an arrow is shown pointing to any training set that produces a significantly lower ($p < 0.05$) error than another, or a cross is shown when no statistically significant differences are found.

Regarding the one-on-one comparative performed in this experiment, more differences can be observed between the different sets. Averaging and grouping the results achieved by the image set used for training (Fig. 6), we can see that the worst results in terms of RNFL MAE_C and MAE_T , as well as

TABLE III

TEST RESULTS ACHIEVED BY THE EXPERIMENT 2 MODELS FOR THE RNFL LAYER. PRECISION, SENSITIVITY AND DICE COEFFICIENT VALUES REPRESENT A RATIO BETWEEN 0 AND 1. MAE_C AND MAE_T VALUES REPRESENT NUMBER OF PIXELS. ALL VALUES REPRESENTED AS MEAN \pm STANDARD DEVIATION

		Retinal Nerve Fibre Layer				
Train	Test	Precision	Sensitivity	Dice	MAE_C	MAE_T
HC	AD	0.850 \pm 0.014	0.929 \pm 0.006	0.882 \pm 0.010	7.17 \pm 0.93	4.52 \pm 0.38
	PD	0.930 \pm 0.003	0.958 \pm 0.002	0.943 \pm 0.002	1.58 \pm 0.14	2.27 \pm 0.08
	MS	0.926 \pm 0.004	0.948 \pm 0.002	0.937 \pm 0.002	1.59 \pm 0.10	2.29 \pm 0.08
	ET	0.896 \pm 0.007	0.928 \pm 0.005	0.912 \pm 0.005	5.04 \pm 0.75	2.84 \pm 0.13
AD	HC	0.923 \pm 0.004	0.926 \pm 0.003	0.924 \pm 0.002	2.32 \pm 0.21	2.72 \pm 0.10
	PD	0.946 \pm 0.002	0.938 \pm 0.003	0.942 \pm 0.001	1.74 \pm 0.20	2.27 \pm 0.07
	MS	0.944 \pm 0.003	0.934 \pm 0.002	0.939 \pm 0.002	1.39 \pm 0.05	2.15 \pm 0.06
	ET	0.938 \pm 0.003	0.925 \pm 0.003	0.932 \pm 0.002	1.51 \pm 0.05	2.41 \pm 0.09
PD	HC	0.867 \pm 0.007	0.927 \pm 0.003	0.896 \pm 0.004	4.67 \pm 0.39	3.69 \pm 0.19
	AD	0.894 \pm 0.007	0.910 \pm 0.008	0.902 \pm 0.007	3.92 \pm 0.51	3.08 \pm 0.17
	MS	0.927 \pm 0.004	0.945 \pm 0.002	0.936 \pm 0.002	1.55 \pm 0.05	2.26 \pm 0.08
	ET	0.914 \pm 0.006	0.913 \pm 0.006	0.914 \pm 0.005	3.53 \pm 0.53	2.70 \pm 0.11
MS	HC	0.925 \pm 0.003	0.942 \pm 0.003	0.934 \pm 0.002	1.60 \pm 0.07	2.45 \pm 0.08
	AD	0.895 \pm 0.005	0.918 \pm 0.007	0.907 \pm 0.006	3.19 \pm 0.33	3.02 \pm 0.12
	PD	0.935 \pm 0.002	0.941 \pm 0.004	0.938 \pm 0.002	1.60 \pm 0.11	2.45 \pm 0.10
	ET	0.882 \pm 0.009	0.923 \pm 0.006	0.902 \pm 0.006	4.65 \pm 0.70	3.22 \pm 0.17
ET	HC	0.911 \pm 0.005	0.936 \pm 0.003	0.923 \pm 0.003	2.88 \pm 0.31	2.78 \pm 0.12
	AD	0.918 \pm 0.004	0.932 \pm 0.006	0.925 \pm 0.004	1.96 \pm 0.15	2.83 \pm 0.14
	PD	0.939 \pm 0.003	0.944 \pm 0.002	0.941 \pm 0.002	1.52 \pm 0.08	2.09 \pm 0.06
	MS	0.937 \pm 0.002	0.950 \pm 0.003	0.943 \pm 0.001	2.10 \pm 0.31	2.30 \pm 0.07

TABLE IV

TEST RESULTS ACHIEVED BY THE EXPERIMENT 2 MODELS FOR THE GCL-BM LAYER. PRECISION, SENSITIVITY AND DICE COEFFICIENT VALUES REPRESENT A RATIO BETWEEN 0 AND 1. MAE_C AND MAE_T VALUES REPRESENT NUMBER OF PIXELS. ALL VALUES REPRESENTED AS MEAN \pm STANDARD DEVIATION

