





ORIGINAL ARTICLE

The macular retinal ganglion cell layer as a biomarker for diagnosis and prognosis in multiple sclerosis: A deep learning approach

Alberto Montolio^{1,2}  | José Cegoñino^{1,2}  | Elena Garcia-Martin^{3,4}  |
Amaya Pérez del Palomar^{1,2} 

¹Biomaterials Group, Aragon Institute of Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain

²Mechanical Engineering Department, University of Zaragoza, Zaragoza, Spain

³Ophthalmology Department, Miguel Servet University Hospital, Zaragoza, Spain

⁴GIMSO Research and Innovation Group, Aragon Institute for Health Research (IIS Aragon), Zaragoza, Spain

Correspondence

Alberto Montolio, Escuela de Ingeniería y Arquitectura, Campus Río Ebro, Edificio Betancourt, C/María de Luna s/n, 50018 Zaragoza, Spain.
Email: amontolio@unizar.es

Funding information

Instituto de Salud Carlos III, Grant/Award Number: PI17/01726; Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: BES-2017-080384; Ministerio de Economía y Competitividad, Grant/Award Number: DPI 2016-79302-R

Abstract

Purpose: The macular ganglion cell layer (mGCL) is a strong potential biomarker of axonal degeneration in multiple sclerosis (MS). For this reason, this study aims to develop a computer-aided method to facilitate diagnosis and prognosis in MS.

Methods: This paper combines a cross-sectional study of 72 MS patients and 30 healthy control subjects for diagnosis and a 10-year longitudinal study of the same MS patients for the prediction of disability progression, during which the mGCL was measured using optical coherence tomography (OCT). Deep neural networks were used as an automatic classifier.

Results: For MS diagnosis, greatest accuracy (90.3%) was achieved using 17 features as inputs. The neural network architecture comprised the input layer, two hidden layers and the output layer with softmax activation. For the prediction of disability progression 8 years later, accuracy of 81.9% was achieved with a neural network comprising two hidden layers and 400 epochs.

Conclusion: We present evidence that by applying deep learning techniques to clinical and mGCL thickness data it is possible to identify MS and predict the course of the disease. This approach potentially constitutes a non-invasive, low-cost, easy-to-implement and effective method.

KEY WORDS

multiple sclerosis, deep learning, optical coherence tomography, retinal ganglion cell layer

1 | INTRODUCTION

Deep learning techniques are increasingly being used in clinical applications (Ker et al., 2018). Most of these applications focus on diagnostic imaging, such as the use of magnetic resonance imaging (MRI) to diagnose multiple sclerosis (MS) (Marzullo et al., 2019; McKinley et al., 2020; Salem et al., 2019). Given the important role of MRI in the diagnosis and treatment of MS, several studies have explored the application of artificial intelligence (AI) to this disease (Bonacchi et al., 2022).

Multiple sclerosis is a chronic neurodegenerative disease that affects the central nervous system (CNS). It is characterized by high clinical variability, as each case is unique (Swanton et al., 2014). MS is diagnosed using the McDonald criteria (Thompson et al., 2018), which entail a lengthy process involving several invasive tests, such

as MRI and lumbar puncture to analyse cerebrospinal fluid (CSF). Moreover, as the number of treatment options continues to increase, selecting the best therapy for each individual at the appropriate stage of the disease is a challenge. The development of reliable biomarkers that can be measured non-invasively is therefore essential.

Multiple sclerosis has been shown to affect the retina, which is an extension of the CNS, particularly affecting the retinal nerve fibre layer (RNFL) and the ganglion cell layer (GCL) (You et al., 2020). The scar tissue formed in this disease damages the nerve fibres of the retinal ganglion cells, interrupting the passage of nerve impulses. As a result, the cells stop transmitting information and cell death occurs (Eslami et al., 2020; Pietroboni et al., 2019). Non-invasive analysis of these retinal layers using optical coherence tomography (OCT) has demonstrated its usefulness as an MS biomarker (Alonso et al., 2018;

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Acta Ophthalmologica* published by John Wiley & Sons Ltd on behalf of Acta Ophthalmologica Scandinavica Foundation.

Frohman et al., 2006), particularly since OCT is an easy-to-use, cost-effective and objective technology that provides high-resolution images.

Although the RNFL has been widely used to identify MS (Montolío et al., 2019; Pueyo et al., 2008; Toledo et al., 2008), the macular GCL (mGCL) also indicates neurodegeneration in MS as it is formed by the cell bodies of the RNFL axons (Britze et al., 2017). The mGCL and the inner plexiform layer (IPL) are usually measured together (referred to as the mGCIPL) because of the low contrast that hinders their segmentation (Petzold et al., 2017). It has been shown that mGCIPL thinning in MS patients is significantly greater than in healthy control subjects, associating faster rates of atrophy of these layers with disability progression (Graham et al., 2016; Saidha et al., 2015). Shi et al. (2019) found a horseshoe-like reduction in mGCIPL thickness at the nasal sector and revealed its relationship to visual dysfunction and disability in MS patients. Lambe et al. (2021) demonstrated that a basal GCIPL thickness $<70\mu\text{m}$ was associated with a four-fold increase in the likelihood of significant worsening of disability. In this sense, other authors have associated a basal mGCIPL thickness $<77\mu\text{m}$ with a higher progression of disability. Moreover, annual mGCIPL thickness loss $\geq 1\mu\text{m}$ served to identify patients with disability progression (Bsteh et al., 2020). Schurz et al. (2021) also established that mGCIPL thinning $\geq 1\mu\text{m}/\text{year}$ represents an increased risk of worsening disability in MS patients.

