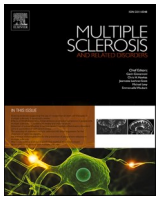
Multiple Sclerosis and Related Disorders 74 (2023) 104725

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

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



Diagnosis of multiple sclerosis using optical coherence tomography supported by artificial intelligence

Miguel Ortiz [a], Victor Mallen [b,c], Luciano Boquete [d], Eva M. Sánchez-Morla [e], Beatriz Cordón [b,c], Elisa Vilades [b,c], Francisco J. Dongil-Moreno [d], Juan M. Miguel-Jiménez [d], Elena Garcia-Martin [b,c,*]

[a] School of Physics, University of Melbourne, Melbourne, VIC 3010, Australia
[b] Department of Ophthalmology, Miguel Servet University Hospital, Zaragoza, Spain
[c] Aragon Institute for Health Research (IIS Aragon). Miguel Servet Ophthalmology Innovation and Research Group (GIMSO), University of Zaragoza, Spain
[d] Biomedical Engineering Group, Department of Electronics, University of Alcalá, Alcalá de Henares, Spain
[e] Faculty of Medicine, Complutense University of Madrid, Madrid 28040, Spain

ARTICLE INFO

Keywords:
Multiple sclerosis
Optical coherence tomography
Biomarker
Convolutional neural network

ABSTRACT

Background: Current procedures for diagnosing multiple sclerosis (MS) present a series of limitations, making it critically important to identify new biomarkers. The aim of the study was to identify new biomarkers for the early diagnosis of MS using spectral-domain optical coherence tomography (OCT) and artificial intelligence.
Methods: Spectral domain OCT was performed on 79 patients with relapsing-remitting multiple sclerosis (RRMS) (disease duration $\leq$ 2 years, no history of optic neuritis) and on 69 age-matched healthy controls using the posterior pole protocol that incorporates the anatomic Positioning System. Median retinal thickness values in both eyes and inter-eye difference in healthy controls and patients were evaluated by area under the receiver operating characteristic (AUROC) curve analysis in the foveal, parafoveal and perifoveal areas and in the overall area spanned by the three rings. The structures with the greatest discriminant capacity — retinal thickness and inter-eye difference — were used as inputs to a convolutional neural network to assess the diagnostic capability.
Results: Analysis of retinal thickness and inter-eye difference in RRMS patients revealed that greatest alteration occurred in the ganglion cell (GCL), inner plexiform (IPL), and inner retinal (IRL) layers. By using the average thickness of the GCL (AUROC = 0.82) and the inter-eye difference in the IPL (AUROC = 0.71) as inputs to a two-layer convolutional neural network, automatic diagnosis attained accuracy = 0.87, sensitivity = 0.82, and specificity = 0.92.
Conclusion: This study adds weight to the argument that neuroretinal structure analysis could be incorporated into the diagnostic criteria for MS.

1. Introduction

Studies demonstrating the potential feasibility of using neuroretinal thickness values, obtained by optical coherence tomography (OCT), to demonstrate dissemination in space and time in multiple sclerosis (MS) diagnosis have increased considerably in recent years (Alonso et al., 2018; Petzold et al., 2017). The neuroaxonal damage produced by MS manifests as thinning of both the peripapillary retinal nerve fibre layer (pRNFL) and the macular ganglion cell and inner plexiform layer (GCIPL) and this neuroaxonal loss correlates with physical disability in patients with relapsing-remitting multiple sclerosis (RRMS) (Garcia-Martin et al., 2017; Petzold et al., 2017). There is also evidence to

Abbreviations: APS, anatomical positioning system; AUROC, area under the receiver operating characteristic; CNN, convolutional neural network; EDSS, expanded disability status scale; ETDRS, early treatment diabetic retinopathy study; GCIPL, ganglion cell and inner plexiform layer; GCL, ganglion cell layer; HC, healthy controls; INL, inner nuclear layer; IPL, inner plexiform layer; IRL, inner retinal layers; ML, machine learning; MS, multiple sclerosis; OCT, optical coherence tomography; ONL, outer nuclear layer; OPL, outer plexiform layer; ORL, outer retinal layers; pRNFL, peripapillary retinal nerve fibre layer; QALY, Quality-Adjusted Life Years; RNFL, retinal nerve fibre layer; RPE, retinal pigment epithelium; RRMS, relapsing-remitting multiple sclerosis; SD, standard deviation.

* Corresponding author.
E-mail address: egmvivax@yahoo.com (E. Garcia-Martin).

https://doi.org/10.1016/j.msard.2023.104725
Received 16 February 2023; Received in revised form 15 March 2023; Accepted 16 April 2023
Available online 17 April 2023

suggest that thickening of the inner nuclear layer (INL) may be associated with inflammatory activity in MS (Bsteh et al., 2020).

As MS progresses, inter-eye structural differences develop both in eyes with a clinical history of optic neuritis and in those without such a history (Kenney et al., 2022). It was therefore proposed to use inter-eye difference cut-offs of pRNFL $\geq$ 5 μm and GCIPL $\geq$ 4 μm to identify unilateral optic neuritis (Nolan et al., 2018; Nolan-Kenney et al., 2019). Values close to these thresholds were subsequently confirmed in further studies (Bsteh et al., 2020).

Nevertheless, further validation is required using other exploratory protocols (e.g. Early Treatment Diabetic Retinopathy Study (ETDRS) and the Posterior Pole protocol) and other technologies such as swept-source OCT (Yasin Alibhai et al., 2018). The Posterior Pole protocol is based on the high reliability provided by applying the Anatomical Positioning System (APS) to OCT imaging. The APS works by locating points on the eye using two fixed structural landmarks: the centre of the fovea and the centre of the Bruch membrane opening. This allows much more precise monitoring, detecting even the slightest alteration or inter-eye difference because using the APS, in conjunction with the True Track (TruTrack, Heidelberg Engineering) eye-tracking system, ensures accurate identification of macula position in each eye based on head tilt and eye cyclotorsion.

Machine learning (ML) techniques are capable of identifying complex relationships or patterns in empirical data and applying that knowledge to new datasets. Conventional neural networks, support vector machines and other traditional automatic classifiers can be used in MS diagnosis supported by OCT if the available information comprises peripapillary measurements or the average thickness values in the regions defined by the ETDRS grid (Cavaliere et al., 2019; Ciftci Kavaklioglu et al., 2022; Garcia-Martin et al., 2021; Kenney et al., 2022). When analysing a greater number of measurements, such as those produced by a 45 $\times$ 60 thicknesses matrix, it is preferable to use a convolutional neural network (CNN) (López-Dorado et al., 2021).