Ganglion Cell Layer - Bruch's Membrane						
Train	Test	Precision	Sensitivity	Dice	MAE_C	MAE_T
HC	AD	0.986 \pm 0.001	0.919 \pm 0.010	0.951 \pm 0.005	4.69 \pm 0.42	9.66 \pm 1.06
	PD	0.987 \pm 0.001	0.983 \pm 0.002	0.985 \pm 0.001	2.15 \pm 0.16	2.90 \pm 0.18
	MS	0.990 \pm 0.000	0.983 \pm 0.001	0.986 \pm 0.000	1.98 \pm 0.14	2.86 \pm 0.11
	ET	0.982 \pm 0.001	0.948 \pm 0.006	0.965 \pm 0.003	3.60 \pm 0.32	7.21 \pm 0.77
AD	HC	0.968 \pm 0.004	0.987 \pm 0.000	0.978 \pm 0.002	4.21 \pm 0.45	5.13 \pm 0.51
	PD	0.982 \pm 0.002	0.986 \pm 0.000	0.984 \pm 0.001	2.58 \pm 0.31	3.20 \pm 0.21
	MS	0.985 \pm 0.001	0.986 \pm 0.001	0.986 \pm 0.000	2.09 \pm 0.11	2.88 \pm 0.09
	ET	0.982 \pm 0.001	0.984 \pm 0.001	0.983 \pm 0.001	2.20 \pm 0.06	3.36 \pm 0.11
PD	HC	0.954 \pm 0.005	0.989 \pm 0.000	0.971 \pm 0.002	7.04 \pm 0.64	6.76 \pm 0.62
	AD	0.986 \pm 0.001	0.939 \pm 0.007	0.962 \pm 0.004	3.58 \pm 0.25	6.05 \pm 0.44
	MS	0.987 \pm 0.001	0.985 \pm 0.001	0.986 \pm 0.000	2.03 \pm 0.06	2.81 \pm 0.10
	ET	0.985 \pm 0.001	0.960 \pm 0.005	0.972 \pm 0.003	2.87 \pm 0.17	5.05 \pm 0.41
MS	HC	0.985 \pm 0.001	0.987 \pm 0.000	0.986 \pm 0.000	2.18 \pm 0.09	2.89 \pm 0.07
	AD	0.982 \pm 0.001	0.948 \pm 0.006	0.965 \pm 0.003	3.60 \pm 0.25	5.79 \pm 0.40
	PD	0.983 \pm 0.001	0.984 \pm 0.001	0.984 \pm 0.000	2.26 \pm 0.12	3.16 \pm 0.17
	ET	0.983 \pm 0.001	0.944 \pm 0.007	0.963 \pm 0.003	2.90 \pm 0.15	5.44 \pm 0.40
ET	HC	0.968 \pm 0.004	0.986 \pm 0.000	0.977 \pm 0.002	5.37 \pm 0.66	5.03 \pm 0.49
	AD	0.983 \pm 0.001	0.971 \pm 0.002	0.977 \pm 0.001	3.12 \pm 0.19	4.53 \pm 0.27
	PD	0.989 \pm 0.000	0.985 \pm 0.001	0.987 \pm 0.000	1.96 \pm 0.11	2.64 \pm 0.07
	MS	0.984 \pm 0.002	0.984 \pm 0.000	0.984 \pm 0.001	2.68 \pm 0.33	3.19 \pm 0.21

GCL-BM MAE_C , are achieved by the models that were trained on HC images. This can be indicative that models trained on images that do not display the visual features that characterise these diseases can have a reduced performance when used to measure the retinal features of ND patients. Furthermore, the models trained on HC images seem to over-segment the RNFL on images of ND patients. On average, these models achieve the highest RNFL Sensitivity (0.941 ± 0.015) while at the same time achieving the lowest Precision (0.901 ± 0.037). A thinning of this layer is most commonly associated with ND progression.

This loss of performance is not seen when grouping these results by the set used for testing, indicating that this may be a lack of representation of the relevant visual features, rather than HC images being visually distinct (i.e. HC images are not representative enough of ND images, while ND images are of HC images). This uni-directional relationship is also apparent with the images belonging to the PD patients. Models that only used PD images for training produce the second highest MAE for both layers and metrics, while the other sets are all separately able to segment PD images with the lowest error.