As detailed above, MRI is the test most commonly combined with AI in MS diagnosis and prognosis (Rocca et al., 2021; Yoo et al., 2019; Zhang et al., 2019). However, researchers have also combined AI with imaging techniques such as OCT (Betzler et al., 2022). Cavaliere et al. (2019) developed a diagnostic method based on the support vector machine (SVM) and RNFL and GCIPL measurements taken with the deep range imaging (DRI) OCT Triton device. These data were also used by Garcia-Martin et al. (2021) when comparing the SVM and artificial neural networks (ANNs). Pérez del Palomar et al. (2019) analysed the ability of machine learning techniques to improve the detection of RNFL and GCIPL damage in MS patients. Another MS diagnosis method was proposed by López-Dorado et al. (2021), who used a convolutional neural network (CNN) to classify OCT images from 48 recently diagnosed MS patients and 48 control subjects. The results of the above-mentioned studies did not provide a clear-cut recommendation on whether it is better to use the RNFL or the GCIPL to diagnose the disease. Thus, in this new analysis we wanted to conduct an additional study to corroborate whether the GCL is as good a biomarker as the RNFL. In addition, this study assessed GCL thickness independently of IPL, in contrast to previous studies.

Regarding MS prognosis using AI, most studies predicted disability progression based on data from the evaluations included in the McDonald criteria, such as MRI, CSF analysis and evoked potentials (EPs) (Seccia et al., 2021). Thus, very few studies based their prediction of disease progression on OCT data. Our previous studies proposed several machine learning techniques to predict long-term disability progression using RNFL

thickness measured by Cirrus HD-OCT and Spectralis OCT (Montolío et al., 2021, 2022). In addition, it is necessary to evaluate the potential offered by GCL thickness for the same purpose. Therefore, in this study ANNs were used to diagnose and predict progression in MS patients based on GCL thickness. The use of the non-invasive OCT test to measure thickness as an MS biomarker would reduce the need for many invasive tests and would facilitate early diagnosis. This would allow earlier initiation of treatment for patients, thus improving their quality of life (Yap et al., 2019).

2 | MATERIALS AND METHODS

2.1 | Participants

This study enrolled 102 participants (72 MS patients and 30 healthy subjects) of white European origin. The study was approved by the Ethics Committee of the Miguel Servet Hospital and adhered to the Declaration of Helsinki. All participants provided written informed consent before enrolment.

This paper combines a cross-sectional study of MS patients and healthy control subjects for diagnosis and a 10-year longitudinal study of 72 MS patients for prediction of disability progression. MS patients were diagnosed by a neurologist based on the 2010 revision of the McDonald criteria (Polman et al., 2011) and patient disability was registered using the Expanded Disability Status Scale (EDSS). One eye was randomly selected to avoid possible bias due to interrelationship between eyes of the same subject. In subjects with exclusion criteria in one eye, the other eye was selected.

Neuro-ophthalmological evaluation and OCT scans were performed at the baseline visit on both MS patients and healthy participants. These were followed by five annual visits plus a final visit for MS patients to complete the 10-year follow-up. The baseline visit constitutes the first evaluation and is the starting point of the study in which each patient is at a different stage of the disease.

Based on our preliminary studies related to MS (Cordon et al., 2020; López-Dorado et al., 2021), the sample size necessary to detect differences of at least $2\mu\text{m}$ in OCT-measured thicknesses was determined by applying a bilateral test with α 5% risk and β 10% risk (i.e. with 90% power). In order to obtain a sufficient sample of MS patients to allow an in-depth study of the natural course of the disease, the unexposed/exposed ratio was determined to be 0.5. From these data, it was concluded that at least 28 eyes would be needed in each group.

2.2 | Inclusion/exclusion criteria

The inclusion criteria applied were as follows: best-corrected visual acuity (BCVA) of 20/40 or higher, refractive error within ± 5.00 dioptres (D) equivalent sphere and ± 2.00 D astigmatism and transparent ocular media (nuclear colour/opalescence, cortical or posterior subcapsular lens opacity < 1), according to the Lens Opacities Classification System III (Chylack

et al., 1993). Exclusion criteria included previous retinal and optic nerve affection, other diseases affecting the visual field or nervous system, refractive errors >5 D of equivalent spherical dioptres or >2 D of astigmatism and intraocular pressure >20 mmHg, as well as exclusion of those eyes with a family history of glaucoma. A previous history of optic neuritis was not an exclusion criterion (in order to analyse its influence on MS progression).

Multiple sclerosis patients were recruited between 2006 and 2010 and followed for 10 years by MS neurologists who referred all patients with EDSS ≤ 8 and disease duration ≤ 35 years at the baseline visit from the neurology department of the Miguel Servet Hospital. All consecutive patients meeting all inclusion criteria and not meeting any exclusion criteria were included in the study, and no retrospective patient selection was performed. There were initially 86 MS patients, but 14 dropped out during follow-up and only those who completed the study were included in the analysis. The reasons for drop-out were change of residence, patient decision or poor mobility (no dropouts were due to death).