Although according to the McDonald criteria there is not yet sufficient evidence to consider OCT tests as biomarkers of MS, further research, development, and clinical implementation of the technology is considered a high priority (Thompson et al., 2018).

The aim of this paper is to evaluate the extent to which inter-eye asymmetry and the thicknesses of different retinal structures obtained using the Posterior Pole protocol are able to discriminate between control subjects and patients with recently diagnosed MS. In a second phase, the dataset with the greatest discriminant capacity is selected and used as the input of a CNN to implement an assisted diagnosis system.

2. Methods

The design of this study adhered to the tenets of the Declaration of Helsinki and all participants provided written informed consent. The study protocol was approved by the Clinical Research Ethics Committee of Aragon (Zaragoza, Spain).

2.1. Study cohort

Two independent samples, one comprising RRMS patients and the other comprising healthy controls (HC), were prospectively recruited from two clinics (ophthalmology clinic specializing in neuro-ophthalmology and neurology clinic specializing in demyelinating diseases).

Based on our preliminary studies in MS patients, we calculated the sample size needed to detect differences of at least 3 μm in OCT-measured thicknesses by applying a bilateral test with α 5% risk and β 10% risk (i.e. with a power of 90%). In order to obtain a sufficient sample of MS patients to allow an in-depth study of the natural history of the disease, the non-exposed/exposed ratio was determined to be 0.5. From these data it was concluded that at least 50 eyes would be needed in each group.

The definitive MS diagnosis was based on standard clinical and neuroimaging criteria (Thompson et al., 2018). To ensure a homogeneous population, only patients with the RRMS phenotype without a history of optic neuritis in either eye were included. The healthy controls had no history of ocular or neurological disease and presented no signs or symptoms of them.

The exclusion criteria were best-corrected visual acuity lower than 0.5 according to 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 lower than 2 according to the Lens Opacities Classification System III) (Chylack, 1993), concomitant ocular disease (including glaucoma or retinal pathology), and other systemic conditions potentially affecting the visual system.

Related medical records were carefully reviewed, including disease duration, Expanded Disability Status Scale (EDSS) score, treatments, and the presence of prior episodes of optic neuritis.

All participants underwent a full ophthalmological examination, including clinical history, visual acuity, biomicroscopy of the anterior segment using a slit lamp, Goldmann applanation tonometry, and ophthalmoscopy of the posterior segment to check that they did not meet the exclusion criteria. OCT measurements of the neuroretinal structure were then taken.

2.2. OCT method

Structural measurements of the retina were obtained using the Spectralis OCT device (Heidelberg Engineering Inc., Germany). The Posterior Pole Retinal Thickness Map protocol (Asrani et al., 2011) was used for all subjects. This protocol incorporates the APS, which describes a horizontal line between the fovea and the Bruch membrane opening. Based on that reference line, 61 parallel B-scans are performed inside an area measuring 25° x 30° Each B-scan consists of 768 A-scans. There are 123 μm of spacing between B-scans and 10 frames are averaged per B-scan location. The APS plus the True Track (TruTrack, Heidelberg Engineering) eye-tracking system ensure accurate identification of the position of the macula in each individual based on head tilt and eye cyclotorsion.

The 25° x 30° analysed area is represented as an 8 $\times$ 8 grid that provides overall retinal thickness, and segments each layer thickness into 64 independent cells. The Spectralis OCT device has an axial (in tissue) resolution of 3.9 μm and the thickness value is obtained from the average of each cell (860 $\times$ 860 μm).

For the 64 cells, the segmentation software (HRA version 6.0.7.0) used by the Spectralis OCT device provided the mean thickness of the following nine retinal layers (Fig. 1): retinal nerve fibre layer (RNFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), and outer nuclear layer (ONL), as well as the retinal pigment epithelium (RPE) and the mean thickness of the neuroretina between the internal and external limiting membranes (inner retinal layers: IRL), and the mean thickness of the outer retinal layers (ORL) between the external limiting membrane and the Bruch membrane.

No manual correction was applied to the OCT output. The quality of the scans was assessed prior to analysis and poor-quality scans or images with less than 25/40 quality points were rejected (Balasubramanian et al., 2009). The quality of the images was checked using the OSCAR-IB guidelines (Petzold et al., 2021a).

Due to the number of cells with data, and in the interest of better visualization and understanding of the information, the descriptive data of 64 cells were grouped into the foveal, parafoveal and perifoveal areas; analysis of the region formed by grouping these three rings together (GLOBAL) is also included.

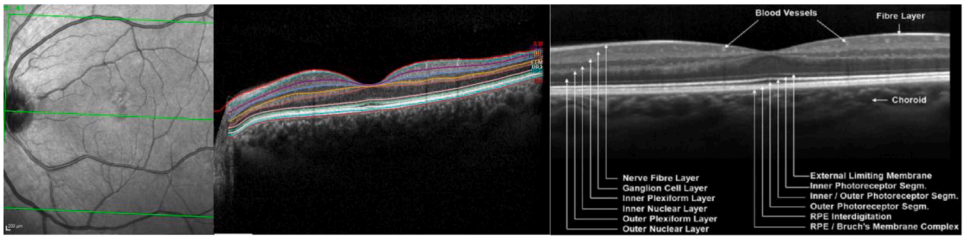


Fig. 1. Visualization of the retinal layers by Spectralis Optical Coherence Tomography (OCT). The left image shows the line of the tomographic slice and its location within the retina, between the optic nerve and the macula of the eye. The middle image shows the retinal layers visualised by OCT in one of our patients and the delineation of the layers made by the OCT software in an automated way, using greyscale differences. Abbreviations: ILM: inner limiting membrane; GCL: ganglion cell layer; IPL: inner plexiform layer; INL: inner nuclear layer; OPL: outer plexiform layer; ELM: external limiting membrane (the structure that separates the inner segments from the outer nuclear layer); PR1: photoreceptor inner segments; PR2: photoreceptor outer segments; RPE: retinal pigment epithelium; BM: Bruch membrane. The right figure shows a schematic of the retinal layers that can be identified by OCT.

2.3. Statistical methods

In continuous variables the normality criterion was evaluated using the Shapiro-Wilk test. Normally distributed data were presented as the mean and the standard deviation (SD). The median and interquartile range (Q3–Q1) were shown for non-Gaussian data. Categorical data were presented as absolute numbers with their corresponding percentages. Difference analysis was performed using Student's *t*-test and $\chi 2$ for the categorical variables.