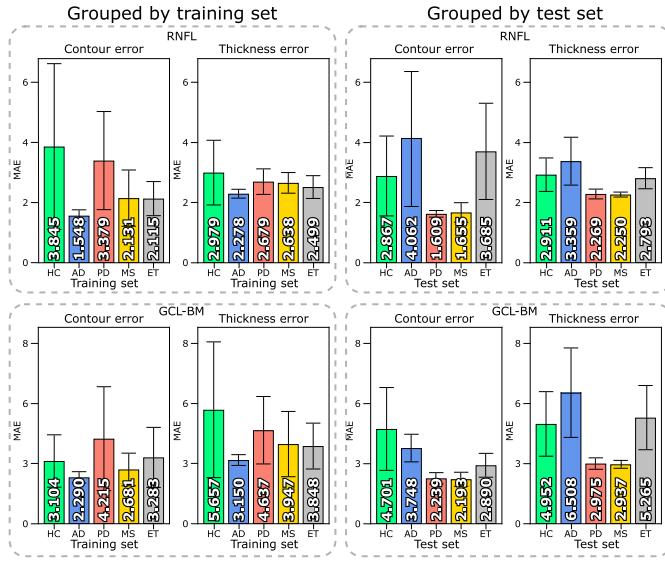


Fig. 6. Comparative of the MAE calculated using layer contours and thickness for the second experiment, averaged and grouped by the set used for training and the set used for testing. Error bars indicate standard deviation. Smaller MAE is better.

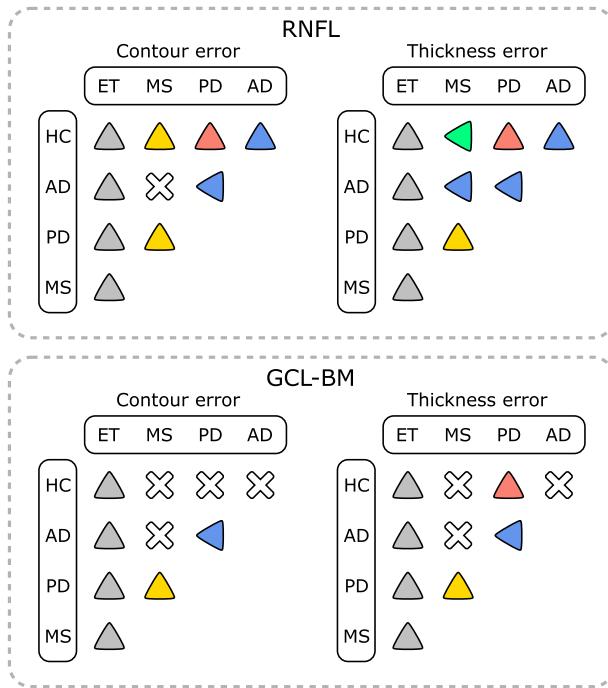


Fig. 7. Results from the Wilcoxon signed-rank test comparing the results of the models based on the images used for training. The images shared in the test set between the two models are used in the comparison. For each point in every matrix, an arrow points to the training set used for the model that produced a lower error with statistical significance with $p < 0.05$, indicative of better performance. A cross indicates no statistically significant difference was found.

On the other hand, AD images seem to allow the models a better representation of the visual features present in all the other NDs. Models trained with AD images achieved the lowest MAE of all those considered for both layers. At the same time, all of the other models achieve the highest MAE_T for both layers and MAE_C for the RNFL when tested on the AD images. It is possible that these images display a higher variability of the neural degeneration associated with the other NDs. This is also supported by the models trained on AD achieving better results on all NDs, but worse on the HC images (Tables III and IV). Furthermore, the worst results achieved by the model trained on HC images are the ones regarding the test on AD images. Other works in the literature aimed at the classification of NDs using other imaging modalities have also reported images of AD patients harder to classify, indicating that this disease may show signs that can be confused with those of other NDs [70], [71]. It bears mentioning that this study only considered early cases of AD, and the results seem to indicate that even in these early stages, there are noticeable changes in the displayed visual features that can have a noticeable effect on model performance.

Models trained with images of ET patients achieve generally well-rounded results when tested on images from the other NDs, while performing slightly worse on HC images. Nevertheless, the statistical tests represented in Fig. 7 show that models trained on ET images are able to achieve significantly lower error than every other training set in all of the comparisons that were performed. As with the AD patients, it is possible that images from ET patients display a wider variability of disease-specific features that allow models trained on them to better adapt to other NDs. Additionally, models trained on images of other NDs perform comparatively worse when tested on ET images, although the magnitude of this error is still lower than when AD images are used for testing. This, in turn, means that models intended for the automatic segmentation of these images may require a greater sample of images to compensate for the lack of representation of other NDs and, specially, that of HC images.