2.3 | Optical coherence tomography

OCT scans were performed on all participants using the Spectralis OCT device (Heidelberg Engineering, Inc.). The eye-tracking function was enabled to achieve greatest accuracy and to use the baseline scan to align subsequent scans with the same area of the retina. Scans were evaluated following the quality control criteria (OSCAR-IB) and the Advised Protocol for OCT Study Terminology and Elements (APOSTEL) (Cruz-Herranz et al., 2016; Petzold et al., 2021). Automated segmentation was performed with the manufacturer's software (HEYEX version 1.9.10.0, VIEWING MODULE version 6.0.9.0) and subsequently corrected by trained graders. In the macular area, the regions analysed are generally defined by the Early Treatment Diabetic Retinopathy Study (ETDRS) grid automatically centred on the fovea by the fovea finder (see Figure 1). This acquisition protocol provides total volume and mGCL thickness in nine sectors: central fovea (CF), inner nasal (IN), outer nasal (ON), inner superior (IS), outer superior (OS), inner temporal (IT), outer temporal (OT), inner inferior (II) and outer inferior (OI).

2.4 | Expanded disability status scale

Disability was assessed at every visit of the 10-year longitudinal study by neurologists specialized in clinical MS studies. Following the standard criteria (Kalincik et al., 2015), which represents a significant worsening of disability, disability progression was defined by the reference EDSS and the EDSS variation (Δ EDSS). Disability progression was considered as follows: Δ EDSS of 1.5 or more points if reference EDSS was 0, Δ EDSS of 1 or more points if reference EDSS was between 1 and 5.5, and Δ EDSS of 0.5 or more points if reference EDSS was above 5.5.

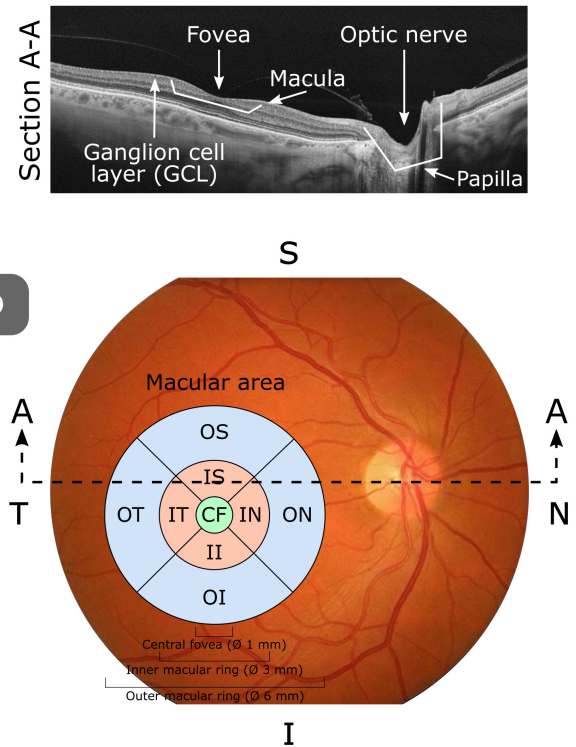


FIGURE 1 Standard Early Treatment Diabetic Retinopathy Study (ETDRS) grid on a right-eye retina. The macular area is divided into nine regions defined by three rings and four quadrants. The rings are central fovea (green), inner macular ring (orange) and outer macular ring (blue). The quadrants are superior (S), nasal (N), inferior (I) and temporal (T). In an optical coherence tomography (OCT) image, section A–A highlights the location of the ganglion cell layer (GCL) (CF, central fovea; II, inner inferior; IN, inner nasal; IS, inner superior; IT, inner temporal; OD, right eye; OI, outer inferior; ON, outer nasal; OS, outer superior; OT, outer temporal).

2.5 | Statistical analysis

Statistical analyses were performed using MATLAB (version 2020b, Mathworks Inc.). All hypothesis tests were evaluated at a significance level of 0.05. Comparison between groups was performed using the Wilcoxon test due to the non-normality of the variables after analysing the normality of the continuous variables using the Kolmogorov–Smirnov test. For categorical variables, *p*-value was obtained with Fisher's exact test since this test is preferable to the Chi-square test because of its accuracy. Moreover, the Chi-square test should be avoided if there are few observations (as is the case of our data).

2.6 | Deep learning

This paper aimed to assess the possibility of diagnosing MS and predicting disability progression in MS patients based on GCL thickness and deep learning techniques. This process was implemented in PYTHON 3.9 using PyTorch.

After filtering the databases by eliminating subjects with incomplete data and discarding variables that had not been collected in a suitably large number of subjects, classes were balanced if necessary. To reduce the risk of overfitting caused by class imbalance, the synthetic minority oversampling technique (SMOTE) was used

(Kuhn & Johnson, 2013; Santos et al., 2018). This method increases the number of minority class subjects by synthesizing new subjects.

In artificial intelligence, variable selection can bring benefits such as reduced overfitting, improved accuracy and reduced computational cost. To test this effect, the neural networks were used with and without the variable reduction performed by the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996). This technique generates regression coefficients to the model features and eliminates those with a regression coefficient equal to zero.