The area under the receiver operating characteristic (AUROC) curve was calculated to test whether two distributions of thickness data or the inter-eye difference were similar or not. This test is identical to the non-parametric Mann–Whitney U statistic (Hanley and McNeil, 1982). The DeLong non-parametric test was performed to evaluate the statistical difference between the AUROC values. P-values < 0.05 were considered to indicate a significant difference. Statistical analysis was performed with Python 3.10.4.

2.4. Convolutional neural network

A CNN is a feed-forward artificial neural network in which the first processing layers perform feature map extraction based on the convolution between input information and the kernels of certain dimensions (Alzubaidi et al., 2021). This process can be configured to span several convolution layers, in which it is also common to implement subsampling, pooling, batch normalization and non-linear transformation operations in order to improve CNN training. If the input information comprises images, the CNN can detect local features such as edges, corners, luminance, etc. After one or more convolution layers, arranged in a pipeline structure, a classification module (e.g. a fully connected neural network) is implemented and the CNN output is generated. The CNN's trainable parameters can be adjusted using one of the variants of the backpropagation algorithm (e.g. Adam optimizer).

3. Results

3.1. Participating subjects

A total of 79 patients (68 women and 11 men) with RRMS and 69 (46 women and 23 men) HCs were enroled. Mean age was 45.64 ± 13.59 years in MS group and 46.94 ± 12.64 years in healthy controls group ($p = 0.604$). Mean disease duration in patients was 1.42 ± 0.72 years and median EDSS score was 1.28 (range: 0–3). Treatment distribution in MS group was 3 patients with Avonex (intramuscular interferon beta-1a), 7 patients with Betaseron (inteferon beta-1b), 6 patients with Rebif (subcutaneous interferon beta-1a), 6 patients with Copaxone (glatiramer acetate), 55 patients received oral last generation treatments: Tecfidera (dimetilfumarato) for 11 patients, Gilenya (fingolimod) for 32 patients, Mayzent (siponimod) for 3 patients and Aubagio (teriflunomida) for 9 patients and 2 patients did not received treatment.

3.2. Retinal structure thicknesses

Fig. 2 shows the median average thickness values for both eyes of the nine retinal structures, represented in heat map form, both for controls and patients. As would be expected, in most of the layers and zones analysed the thicknesses are greater in the control subjects.

Fig. 3 presents a graphical representation of the AUROC value in each of the cells analysed, revealing an annular distribution of the differences found in certain layers of the neuroretina in both cohorts. Table A.1 (supplementary appendix) shows the numerical values of the thicknesses in the foveal, parafoveal and perifoveal rings and for all three rings overall (GLOBAL). The discriminant capacity in each grouping is evaluated using the AUROC.

Analysis of the grouping of the three rings reveals that the most pronounced thinning occurs in the inner layers of the patients' retina: in the GCL (composed of the ganglion cell somas or nuclei) and the IPL, both of which have an AUROC value of 0.82. The second-most affected structure is the IRL (AUROC = 0.78). There is no significant difference between the ROC curves of these three variables. The thinning of these inner layers of the retina confirms that neuroaxonal degeneration associated with development of the disease occurs, even when patients do not present any signs or symptoms of ophthalmological disease or optic neuritis.

Fig. 4 shows in heat map form the inter-eye difference for control subjects and patients in the 64 cells of each of the nine retinal structures. The image shows that, in general, inter-eye difference is greater in patients than in control subjects, especially in the GCL and the IPL. The numerical results are presented in Table A.2 (supplementary appendix).

It is still the inner layers of the retina, specifically the GCL (AUROC = 0.75), IRL (AUROC = 0.74) and IPL (AUROC = 0.71), which provide the greatest discriminant capacity when evaluating the inter-eye difference between the healthy and patient cohorts. No tomographic area within these three layers presents significantly greater alteration due to the disease than any other, and discriminant capacity is fairly similar in the foveal, parafoveal, and perifoveal zones (Fig. 3). Overall affectation of the ganglion cells and of the thickness of their axons therefore appears to occur equally in the different rings into which the neuroretinal analysis is divided.

3.3. CNN input selection

To determine which are the most appropriate inputs to the CNN, the AUROC value of the GLOBAL zone, which spans the foveal, parafoveal and perifoveal rings, is taken as reference. The variables that meet the condition AUROC > 0.70, ranked from highest to lowest discriminant capacity, are as follows: GCL (AUROC = 0.82, CI: 0.89–0.75), IPL (AUROC = 0.82, CI: 0.89–0.75), IRL (AUROC = 0.78, CI = 0.85–0.70) in measurements of the thicknesses of both eyes and GCL (AUROC = 0.75, CI = 0.83–0.67), IRL (AUROC = 0.74, CI = 0.82–0.66) and IPL (AUROC = 0.71, CI = 0.79–0.63) in measurements of inter-eye difference.

Table A.3 (supplementary appendix) analyses the statistical

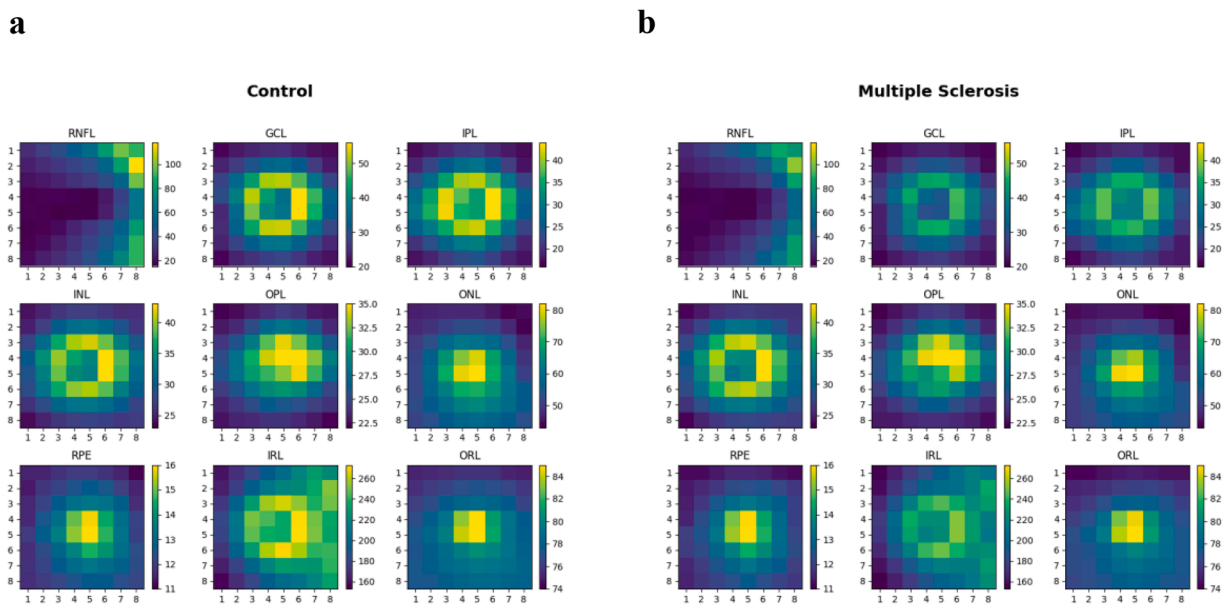


Fig. 2. Heat map visualization of median average thickness for both eyes in μm. The figure shows the average thickness value in each of the 64 cells measured using the exploratory protocol in healthy controls subjects (a) and in MS patients (b).