Overall, the results show that the choice of training images can have a noticeable effect on retinal layer segmentation models. The use of HC images for training can produce models that may mask the effects and progression of ND. These effects are subtle, and arise when taking into account the general performance of each model. A qualitative comparison of the segmentation masks produced by the models of the first and second experiment can be seen in Fig. 8. This figure shows that the models provide a robust and accurate segmentation of the retinal layers while showing minor differences between them. Although not considered in this study, other ocular pathologies such as diabetic macular oedema or age-related macular degeneration may interact with these diseases and further affect segmentation models, producing more deviation in overall thickness measurements.

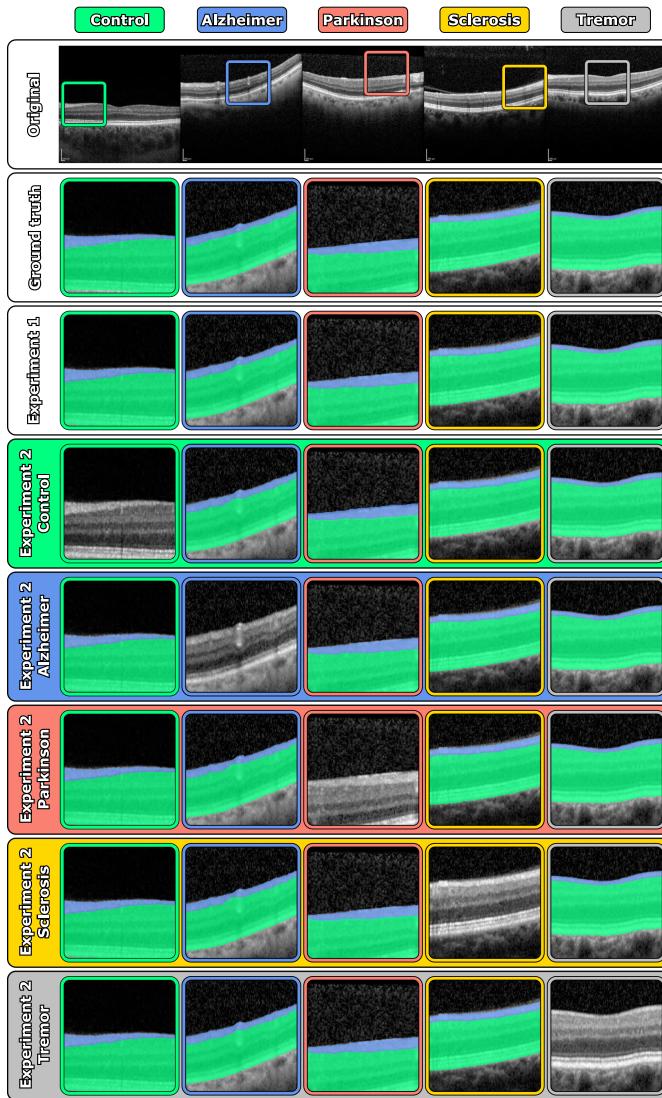


Fig. 8. Qualitative comparison of a series of samples of ground truth and automatic segmentation masks for each of the pathology sets. Every column displays an image representative of each image set. Every row shows the corresponding segmentation masks, expanded for ease of comparison. Experiment 2 rows show the model that was trained on each image set. The training set for each model is not represented.

IV. CONCLUSION

Retinal OCT imaging can provide a non-invasive window of analysis of several NDs. The degeneration caused by these diseases can have a measurable effect on the thickness of several retinal layers. Different NDs have distinct causes and can have a different impact on neural tissue, which can translate into a loss of performance of automated retinal layer segmentation models that were not trained using images representative of these changes. In this work, we have presented a fully automatic approach for the automatic segmentation of retinal layers in multiple disease scenarios of the four most representative NDs (namely, AD, PD, MS, and ET), as well as a set of HC patients. The proposed approach has been meticulously developed and extensively evaluated on a dataset comprising images extracted