The architecture proposed in this study was composed of one fully connected neural network (FCNN) with one or more hidden layers (HLs), in particular, a feedforward backpropagation neural network. Feedforward is the process neural networks use to turn the input into an output. Backpropagation is used to train the FCNN via the following steps: perform a feedforward operation, compare the model output with the desired output, calculate the error and run the feedforward operation backwards (backpropagation) to spread the error to each of the weights. This method updates the weights and provides a better model. In practice, the activation function governing the output of neurons in HLs is the rectified linear unit (ReLU) function. The FCNN outputs were normalized using the softmax function that assigns probabilities to them. The error function was the negative log likelihood (NLL) loss. The Adam optimization algorithm, which is an extension of stochastic gradient descent, was used with a learning rate of 0.003. The optimizer decides by how much and in which direction to change the layer parameters. Finally, drop-out with 0.2 drop probability was used to reduce overfitting (Srivastava et al., 2014).

Another method used to minimize the risk of overfitting, given the difficulty of creating large clinical databases, was 10-fold cross-validation (Santos et al., 2018). Moreover, with k-fold cross-validation the final results are independent of the initial split (Rodríguez et al., 2010). For the numerical features, the training set was normalized with mean of 0 and a standard deviation of 1, and the validation set was normalized with the mean and standard deviation of the training set. Thus, during training our FCNN has no information from the validation set. One-hot encoding was used to encode the categorical features as numerical features (Potdar et al., 2017).

For the MS diagnosis model, as can be seen in Table 1, the raw data set from our cross-sectional study had 17 numerical features: 15 numerical features plus one categorical feature (sex) encoded into two numerical features. For the MS prognosis model, our 10-year longitudinal study was used with data from the first 3 years as input (see Table 2 and Figure 2) and the disability progression 8 years later as output. This decision was based on our previous study, in which we compared model performance using 2 or 3 years of follow-up (Montolío et al., 2022). In this case, the raw data set from 36 MS patients with disability progression (DP) and 36 MS patients with no disability progression (nDP) had 27 numerical features at each visit: 18 numerical features and four categorical features (sex, MS subtype, optic neuritis antecedent and

TABLE 1 Clinical characteristics and macular ganglion cell layer (mGCL) thicknesses from 72 multiple sclerosis (MS) patients and 30 healthy subjects. P-value comes from the Wilcoxon test for continuous variables and from Fisher's exact test for categorical variables. Statistically significant differences ($p < 0.05$) are highlighted in bold. (BCVA: best-corrected visual acuity; th: thickness).

	MS patients (n=72)	Healthy subjects (n=30)	P-value
Clinical characteristics			
Age [years]	47.45±10.62	47.62±14.07	0.794
Sex, n [%]			
Male	19 (26.4)	5 (16.7)	0.442
Female	53 (73.6)	25 (83.3)	
BCVA [Snellen]	1.01±0.24	1.00±0.11	0.024
mGCL thickness			
Total volume [mm ³]	0.88±0.14	1.10±0.10	<0.001
Central fovea th. [µm]	14.14±4.68	15.63±4.17	0.084
Inner nasal th. [µm]	39.36±9.82	52.47±6.71	<0.001
Outer nasal th. [µm]	31.54±4.58	37.63±3.67	<0.001
Inner superior th. [µm]	40.90±9.32	52.33±5.64	<0.001
Outer superior th. [µm]	30.55±7.32	36.63±3.52	<0.001
Inner temporal th. [µm]	35.67±10.09	46.93±4.68	<0.001
Outer temporal th. [µm]	29.44±6.34	36.50±3.99	<0.001
Inner inferior th. [µm]	40.42±9.60	53.00±4.57	<0.001
Outer inferior th. [µm]	28.97±7.50	33.30±4.52	<0.001
Centre th. [µm]	5.10±6.14	3.87±2.53	0.847
Centre min. th. [µm]	1.53±2.70	1.17±1.09	0.631
Centre max. th. [µm]	33.29±10.18	41.53±8.96	<0.001

relapse in preceding year) encoded into their respective numerical features. As we used data from the first 3 years of the follow-up, this raw data set contained 81 numerical features.

3 | RESULTS

Several neural network architectures were tested to analyse the performance of the MS diagnosis and prognosis models based on GCL thickness measured using Spectralis OCT.

The mean total volume and mean thickness in each sector of the mGCL were computed for MS patients and healthy controls. It was observed that this volume and these thicknesses were lesser in MS patients than in healthy subjects. Table 1 shows the results of the Wilcoxon test for continuous features and of Fisher's exact test for categorical features. Statistically significant differences were obtained for BCVA in clinical characteristics and for all OCT data except for central fovea, centre and centre minimum thicknesses.

Since our cross-sectional study contained class-imbalanced data, the SMOTE was used to create 42 synthetic healthy controls in order to have 72 MS patients and 72 healthy controls. After applying variable selection using the LASSO, the selected features were: total volume, IN, OS and II thicknesses (regression

TABLE 2 Clinical characteristics and macular ganglion cell layer (mGCL) thicknesses from 36 multiple sclerosis (MS) patients with disability progression (DP) and 36 MS patients with no disability progression (nDP) at the first 3 years of the 10-year follow-up. P-value comes from the Wilcoxon test for continuous variables and from Fisher's exact test for categorical variables. Statistically significant differences ($p < 0.05$) are highlighted in bold. (BCVA: best-corrected visual acuity; RRMS: relapsing–remitting multiple sclerosis; SPMS: secondary-progressive multiple sclerosis; PPMS: primary-progressive multiple sclerosis; EDSS: expanded disability status scale; th: thickness).