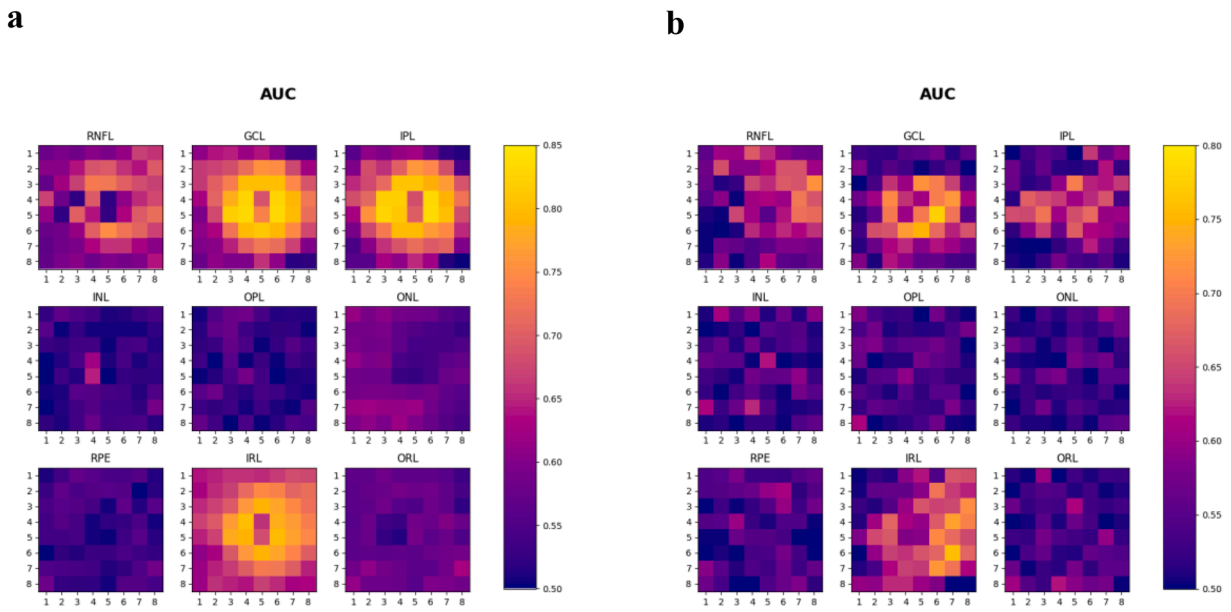


Fig. 3. Heat map visualization of the AUROC (area under ROC curve) values between HC and MS patients. **a** AUROC of median average thickness for both eyes. **b** AUROC of inter-eye difference.

difference between these AUROC values using the DeLong test.

The input to the neural classifier will comprise the information with the highest AUROC value: GCL (thickness). Although there is no statistical difference with the next four datasets ($P > 0.05$), there is a difference with the IPL value (inter-eye difference). The classifier inputs will therefore be GCL (thickness) and IPL (inter-eye difference).

3.4. Automatic classification

Fig. 5a shows the architecture of the CNN. The $8 \times 8 \times 82$ input is made up of images of the two retinal structures with greatest discriminant capacity. The best classification results were obtained with a CNN comprising two convolution layers. The first convolution layer uses 16 kernels (dimensions: 3×3; stride $= 1$, padding $= 1$) producing 16 feature maps (dimensions: 8×8). The second convolution layer implements 32 kernels (dimensions: 3×3; stride $= 2$, padding $= 1$) generating 32 feature maps (dimensions: 4×4). Batch normalization and a non-linear activation function are implemented in both layers after the convolution function. The output information of the second convolution layer ($4 \times 4 \times 32$) is applied to a fully connected network with 128 inputs and two outputs indicating the probability of belonging to either the control subject or patient cohort, respectively.

The number of trainable parameters is 71,506 and the error is defined as the CrossEntropyLoss. The training parameters for the Adam optimizer are learning rate $\eta = 1e\text{-}5$, momentum rate $\gamma = 0.9$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1e{-}8$, and the training epochs were set to 200.

The performance of the classification is represented by a confusion matrix obtained using leave-one-out cross validation: 69+79 CNN

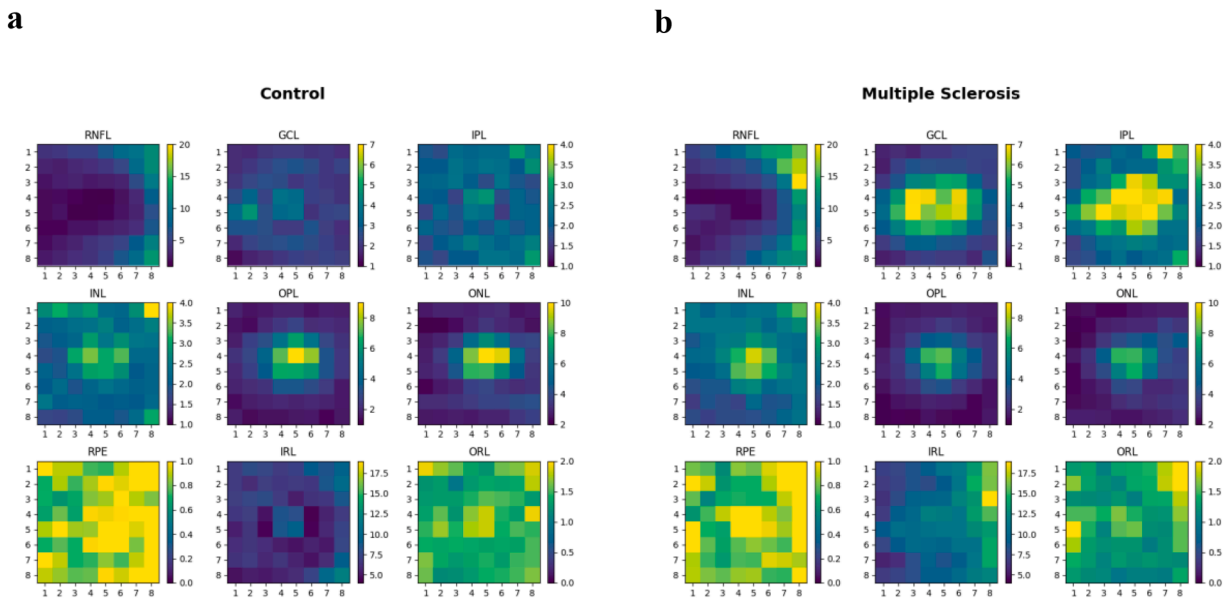


Fig. 4. Heat map visualization of the inter-eye difference in μm. It shows the mean value in each of the 64 cells measured using the exploratory protocol in **a** healthy controls subjects and **b** in MS patients.