under identical conditions for these four representative NDs. These images have been accurately annotated, providing the precise location of the retinal layers. So far, this dataset represents the most complete collection of retinal OCT imaging extracted from ND patients, ensuring a robust and comprehensive analysis of the proposed methods. Taking advantage of this dataset, the proposed approach has been analysed and validated in a two-part in-depth study aimed at comprehensively examining the effects that the presence or absence of representative images of each ND can have on deep learning models intended for retinal segmentation. The two experiments that were conducted allow a comprehensive analysis of the impact that the different features that characterise these diseases can have on models trained without enough representation of said features. The results show that training models in images of HC patients can translate into a loss of performance when segmenting images of ND patients. Furthermore, these models tend to mask the sign of disease, which can be detrimental to the diagnosis process of NDs. The models that were trained on images of all NDs but one are able to perform at a level similar to the baseline models which were allowed to see images of every ND during training. These results indicate that it may be necessary to take into account the presence or absence of patients of these diseases in datasets used to train automatic retinal layer segmentation models, since it can have an impact on performance. By exposing the models to the multiple scenarios of ND during the training, the proposed approach can aid in the diagnosis of these diseases in a reliable way, helping improve patient care in clinical practice.

Early studies attempting to achieve an early diagnosis of NDs through neuro-retinal measurements have primarily relied on numerical thickness data provided by OCT. This study, however, takes a different approach by focusing on the analysis of images without relying on any particular commercial software. The proposed approach has effectively addressed the challenges associated with the lack of representativity in HC image. This enables the development of robust and ND-oriented retinal layer segmentation models, which can be crucial for accurately tracking disease progression in patients affected by these conditions. Moreover, these findings highlight that neuro-retinal images exhibit distinct characteristics that depend on the specific ND that affects the patient. This opens the possibility for differential diagnosis between different neurological disorders. By facilitating an accurate and early differentiation, this can assist clinicians in reaching a definitive diagnosis much sooner which, in turn, allows for the timely establishment of appropriate therapies, ultimately improving overall patient quality of life while slowing down the progression of the disease.

Future plans include a more detailed analysis of the different retinal layers that were included in the GCL-BM set such as the GCL-inner plexiform layer complex, as well as the separate study of the sectors that conform the parafoveal area, which can enable a more comprehensive analysis of the effects of these diseases and the impact that these may have on computer-aided diagnosis systems that make use of automatic retinal layer segmentation. Additionally, the approach and study proposed in this work could be extended to other retinal layer segmentation architectures based on different techniques such as transformers.

Moreover, the development of an automated tool that can enable the differential diagnosis of these diseases can be of great benefit in order to achieve an early treatment of ND.