	MS patients with DP (n=36)		MS patients with nDP (n=36)		MS patients with DP (n=36)		MS patients with nDP (n=36)		MS patients with DP (n=36)		MS patients with nDP (n=36)	
	Baseline		1 year		2 years		1 year		2 years		2 years	
		P-value		P-value		P-value		P-value		P-value		P-value
Clinical characteristics												
Age [years]	43.98±10.11	0.570	46.24±11.38	0.285	44.80±9.98	0.510	47.29±11.50	0.285	46.12±10.28	0.510	48.88±11.35	0.401
Sex, n [%]												
Male	12 (33.3)		7 (19.4)		12 (33.3)		7 (19.4)		12 (33.3)		7 (19.4)	
Female	24 (66.7)		29 (80.6)		24 (66.7)		29 (80.6)		24 (66.7)		29 (80.6)	
BCVA [Snellen]	0.94±0.24	0.485	0.89±0.22	0.146	0.94±0.21	0.942	0.92±0.23	0.834	0.94±0.21	0.942	0.93±0.22	0.815
MS duration [years]	11.25±6.07	1.000	14.25±8.58	0.799	11.89±6.10	1.000	15.06±8.75	0.834	12.78±6.00	0.159	16.60±9.28	0.087
MS subtype, n [%]												
RRMS	33 (91.7)		32 (88.9)		33 (91.7)		31 (86.1)		33 (91.7)		31 (86.1)	
SPMS	2 (5.5)		3 (8.3)		2 (5.5)		4 (11.1)		2 (5.5)		4 (11.1)	
PPMS	1 (2.8)		1 (2.8)		1 (2.8)		1 (2.8)		1 (2.8)		1 (2.8)	
Optic neuritis antecedent, n [%]		0.799		0.799		0.799		0.799		0.799		1.000
Yes	12 (33.3)		10 (27.8)		12 (33.3)		10 (27.8)		12 (33.3)		11 (30.6)	
No	24 (66.7)		26 (72.2)		24 (66.7)		26 (72.2)		24 (66.7)		25 (69.4)	
Relapse in preceding year, n [%]		1.000		1.000		1.000		1.000		1.000		0.429
Yes	5 (13.9)		4 (11.1)		5 (13.9)		6 (16.7)		2 (5.6)		5 (13.9)	
No	31 (86.1)		32 (88.9)		31 (86.1)		30 (83.3)		34 (94.4)		31 (86.1)	
EDSS	1.88±2.12	0.013	3.22±2.36	0.013	1.79±2.12	0.041	3.06±2.60	0.041	1.86±2.38	0.041	3.31±2.55	0.007
Visual EDSS	0.81±0.82	0.462	1.14±1.31	0.462	0.78±0.72	0.626	1.08±1.30	0.626	0.78±0.76	0.626	1.08±1.30	0.589
mGCL thickness												
Total volume [mm ³]	0.93±0.17	0.377	0.89±0.15	0.377	0.91±0.17	0.437	0.89±0.14	0.437	0.90±0.16	0.437	0.88±0.17	0.783
Central fovea th. [µm]	15.11±4.10	0.286	14.50±4.69	0.286	14.72±4.45	0.462	14.22±4.67	0.462	14.03±4.58	0.462	14.64±4.41	0.542
Inner nasal th. [µm]	41.14±11.89	0.951	41.31±8.15	0.951	40.03±11.71	0.879	40.94±7.64	0.879	38.89±11.75	0.879	40.94±10.27	0.530
Outer nasal th. [µm]	33.67±6.46	0.631	32.42±3.89	0.631	32.64±5.74	0.531	32.11±3.99	0.531	31.47±5.21	0.531	31.91±6.41	0.734
Inner superior th. [µm]	44.08±10.70	0.589	42.92±7.75	0.589	41.81±11.17	1.000	42.57±10.16	1.000	40.03±10.33	1.000	41.89±7.67	0.535
Outer superior th. [µm]	33.44±5.54	0.334	32.06±11.14	0.334	32.03±4.93	0.868	31.47±10.80	0.868	30.89±4.26	0.868	31.36±9.35	0.772
Inner temporal th. [µm]	38.55±11.07	0.443	36.72±9.32	0.443	37.33±11.15	0.501	35.83±11.03	0.501	35.92±10.80	0.501	35.17±8.91	0.813

(Continues)

TABLE 2 (Continued)

	MS patients with DP (n=36)		MS patients with nDP (n=36)		MS patients with DP (n=36)		MS patients with nDP (n=36)		MS patients with DP (n=36)		MS patients with nDP (n=36)	
	P-value		P-value		P-value		P-value		P-value		P-value	
	1 year		2 years		1 year		2 years		1 year		2 years	
Outer temporal th. [μm]	31.47 \pm 6.20	29.69 \pm 5.04	0.254	0.371	30.49 \pm 8.24	29.23 \pm 6.87	0.371	0.371	30.11 \pm 4.80	30.28 \pm 5.83	0.910	0.910
Inner inferior th. [μm]	43.81 \pm 11.72	42.00 \pm 8.33	0.410	0.569	42.94 \pm 10.45	41.75 \pm 7.59	0.410	0.569	40.31 \pm 8.03	40.83 \pm 10.59	0.681	0.681
Outer inferior th. [μm]	31.91 \pm 9.92	31.03 \pm 9.28	0.524	0.768	30.59 \pm 8.33	30.23 \pm 5.81	0.524	0.768	29.65 \pm 7.48	29.06 \pm 4.01	0.608	0.608
Centre th. [μm]	3.39 \pm 2.65	5.47 \pm 6.05	0.106	0.852	4.89 \pm 3.74	5.61 \pm 6.42	0.106	0.852	5.72 \pm 6.90	4.72 \pm 5.42	0.462	0.462
Centre min. th. [μm]	1.22 \pm 1.57	2.00 \pm 3.25	0.386	0.370	1.44 \pm 1.80	2.28 \pm 3.51	0.386	0.370	1.64 \pm 3.27	1.28 \pm 1.61	0.967	0.967
Centre max. th. [μm]	33.89 \pm 12.33	33.22 \pm 9.07	0.689	0.430	34.86 \pm 11.78	32.50 \pm 8.12	0.689	0.430	34.33 \pm 8.05	34.33 \pm 12.94	0.879	0.879