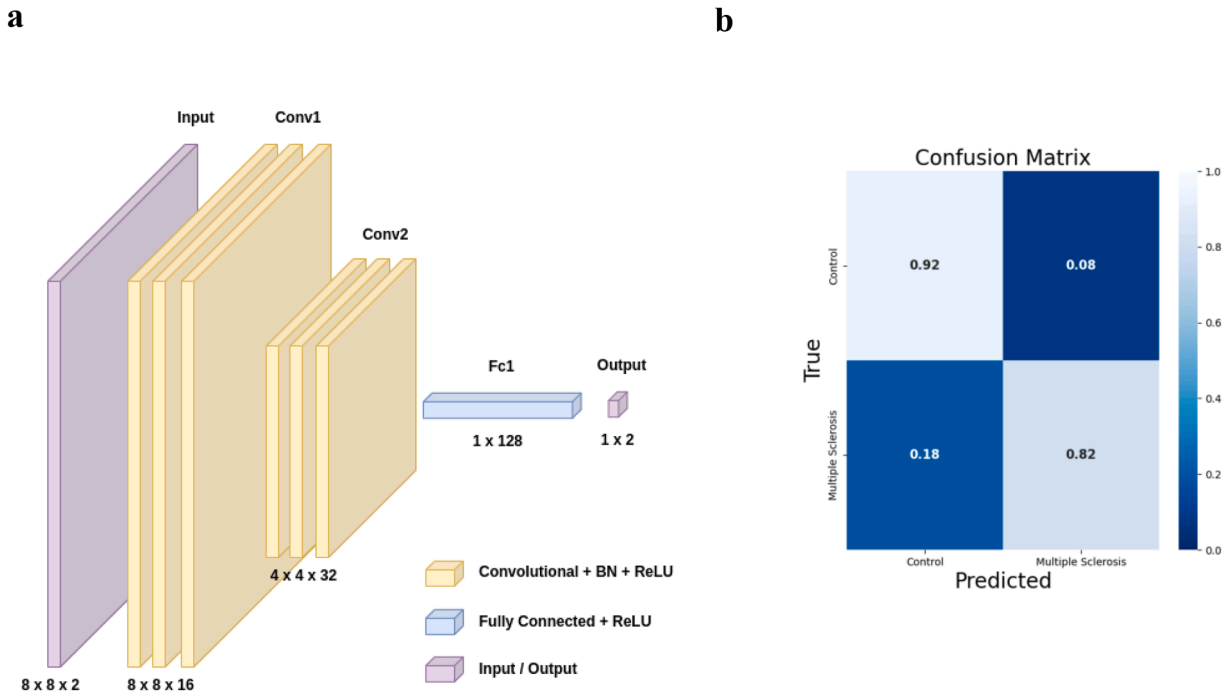


Fig. 5. Classifier architecture and confusion matrix obtained in the leave-one-out cross validation process. **a** Convolutional neural network (CNN) architecture. **b** Confusion matrix.

training processes were executed from scratch (random initialization of the weights of the fully connected layer and in the coefficients of the kernels). Under these conditions, the assisted diagnostic system implemented attains accuracy = 0.87, sensitivity = 0.82, and specificity = 0.92 (Fig. 5b).

4. Discussion

In this work we have used state-of-the-art SD-OCT technology that allows us to obtain highly accurate structural measurements thanks to an automatic positioning system (APS) and the separate segmentation of the GCL and IPL layers. Under these conditions, we have obtained results relating to 1) analysis of neuroretinal thinning in patients recently diagnosed with RRMS versus age-matched control subjects; 2) variation in inter-eye thickness in both cohorts (healthy controls and patients); 3) analysis of the discriminant capacity of thickness and inter-eye difference; and 4) the creation and application of a CNN to assess the ability to distinguish between patients and healthy controls using the retinal structures with the greatest discriminant capacity (thickness and inter-eye difference).

An important innovation of this study is that the Posterior Pole protocol has been used to improve measurement reliability with the

incorporation of the APS to locate the centres of the fovea and the Bruch membrane opening, respectively, as fixed structural landmarks, thus ensuring that the tomographic analysis is performed on the macula and not on another nearby area of the retina. In addition to being more accurate as regards following up and monitoring changes in patients over time, this protocol constitutes a significant advance in our study because it reduces inter-eye variability not attributable to the pathology itself. The APS ensures that the right and left eye measurements match exactly. This is achieved by using the same criteria to align the centre of the fovea in both eyes and, therefore, ensuring the measurements are precisely mirrored in both eyes. This provides a significant advantage when evaluating inter-eye thickness differences by ensuring that the differences found are due either to the pathology or to possible episodes of subclinical optic neuritis. Without the APS, differences could be due to different degrees of vergence between a patient's eyes, to the position of the patient's head in the scanner, or even to minor anatomical differences such as refractive differences, peripapillary atrophy, etc.

As regards the first analysis, we found that the inner retinal layers (GCL, IPL, and IRL), which contain the ganglion cell somas and axons, showed the highest affectation and so possessed the greatest power to discriminate between control subjects and patients. The AUROC values obtained in our study are consistent with those reported in previous papers (including those employing other OCT technologies and protocols) and suggest that the neuroaxonal loss that occurs in RRMS causes a reduction in ganglion cells, an outcome that previous authors have observed correlates with the functional disability associated with this disease (Garcia-Martin et al., 2017; Petzold et al., 2017; Montolío et al., 2021, 2022).

One of the current issues in clinical care and clinical research in MS is the concept of disease progression without evidence of clinical or neuroradiological relapse. The patients in our study never experienced a history of optic neuritis, so our results demonstrate the recorded presence of neurodegeneration in the absence of inflammatory attacks.