REFERENCES

- [1] J. Bradshaw, "Fluctuating cognition in dementia with Lewy bodies and Alzheimers disease is qualitatively distinct," *J. Neurol. Neurosurg. Psychiatry*, vol. 75, no. 3, pp. 382–387, Mar. 2004.
- [2] T. A. Salthouse, J. R. Nesselroade, and D. E. Berish, "Short-term variability in cognitive performance and the calibration of longitudinal change," *J. Gerontol. Ser. B: Psychol. Sci. Social Sci.*, vol. 61, no. 3, pp. P144–P151, May 2006.
- [3] C. K. Glass, K. Saijo, B. Winner, M. C. Marchetto, and F. H. Gage, "Mechanisms underlying inflammation in neurodegeneration," *Cell*, vol. 140, no. 6, pp. 918–934, Mar. 2010.
- [4] A. J. Mejia-Vergara, R. Karanjia, and A. A. Sadun, "OCT parameters of the optic nerve head and the retina as surrogate markers of brain volume in a normal population, a pilot study," *J. Neurological Sci.*, vol. 420, Jan. 2021, Art. no. 117213.
- [5] T. E. Yap, S. I. Balendra, M. T. Almonte, and M. F. Cordeiro, "Retinal correlates of neurological disorders," *Therapeutic Adv. Chronic Dis.*, vol. 10, Jan. 2019, Art. no. 204062231988220.
- [6] D. Huang et al., "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, Nov. 1991.
- [7] T. C. Chen, "Spectral domain optical coherence tomography," *Arch. Ophthalmol.*, vol. 123, no. 12, Dec. 2005, Art. no. 1715.
- [8] S. Wolf and U. Wolf-Schnurrbusch, "Spectral-domain optical coherence tomography use in macular diseases: A review," *Ophthalmologica*, vol. 224, no. 6, pp. 333–340, 2010.
- [9] A. Ha and K. H. Park, "Optical coherence tomography for the diagnosis and monitoring of glaucoma," *Asia-Pacific J. Ophthalmol.*, vol. 8, no. 2, pp. 135–145, 2019.
- [10] T. C. Chen, A. Zeng, W. Sun, M. Mujat, and J. F. de Boer, "Spectral domain optical coherence tomography and glaucoma," *Int. Ophthalmol. Clin.*, vol. 48, no. 4, pp. 29–45, 2008.
- [11] A. Invernizzi, M. Cozzi, and G. Staurenghi, "Optical coherence tomography and optical coherence tomography angiography in uveitis: A review," *Clin. Exp. Ophthalmol.*, vol. 47, no. 3, pp. 357–371, Mar. 2019.
- [12] G. Virgili et al., "Optical coherence tomography versus stereoscopic fundus photography or biomicroscopy for diagnosing diabetic macular edema: A systematic review," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 11, pp. 4963–4973, Nov. 2007.
- [13] A. A. Khanifar, A. F. Koreishi, J. A. Izatt, and C. A. Toth, "Drusen ultrastructure imaging with spectral domain optical coherence tomography in age-related macular degeneration," *Ophthalmology*, vol. 115, no. 11, pp. 1883–1890, Nov. 2008.
- [14] M. Satue et al., "Optical coherence tomography as a biomarker for diagnosis, progression, and prognosis of neurodegenerative diseases," *J. Ophthalmol.*, vol. 2016, pp. 1–9, 2016.
- [15] P. J. Snyder et al., "Retinal imaging in Alzheimers and neurodegenerative diseases," *Alzheimer's Dement.*, vol. 17, no. 1, pp. 103–111, Oct. 2020.
- [16] V. T. Chan et al., "Spectral-domain OCT measurements in Alzheimer's disease," *Ophthalmology*, vol. 126, no. 4, pp. 497–510, Apr. 2019.
- [17] K. L. Thomson, J. M. Yeo, B. Waddell, J. R. Cameron, and S. Pal, "A systematic review and meta-analysis of retinal nerve fiber layer change in dementia, using optical coherence tomography," *Alzheimers Dementia: Diagnosis, Assessment Dis. Monit.*, vol. 1, no. 2, pp. 136–143, Apr. 2015.
- [18] B. M. Wong et al., "Validation of optical coherence tomography retinal segmentation in neurodegenerative disease," *Transl. Vis. Sci. Technol.*, vol. 8, no. 5, Sep. 2019, Art. no. 6.
- [19] E. M. Lad et al., "Evaluation of inner retinal layers as biomarkers in mild cognitive impairment to moderate Alzheimer's disease," *PLoS One*, vol. 13, no. 2, Feb. 2018, Art. no. e0192646.
- [20] E. Salobrar-Garcia et al., "Analysis of retinal peripapillary segmentation in early Alzheimer's disease patients," *Biomed Res. Int.*, vol. 2015, pp. 1–8, 2015.
- [21] E. M. Shrier, C. R. Adam, B. Spund, S. Glazman, and I. Bodis-Wollner, "Interocular asymmetry of foveal thickness in Parkinson disease," *J. Ophthalmol.*, vol. 2012, pp. 1–6, 2012.
- [22] S. Slotnick et al., "A novel retinal biomarker for Parkinsons disease: Quantifying the foveal pit with optical coherence tomography," *Movement Disord.*, vol. 30, no. 12, pp. 1692–1695, Sep. 2015.