coefficients: 0.0428, 0.0029, 0.0311 and 0.0059, respectively). These features showed a statistically significant difference between MS patients and healthy controls (see Table 1). The influence of the reduced data set on model performance was thus analysed.

For MS diagnosis, the best neural network architecture for both raw and reduced data is shown in Figure 3. With raw data (17 features), the greatest accuracy (90.3%) was obtained with two HLs and ReLU activation and softmax activation in the output layer. As can be seen in the confusion matrix (Figure 5), there were six false positives (FPs) and eight false negatives (FNs). After variable selection, the best accuracy (82.6%) was achieved with the FCNN architecture as follows: four input units, two HLs with 4 units and ReLU activation, and softmax activation in the output layer. In this case, the confusion matrix shows 15 FPs and 10 FN. Both results were obtained using 400 epochs.

The statistical analysis performed on our raw data used for MS prognosis revealed that only EDSS showed statistically significant differences between MS patients with DP and MS patients with nDP at baseline, 1 and 2 years (see Table 2).

Our 10-year longitudinal study for DP prediction was balanced as there were 36 MS patients with DP (EDSS at 10 years=3.78 \pm 2.28) and 36 MS patients with nDP (EDSS at 10 years=3.49 \pm 2.41). Variable selection was applied to the 81 raw features and the reduced data set contained 12 features. The regression coefficient for each feature is detailed below in square brackets. These features were as follows: five at baseline (EDSS [0.1087], MS duration [0.0034], OS [0.0098], OT [0.0018] and centre thicknesses [0.0380]), three at 1 year (EDSS [0.0845], MS duration [0.0043] and OS thickness [0.0483]) and four at 2 years (EDSS [0.0768], MS duration [0.0023], IS [0.0956] and OS thicknesses [0.0007]). As can be seen, EDSS ($p < 0.05$), MS duration and OS thickness were chosen by LASSO for all time points.

We evaluated the ability of different FCNNs to predict whether a MS patient would suffer long-term DP based on GCL data collected in the first 3 years. Using the raw data and the FCNN architecture detailed at the top of Figure 4, an accuracy of 65.3% was obtained for 300 epochs. With a reduced data set and the FCNN structure indicated at the bottom of Figure 4, the accuracy was 81.9% for 400 epochs. The confusion matrices for both MS prognosis models can be seen in Figure 5, where the model before variable selection presented 14 FPs and 12 FN, while the model with reduced data obtained seven FPs and six FN.

4 | DISCUSSION

In this paper, we propose and implement a deep learning approach to MS diagnosis and prognosis that analyses GCL data obtained using Spectralis OCT. Very few studies have investigated AI model performance in the diagnosis and prediction of disability progression of this disease using OCT data because most papers focus on MRI data. In recent years, several authors have tested the feasibility of using GCIPL measurements