Our analyses also show that affectation of the ganglion cells and of the thickness of their axons occurs in all the different areas of the retina and is not greater in the foveal area than elsewhere. Detailed study of the papillomacular bundle may find some differences, since this bundle contains the greatest number of fibres originating from the ganglion cell somas that converge to form the optic nerve.

In relation to the analysis of the difference in inter-eye thickness in each subject's two eyes, we observed that once more the inner retinal layers (GCL, IRL, and IPL) exhibited the greatest capacity to discriminate between patients and control subjects. According to their medical histories, our patients had not previously suffered episodes of optic neuritis, although it is possible that they may have suffered subclinical episodes or may simply exhibit different axonal affectation between the two eyes. This finding suggests that axonal affectation at neuroretina level in patients is an asymmetrical process that affects each eye of each subject differently even when there is no clinical history of affectation of visual acuity in one of the eyes in particular. Our findings are consistent with those reported by (Petzold et al., 2021b) where OCT images of 71,939 control subjects and 144 patients without optic neuritis were analysed, finding that the highest AUROC value was achieved for inter-eye difference in the mGCIPL (AUC = 0.71, CI: 0.6646–0.7575). Our study has improved on these results, achieving AUROC values of 0.75 (CI: 0.83–0.67) in the GCL and of 0.74 (CI: 0.82–0.66) in the IRL. This may be because our analysis uses the APS, which increases the reproducibility of the OCT measurements, and also because our study separates the GCL and IPL while Petzold et al. analysed both layers (mGCIPL) together, which may produce greater variability because it includes more cells in the measurements. Likewise, and in line with the conclusions of (Petzold et al., 2021b), in our study we observed that inter-eye difference in the RNFL possesses low discriminatory power (AUROC = 0.68).

In another study, (Nij Bijvank et al., 2022) found that inter-eye difference in the mGCIPL proved highly accurate in diagnosing patients with multiple sclerosis suffering both unilateral and bilateral optic neuritis over time. The diagnostic accuracy of using inter-eye difference to differentiate between patients with multiple sclerosis without optic neuritis and control subjects was moderate to good (AUC: 0.66–0.73, disease duration: 20.3 years; this study includes different types of patients with multiple sclerosis), attaining better results and at earlier stages of relapsing-remitting multiple sclerosis. Bijvank et al. come to the conclusion, as we did, that some of the patients had subclinical optic neuritis or asymmetrical pathologies further along the visual pathway causing trans-synaptic retrograde degeneration of the visual pathway.

In our study, which drew on patients who had not been diagnosed with optic neuritis, we found that the average thickness of the inner retinal layers of both eyes had greater discriminant capacity than the inter-eye difference measurement in the same subject. According to the results presented in Table A.3 (supplementary appendix), however, the only significant difference exists between the GCL (thickness) and the IPL (inter-eye difference). To the best of our knowledge, this is the first time this analysis has been conducted.

As regards the fourth and final analysis in this study, recent papers have explored automatic diagnosis of MS in early and incipient stages of the disease using ML techniques. Analysing data obtained with a swept-source OCT device (45 × 60 matrix) with an effect size metric (Cohen's distance) to detect the regions most altered by the disease and a feed-forward neural network as a classifier, accuracy = 0.98 was attained (disease duration: 7.35 ± 1.95 months, EDSS score = 1.07) (Garcia--Martin et al., 2021). With the same database but using convolutional neural networks to implement data augmentation using generative adversarial networks and a classifier, a perfect classification was obtained (accuracy = 1.0) (López-Dorado et al., 2021). In a paediatric population (disease duration: 0.6 years) it yielded accuracy = 0.80 when classifying multiple sclerosis patients versus controls using a Random Forest classifier (Ciftci Kavaklioglu et al., 2022). Since the results reported to date depend on factors such as the patient cohort characteristics, the technology used to take the readings (spectral domain OCT versus swept-source OCT), the exploratory protocol employed, the data analysis and classification method, as well as consideration of the subjects' other biological variables, etc., further studies are needed to determine the most appropriate procedure.

The best results in the classification were obtained using combined thickness and inter-eye difference data values as inputs to the CNN. Importantly, the sensitivity and specificity of OCT when diagnosing MS using the new CNN-based analysis methodology come close to that of the current diagnostic gold standard, MRI, and may represent an extra diagnostic tool. Some authors propose the incorporation of OCT analysis in the diagnostic criteria for multiple sclerosis (Kenney et al., 2022; Petzold et al., 2017).

Diagnosing neurodegenerative diseases rapidly and in their early stages is one of the main challenges facing neurodegenerative disease clinicians worldwide and, in the case of MS, OCT may help achieve this goal. Earlier disease detection allows doctors to provide earlier treatment to slow down the progression of the disease, thus achieving an increase in Quality-Adjusted Life Years (QALY), namely a higher number of years with a higher quality of life for these patients.

Author contributions

MO, LB, EMSM, EGM contributed to the conception and design of the study.

VM, EV, BC contributed to the acquisition and analysis of data.

LB, AMSM, FJD, JMMJ, EGM contributed to analysis and interpretation of results or preparing the figures.

MO, LB, EGM contributed to drafting the text.

All authors reviewed the results and approved the final version of the manuscript.

Funding

This study was supported by Carlos III Health Institute grants PI17/01726 and PI20/00437, by the Inflammatory Disease Network (RICORS) (RD21/0002/0050) (Carlos III Health Institute), and by project reference UAH-GP2022–2 funded by the University of Alcalá Proprietary Research Programme.

The funding organizations had no role in the design or conduct of this research.

Institutional review board statement

The design of this study adhered to the tenets of the Declaration of Helsinki and all participants provided written informed consent. The study protocol was approved by the Clinical Research Ethics Committee of Aragon (Zaragoza, Spain).

Data availability statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

CRediT authorship contribution statement

Miguel Ortiz: Conceptualization, Methodology, Software, Writing – review & editing. **Victor Mallen:** Data curation, Visualization, Writing – review & editing. **Luciano Boquete:** Conceptualization, Methodology, Software, Writing – original draft, Supervision. **Eva M. Sánchez-Morla:** Visualization, Investigation, Software, Writing – review & editing. **Beatriz Cordón:** Visualization, Investigation, Software, Writing – review & editing. **Elisa Vilades:** Visualization, Investigation, Software, Writing – review & editing. **Francisco J. Dongil-Moreno:** Visualization, Investigation, Software, Writing – review & editing. **Juan M. Miguel-Jiménez:** Visualization, Investigation, Software, Writing – review & editing. **Elena Garcia-Martin:** Conceptualization, Methodology, Software, Writing – original draft, Supervision.