- [23] M. Pulicken, E. Gordon-Lipkin, L. J. Balcer, E. Frohman, G. Cutter, and P. A. Calabresi, "Optical coherence tomography and disease subtype in multiple sclerosis," *Neurology*, vol. 69, no. 22, pp. 2085–2092, Nov. 2007.
- [24] C. H. Polman et al., "Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria," *Ann. Neurol.*, vol. 69, no. 2, pp. 292–302, Feb. 2011.
- [25] S. Motamedi et al., "Normative data and minimally detectable change for inner retinal layer thicknesses using a semi-automated OCT image segmentation pipeline," *Front. Neurol.*, vol. 10, Nov. 2019, Art. no. 1117.
- [26] P. J. Patel, F. K. Chen, L. da Cruz, and A. Tufail, "Segmentation error in stratus optical coherence tomography for neovascular age-related macular degeneration," *Invest. Ophthalmol. Vis. Sci.*, vol. 50, no. 1, pp. 399–404, Jan. 2009.
- [27] A. Montuoro, S. M. Waldstein, B. S. Gerendas, U. Schmidt-Erfurth, and H. Bogunović, "Joint retinal layer and fluid segmentation in OCT scans of eyes with severe macular edema using unsupervised representation and auto-context," *Biomed. Opt. Exp.*, vol. 8, no. 3, pp. 1874–1888, Feb. 2017.
- [28] H. Ishikawa, D. M. Stein, G. Wollstein, S. Beaton, J. G. Fujimoto, and J. S. Schuman, "Macular segmentation with optical coherence tomography," *Invest. Ophthalmol. Vis. Sci.*, vol. 46, no. 6, pp. 2012–2017, Jun. 2005.
- [29] S. J. Chiu, M. J. Allingham, P. S. Mettu, S. W. Cousins, J. A. Izatt, and S. Farsiu, "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Exp.*, vol. 6, no. 4, pp. 1172–1194, Mar. 2015.
- [30] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Express*, vol. 18, no. 18, pp. 19413–19428, Aug. 2010.
- [31] L. Fang, D. Cunefare, C. Wang, R. H. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Exp.*, vol. 8, no. 5, pp. 2732–2744, Apr. 2017.
- [32] K. Hu, B. Shen, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Automatic segmentation of retinal layer boundaries in OCT images using multi-scale convolutional neural network and graph search," *Neurocomputing*, vol. 365, pp. 302–313, Nov. 2019.
- [33] T. Hassan, M. U. Akram, M. F. Masood, and U. Yasin, "Deep structure tensor graph search framework for automated extraction and characterization of retinal layers and fluid pathology in retinal SD-OCT scans," *Comput. Biol. Med.*, vol. 105, pp. 112–124, Feb. 2019.
- [34] J. Kugelman, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Automatic segmentation of OCT retinal boundaries using recurrent neural networks and graph search," *Biomed. Opt. Exp.*, vol. 9, no. 11, pp. 5759–5777, Oct. 2018.
- [35] T. Fu, X. Liu, D. Liu, and Z. Yang, "A deep convolutional feature based learning layer-specific edges method for segmenting OCT image," in *Proc. SPIE*, vol. 10420, 2017, Art. no. 1042029.
- [36] A. Shah, M. D. Abramoff, and X. Wu, "Simultaneous multiple surface segmentation using deep learning," in *Proc. Int. Workshop Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, 2017, pp. 3–11.
- [37] A. Shah, L. Zhou, M. D. Abrámoff, and X. Wu, "Multiple surface segmentation using convolution neural nets: Application to retinal layer segmentation in OCT images," *Biomed. Opt. Exp.*, vol. 9, no. 9, pp. 4509–4526, Aug. 2018.
- [38] J. Hamwood, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers," *Biomed. Opt. Exp.*, vol. 9, no. 7, pp. 3049–3066, Jun. 2018.
- [39] J. A. Sousa et al., "Automatic segmentation of retinal layers in OCT images with intermediate age-related macular degeneration using u-net and DexiNed," *PLoS One*, vol. 16, no. 5, May 2021, Art. no. e0251591.
- [40] A. Lang et al., "Retinal layer segmentation of macular OCT images using boundary classification," *Biomed. Opt. Exp.*, vol. 4, no. 7, pp. 1133–1152, Jun. 2013.
- [41] Y. Liu et al., "Layer boundary evolution method for macular OCT layer segmentation," *Biomed. Opt. Exp.*, vol. 10, no. 3, pp. 1064–1080, Feb. 2019.
- [42] Y. He et al., "Deep learning based topology guaranteed surface and MME segmentation of multiple sclerosis subjects from retinal OCT," *Biomed. Opt. Exp.*, vol. 10, no. 10, pp. 5042–5058, Sep. 2019.
- [43] B. Wang, W. Wei, S. Qiu, S. Wang, D. Li, and H. He, "Boundary aware u-net for retinal layers segmentation in optical coherence tomography images," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 8, pp. 3029–3040, Aug. 2021.
- [44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.