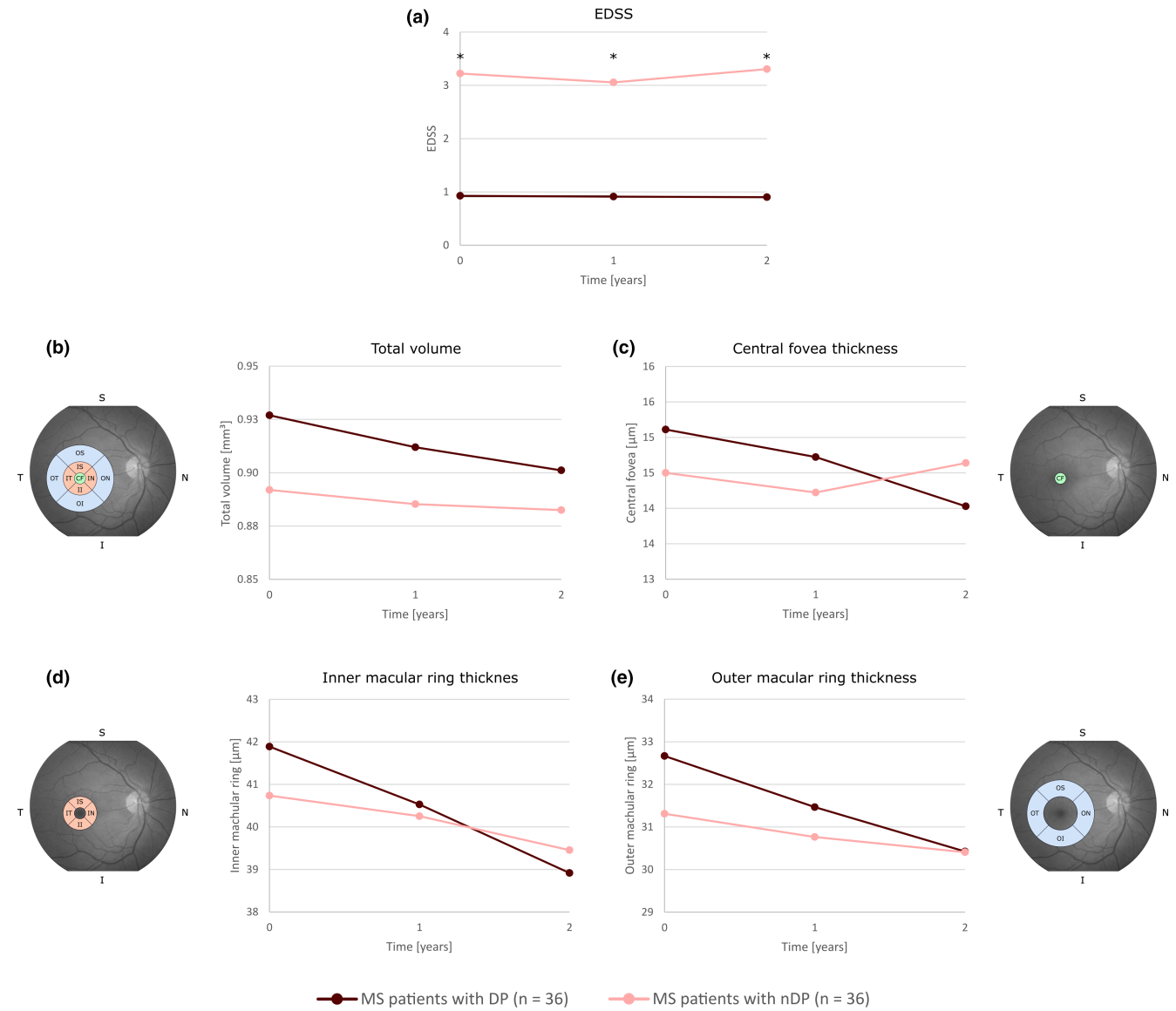


FIGURE 2 Changes over time in several characteristics from 36 multiple sclerosis (MS) patients with disability progression (DP) and 36 MS patients no disability progression (NDP). (a) Expanded Disability Status Scale (EDSS). (b) Total volume of the macular ganglion cell layer (mGCL). (c, d) Thickness of the mGCL in rings defined by the standard Early Treatment Diabetic Retinopathy Study (ETDRS): central fovea (green), inner macular ring (orange) and outer macular ring (blue). * indicates statistically significant differences ($p < 0.05$) (CF, central fovea; I, inferior; II, inner inferior; IN, inner nasal; IS, inner superior; IT, inner temporal; N, nasal; OI, outer inferior; ON, outer nasal; OS, outer superior; OT, outer temporal; S, superior; T, temporal).

to diagnose MS with techniques such as the SVM or ANNs (Cavaliere et al., 2019; Garcia-Martin et al., 2021). However, when predicting disability progression, GCL thickness has only been analysed from a statistical point of view (Lambe et al., 2021; Schurz et al., 2021).

GCIPL thinning in MS is caused by axonal degeneration and is related to worsening physical and cognitive disability and brain atrophy (Bsteh et al., 2020; Coric et al., 2018; Zimmermann et al., 2018). As can be seen in the statistical analysis between MS patients and healthy controls (Table 1), both macular volume and the different mGCL thicknesses were lower in the MS patients and exhibited significant difference (except for CF). The MS patients were divided into patients with DP and patients with NDP following the standard criteria defined above. Of all the characteristics, the difference was only significant in the EDSS in the 3 years used (see Table 2). In line with this result, Dekker et al. (2019) showed the power of the EDSS for disability progression prediction. Figure 2a

shows how the EDSS score for MS patients with NDP is higher than for MS patients with DP. On the contrary, macular volume and mGCL thickness were higher for MS patients with DP and their decrease in the first 3 years was also greater (see Figure 2b–e). These findings are in line with those of previous studies that corroborated that axonal damage occurs cumulatively from the onset of MS and that most retinal thinning occurs in the early stages of the disease (Montolío et al., 2021, 2022). However, while the differences between MS patients with DP and MS patients with NDP were significant in the RNFL in previous studies, these differences were not significant in the GCL.

For MS diagnosis purposes, the best accuracy was 90.3% (AUC: 0.9028), achieved with our raw data set. Pérez del Palomar et al. (2019) obtained an accuracy lower than 75% for macular area using GCIPL thickness measured by DRI OCT Triton. Although Cavaliere et al. (2019) did not use mGCL separately, a result