Declaration of Competing Interest

The authors report no competing interests.

Acknowledgments

We thank Dr Luis E Pablo for help cooperation in the use of devices.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2023.104725.

References

Alonso, R., Gonzalez-Moron, D., Garcea, O., 2018. Optical coherence tomography as a biomarker of neurodegeneration in multiple sclerosis: a review. Mult. Scler. Relat. Disord. 22, 77–82. https://doi.org/10.1016/j.msard.2018.03.007.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 53. https://doi.org/10.1186/s40537-021-00444-8.

Asrani, S., Rosdahl, J.A., Allingham, R.R., 2011. Novel software strategy for glaucoma diagnosis: asymmetry analysis of retinal thickness. Arch. Ophthalmol. 129, 1205–1211. https://doi.org/10.1001/archophthalmol.2011.242.

Balasubramanian, M., Bowd, C., Vizzeri, G., Weinreb, R.N., Zangwill, L.M., 2009. Effect of image quality on tissue thickness measurements obtained with spectral domain-optical coherence tomography. Opt. Express 17, 4019. https://doi.org/10.1364/OE.17.004019.

Bsteh, G., Hegen, H., Altmann, P., Berek, K., Auer, M., Zinganell, A., di Pauli, F., Rommer, P., Leutmezer, F., Deisenhammer, F., Berger, T., 2020. Inner nuclear layer and olfactory threshold are interlinked and reflect inflammatory activity in multiple sclerosis. Mult. Scler. J. Exp. Transl. Clin. 6, 205521732094573 https://doi.org/10.1177/2055217320945738.

Cavaliere, C., Vilades, E., Alonso-Rodríguez, M.C., Rodrigo, M.J., Pablo, L.E., Miguel, J. M., López-Guillén, E., Morla, E.M.S., Boquete, L., Garcia-Martin, E., 2019. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. Sensors (Basel) 19. https://doi.org/10.3390/s19235323.

Chylack, L.T., 1993. The lens opacities classification system III. Arch. Ophthalmol. 111, 831. https://doi.org/10.1001/archopht.1993.01090060119035.

Ciftci Kavaklioglu, B., Erdman, L., Goldenberg, A., Kavaklioglu, C., Alexander, C., Oppermann, H.M., Patel, A., Hossain, S., Berenbaum, T., Yau, O., Yea, C., Ly, M., Costello, F., Mah, J.K., Reginald, A., Banwell, B., Longoni, G., Ann Yeh, E., 2022. Machine learning classification of multiple sclerosis in children using optical coherence tomography. Multiple Sclerosis J. 28, 2253–2262. https://doi.org/10.1177/13524585221112605.

Garcia-Martin, E., Ara, J.R., Martin, J., Almarcegui, C., Dolz, I., Vilades, E., Gil-Arribas, L., Fernandez, F.J., Polo, V., Larrosa, J.M., Pablo, L.E., Satue, M., 2017. Retinal and optic nerve degeneration in patients with multiple sclerosis followed up for 5 years. Ophthalmology 124, 688–696. https://doi.org/10.1016/j.ophtha.2017.01.005.

Garcia-Martin, E., Ortiz, M., Boquete, L., Sánchez-Morla, E.M., Barea, R., Cavaliere, C., Vilades, E., Orduna, E., Rodrigo, M.J., 2021. Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier. Comput. Biol. Med. 129, 104165 https://doi.org/10.1016/j.compbiomed.2020.104165.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Kenney, R.C., Liu, M., Hasanaj, L., Joseph, B., Abu Al-Hassan, A., Balk, L.J., Behbehani, R., Brandt, A., Calabresi, P.A., Frohman, E., Frohman, T.C., Havla, J., Hemmer, B., Jiang, H., Knier, B., Korn, T., Leocani, L., Martinez-Lapiscina, E.H., Papadopoulou, A., Paul, F., Petzold, A., Pisa, M., Villoslada, P., Zimmermann, H., Thorpe, L.E., Ishikawa, H., Schuman, J.S., Wollstein, G., Chen, Y., Saidha, S., Galetta, S., Balcer, L.J., 2022. The role of optical coherence tomography criteria and machine learning in multiple sclerosis and optic neuritis diagnosis. Neurology 99, e1100–e1112. https://doi.org/10.1212/WNL.0000000000200883.

López-Dorado, A., Ortiz, M., Satue, M., Rodrigo, M.J., Barea, R., Sánchez-Morla, E.M., Cavaliere, C., Rodríguez-Ascariz, J.M., Orduna-Hospital, E., Boquete, L., Garcia-Martin, E., 2021. Early diagnosis of multiple sclerosis using swept-source optical coherence tomography and convolutional neural networks trained with data augmentation. Sensors 22, 167. https://doi.org/10.3390/s22010167.

Montolío, A., Cegoñino, J., Garcia-Martin, E., Pérez del Palomar, A., 2022. Comparison of machine learning methods using spectralis OCT for diagnosis and disability progression prognosis in multiple sclerosis. Ann. Biomed. Eng. 50, 507–528. https://doi.org/10.1007/s10439-022-02930-3.

Montolío, A., Martín-Gallego, A., Cegoñino, J., Orduna, E., Vilades, E., Garcia-Martin, E., del Palomar, A.P., 2021. Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. Comput. Biol. Med. 133, 104416 https://doi.org/10.1016/j.compbiomed.2021.104416.

Nij Bijvank, J., Uitdehaag, B.M.J., Petzold, A., 2022. Retinal inter-eye difference and atrophy progression in multiple sclerosis diagnostics. J. Neurol. Neurosurg. Psychiatry 93, 216–219. https://doi.org/10.1136/jnnp-2021-327468.

Nolan, R.C., Galetta, S.L., Frohman, T.C., Frohman, E.M., Calabresi, P.A., Castrillo-Viguera, C., Cadavid, D., Balcer, L.J., 2018. Optimal intereye difference thresholds in retinal nerve fiber layer thickness for predicting a unilateral optic nerve lesion in multiple sclerosis. J. Neuro Ophthalmol. 38, 451–458. https://doi.org/10.1097/WNO.0000000000000629.

Nolan-Kenney, R.C., Liu, M., Akhand, O., Calabresi, P.A., Paul, F., Petzold, A., Balk, L., Brandt, A.U., Martínez-Lapiscina, E.H., Saidha, S., Villoslada, P., Al-Hassan, A.A., Behbehani, R., Frohman, E.M., Frohman, T., Havla, J., Hemmer, B., Jiang, H., Knier, B., Korn, T., Leocani, L., Papadopoulou, A., Pisa, M., Zimmermann, H., Galetta, S.L., Balcer, L.J., 2019. Optimal intereye difference thresholds by optical coherence tomography in multiple sclerosis: an international study. Ann. Neurol. 85, 618–629. https://doi.org/10.1002/ana.25462.