- [45] Y. He et al., "Structured layer surface segmentation for retina OCT using fully convolutional regression networks," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101856.
- [46] S. K. Yadav et al., "Intraretinal layer segmentation using cascaded compressed U-Nets," *J. Imag.*, vol. 8, no. 5, May 2022, Art. no. 139.
- [47] Y. Wu et al., "Training deep learning models to work on multiple devices by cross-domain learning with no additional annotations," *Ophthalmology*, vol. 130, no. 2, pp. 213–222, Sep. 2022.
- [48] M. Satue 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*, vol. 37, no. 7, pp. 1314–1319, Jun. 2023.
- [49] B. Tugeu et al., "Evaluation of retinal alterations in Parkinson disease and tremor diseases," *Acta Neurologica Belgica*, vol. 120, no. 1, pp. 107–113, Nov. 2019.
- [50] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*, 4th ed. Washington, DC, USA: Autor, 1994.
- [51] A. J. Thompson et al., "Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria," *Lancet Neurol.*, vol. 17, no. 2, pp. 162–173, Feb. 2018.
- [52] D. J. Gelb, E. Oliver, and S. Gilman, "Diagnostic criteria for Parkinson disease," *Arch. Neurol.*, vol. 56, no. 1, pp. 33–39, Jan. 1999.
- [53] M. M. Hoehn and M. D. Yahr, "Parkinsonism: Onset, progression, and mortality," *Neurology*, vol. 17, no. 5, pp. 427–427, May 1967.
- [54] H. Reichmann, "Clinical criteria for the diagnosis of Parkinson's disease," *Neurodegener. Dis.*, vol. 7, no. 5, pp. 284–290, 2010.
- [55] G. Deuschl, P. Bain, and M. Brin, "Consensus statement of the movement disorder society on tremor," *Movement Disord.*, vol. 13, no. S3, pp. 2–23, Oct. 2008.
- [56] S. Siik et al., "Lens autofluorescence and light scatter in relation to the lens opacities classification system, LOCS III," *Acta Ophthalmologica Scandinavica*, vol. 77, no. 5, pp. 509–514, Oct. 1999.
- [57] A. Petzold et al., "Artificial intelligence extension of the OSCAR-IB criteria," *Ann. Clin. Transl. Neurol.*, vol. 8, no. 7, pp. 1528–1542, May 2021.
- [58] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.
- [59] G. Bradski, "The openCV library," *Dr. Dobb's J. Softw. Tools*, vol. 25, pp. 120–125, 2000.
- [60] J. Li et al., "Multi-scale GCN-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary OCT images," *Biomed. Opt. Exp.*, vol. 12, no. 4, pp. 2204–2220, 2021.
- [61] A. G. Roy et al., "ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Exp.*, vol. 8, no. 8, pp. 3627–3642, Jul. 2017.
- [62] S. K. Devalla et al., "DRUNet: A dilated-residual u-net deep learning network to segment optic nerve head tissues in optical coherence tomography images," *Biomed. Opt. Exp.*, vol. 9, no. 7, pp. 3244–3265, Jul. 2018.
- [63] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [64] G. H. Chen, "Deep kernel survival analysis and subject-specific survival time prediction intervals," in *Proc. 5th Mach. Learn. Healthcare Conf.*, 2020, vol. 126, pp. 537–565.
- [65] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," *NPJ Digit. Med.*, vol. 1, no. 1, May 2018, Art. no. 18.
- [66] N. J. Hart, Y. Koronyo, K. L. Black, and M. Koronyo-Hamaoui, "Ocular indicators of Alzheimer's: Exploring disease in the retina," *Acta Neuropathol.*, vol. 132, no. 6, pp. 767–787, Sep. 2016.
- [67] L. G. M. Mello, L. B. Bissoli, F. P. Saraiva, R. de Paula Doyle Maia, and M. L. R. Monteiro, "Retinal layers and choroid measurements in Parkinsons disease with or without pramipexole treatment," *Movement Disord.*, vol. 35, no. 12, pp. 2357–2359, Sep. 2020.
- [68] S. Saidha et al., "Visual dysfunction in multiple sclerosis correlates better with optical coherence tomography derived estimates of macular ganglion cell layer thickness than peripapillary retinal nerve fiber layer thickness," *Mult. Scler. J.*, vol. 17, no. 12, pp. 1449–1463, Aug. 2011.
- [69] G. Mostile et al., "Clinical-instrumental patterns of neurodegeneration in essential tremor: A data-driven approach," *Parkinsonism Related Disord.*, vol. 87, pp. 124–129, Jun. 2021.
- [70] D. Payares-Garcia, J. Mateu, and W. Schick, "Spatially informed Bayesian neural network for neurodegenerative diseases classification," *Statist. Med.*, vol. 42, no. 2, pp. 105–121, Nov. 2022.
- [71] K. R. Bhatele, A. Jha, K. Kapoor, and D. Tiwari, "Neurodegenerative diseases-caps: A capsule network based early screening system for the classification of neurodegenerative diseases," *Cogn. Neurodyn.*, vol. 16, no. 6, pp. 1361–1377, Feb. 2022.

Open Access provided by 'Universidade da Coruña/CISUG' within the CRUI-CARE Agreement.