Petzold, A., Albrecht, P., Balcer, L., Bekkers, E., Brandt, A.U., Calabresi, P.A., Deborah, O. G., Graves, J.S., Green, A., Keane, P.A., Nij Bijvank, J.A., Sander, J.W., Paul, F., Saidha, S., Villoslada, P., Wagner, S.K., Yeh, E.A., Aktas, O., Antel, J., Asgari, N., Audo, I., Avasarala, J., Avril, D., Bagnato, F.R., Banwell, B., Bar-Or, A., Behbehani, R., Manterola, A.B., Bennett, J., Benson, L., Bernard, J., Bremond-Gignac, D., Britze, J., Burton, J., Calkwood, J., Carroll, W., Chandratheva, A., Cohen, J., Comi, G., Cordano, C., Costa, S., Costello, F., Courtney, A., Cruz-Herranz, A., Cutter, G., Crabb, D., Delott, L., de Seze, J., Diem, R., Dollfuss, H., el Ayoubi, N.K., Fasser, C., Finke, C., Fischer, D., Fitzgerald, K., Fonseca, P., Frederiksen, J.L., Frohman, E., Frohman, T., Fujihara, K., Cuellar, I.G., Galetta, S., Garcia-Martin, E., Giovannoni, G., Glebauskiene, B., Suárez, I.G., Jensen, Gorm Pihl, Hamann, S., Hartung, H., Havla, J., Hemmer, B., Huang, S., Imitola, J., Jasinskas, V., Jiang, H., Kafieh, R., Kappos, L., Kardon, R., Keegan, D., Kildebeck, E., Kim, U.S., Klistorner, S., Knier, B., Kolbe, S., Korn, T., Krupp, L., Lagrèze, W., Leocani, L., Levin, N., Liskova, P., Preiningerova, J.L., Lorenz, B., May, E., Miller, D., Mikolajczak, J., Saïd, S.M., Montalban, X., Morrow, M., Mowry, E., Murta, J., Navas, C., Nolan, R., Nowomiejska, K., Oertel, F.C., Oh, J., Oreja-Guevara, C., Orssaud, C., Osborne, B., Outteryck, O., Paiva, C., Palace, J., Papadopoulou, A., Patsopoulos, N., Preiningerova, J.L., Pontikos, N., Preising, M., Prince, J., Reich, D., Rejdak, R., Ringelstein, M., Rodriguez de Antonio, L., Sahel, J., Sanchez-Dalmau, B., Sastre-Garriga, J., Schippling, S., Schuman, J., Shindler, K., Shin, R., Shuey, N., Soelberg, K., Specovius, S., Suppiej, A., Thompson, A., Toosy, A., Torres, R., Touitou, V., Trauzettel-Klosinski, S., van der Walt, A., Vermersch, P., Vidal-

Jordana, A., Waldman, A.T., Waters, C., Wheeler, R., White, O., Wilhelm, H., Winges, K.M., Wiegerinck, N., Wiehe, L., Wisnewski, T., Wong, S., Würfel, J., Yaghi, S., You, Y., Yu, Z., Yu-Wai-Man, P., Žemaitien≐, R., Zimmermann, H., 2021a. Artificial intelligence extension of the OSCAR-IB criteria. Ann. Clin. Transl. Neurol. 8, 1528–1542. https://doi.org/10.1002/acn3.51320.

Petzold, A., Balcer, L.J., Calabresi, P.A., Costello, F., Frohman, T.C., Frohman, E.M., Martinez-Lapiscina, E.H., Green, A.J., Kardon, R., Outteryck, O., Paul, F., Schippling, S., Vermersch, P., Villoslada, P., Balk, L.J., ERN-EYE IMSVISUAL, 2017. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. Lancet Neurol. 16, 797–812. https://doi.org/10.1016/S1474-4422(17)30278-8.

Petzold, A., Chua, S.Y.L., Khawaja, A.P., Keane, P.A., Khaw, P.T., Reisman, C., Dhillon, Baljean, Strouthidis, N.G., Foster, P.J., Patel, P.J., Atan, D., Aslam, T., Barman, S.A., Barrett, J.H., Bishop, P., Bunce, C., Carare, R.O., Chakravarthy, U., Chan, M., Chua, S.Y.L., Crabb, D.P., Day, A., Desai, P., Dhillon, Bal, Dick, A.D., Egan, C., Ennis, S., Ennis, S., Foster, P.J., Fruttiger, M., Gallacher, J.E.J., Garway-Heath, D.F., Gibson, J., Gore, D., Guggenheim, J.A., Hammond, C.J., Hardcastle, A., Harding, S.P., Hogg, R.E., Hysi, P., Keane, P.A., Khaw, S.P.T., Khawaja, A.P., Lascaratos, G., Lotery, A.J., Macgillivray, T., Mackie, S., McGaughey, M., McGuinness, B., McKay, G.J., McKibbin, M., Moore, T., Morgan, J.E., Muthy, Z.A., O'Sullivan, E., Owen, C.G., Patel, P., Paterson, E., Peto, T., Petzold, A., Rahi, J.S., Rudnikca, A.R., Self, J., Sivaprasad, S., Steel, D., Stratton, I., Strouthidis, N., Sudlow, C., Thomas, D., Trucco, E., Tufail, A., Vitart, V., Vernon, S.A., Viswanathan, A.C., Williams, C., Williams, K., Woodside, J.v, Yates, M.M., Zheng, Y., 2021b. Retinal asymmetry in multiple sclerosis. Brain 144, 224–235. https://doi.org/10.1093/brain/awaa361.

Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., Fujihara, K., Galetta, S.L., Hartung, H.P., Kappos, L., Lublin, F.D., Marrie, R.A., Miller, A.E., Miller, D.H., Montalban, X., Mowry, E.M., Sorensen, P.S., Tintoré, M., Traboulsee, A.L., Trojano, M., Uitdehaag, B.M.J., Vukusic, S., Waubant, E., Weinshenker, B.G., Reingold, S.C., Cohen, J.A., 2018. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 17, 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2.

Yasin Alibhai, A., Or, C., Witkin, A.J., 2018. Swept source optical coherence tomography: a review. Curr. Ophthalmol. Rep. 6, 7–16. https://doi.org/10.1007/s40135-018-0158-3.