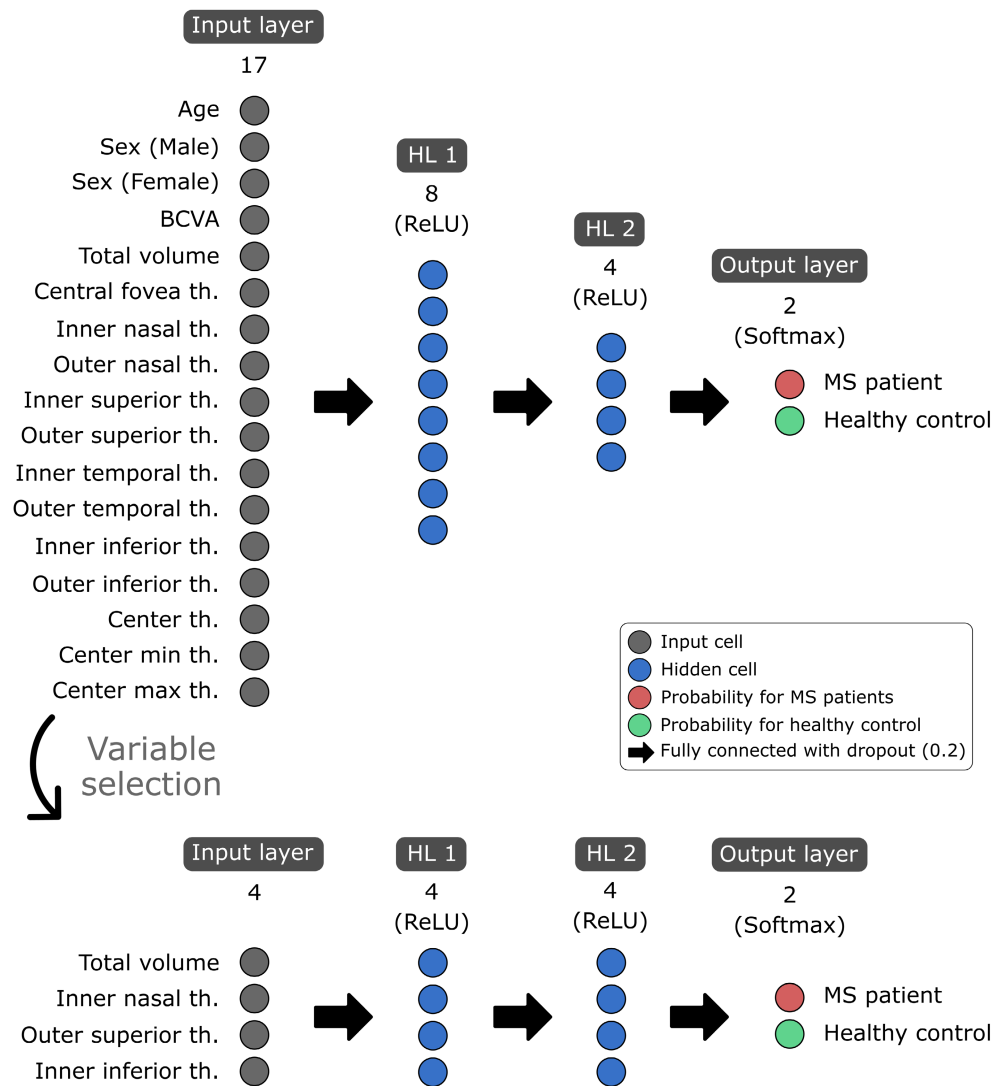


FIGURE 3 Neural network architecture for MS diagnosis model. With the raw data set, a fully connected neural network (FCNN) was used with 17 input units, a hidden layer (HL) with 8 hidden units and rectified linear unit (ReLU) activation, then an HL with 4 units and ReLU activation, and finally an output layer with softmax activation. After variable selection using the least absolute shrinkage and selection operator (LASSO), the FCNN was as follows: 4 input units, 2 HLs with 4 units and ReLU activation, and softmax activation in the output layer (BCVA, best-corrected visual acuity; MS, multiple sclerosis; th, thickness).

comparable with this study is the AUC of 0.835 obtained using the retinal thickness in macular and peripapillary areas by SVM. An AUC of 0.83 was also obtained by Garcia-Martin et al. (2021) using the GCL thickness in a 9×12 mm area encompassing the macula and the optic nerve. Therefore, using only the mGCL thickness, our result was better than those obtained in previous papers.

In the case of MS prognosis, as can be seen in Figure 5b, the best result (acc: 81.9; AUC: 0.8194) was obtained with our reduced data set. In this case, in contrast to the MS diagnosis model, the feature selection performed using the LASSO increased model performance. Therefore, the feature-to-sample ratio in the raw data set was too high (81 features and 36 samples per class). With feature reduction, this ratio decreased (12 features shown in Figure 4) to a more effective value (Peduzzi et al., 1996). Given the lack of AI models that use GCL thickness to predict disability progression, another alternative is to compare our model with those models that used RNFL thickness for the same purpose. In our previous paper based on the RNFL, and also measured using Spectralis OCT, the best results were an accuracy

of 95.8% for MS diagnosis and of 91.3% for MS prognosis (Montolío et al., 2022). In view of the findings, we can conclude that the RNFL performed better than the GCL for these tasks and that both retinal layers are valuable biomarkers in MS.

Other biomarkers, in combination with artificial intelligence techniques, were used for the same purpose. Most of the studies are related to diagnostic imaging such as brain MRI, which has become one of the main clinical tools for diagnosing and monitoring MS (Hashemi et al., 2019; Salem et al., 2019). The models developed by Zurita et al. (2018) were able to distinguish between patients and healthy subjects, reaching accuracy levels of 89%. Pinto et al. (2020) predicted disability progression after 4 years based on 2-year follow-up with MRI, CSF analysis and EP data; and the result was an AUC of 0.89. Work by Yperman et al. (2020) obtained an AUC of 0.75 when predicting disability progression after 2 years using EP time series. It was shown that the mGCL performed as well as the previous biomarkers analysed.

This study makes an important contribution to the available literature, since we used a real-life cohort in

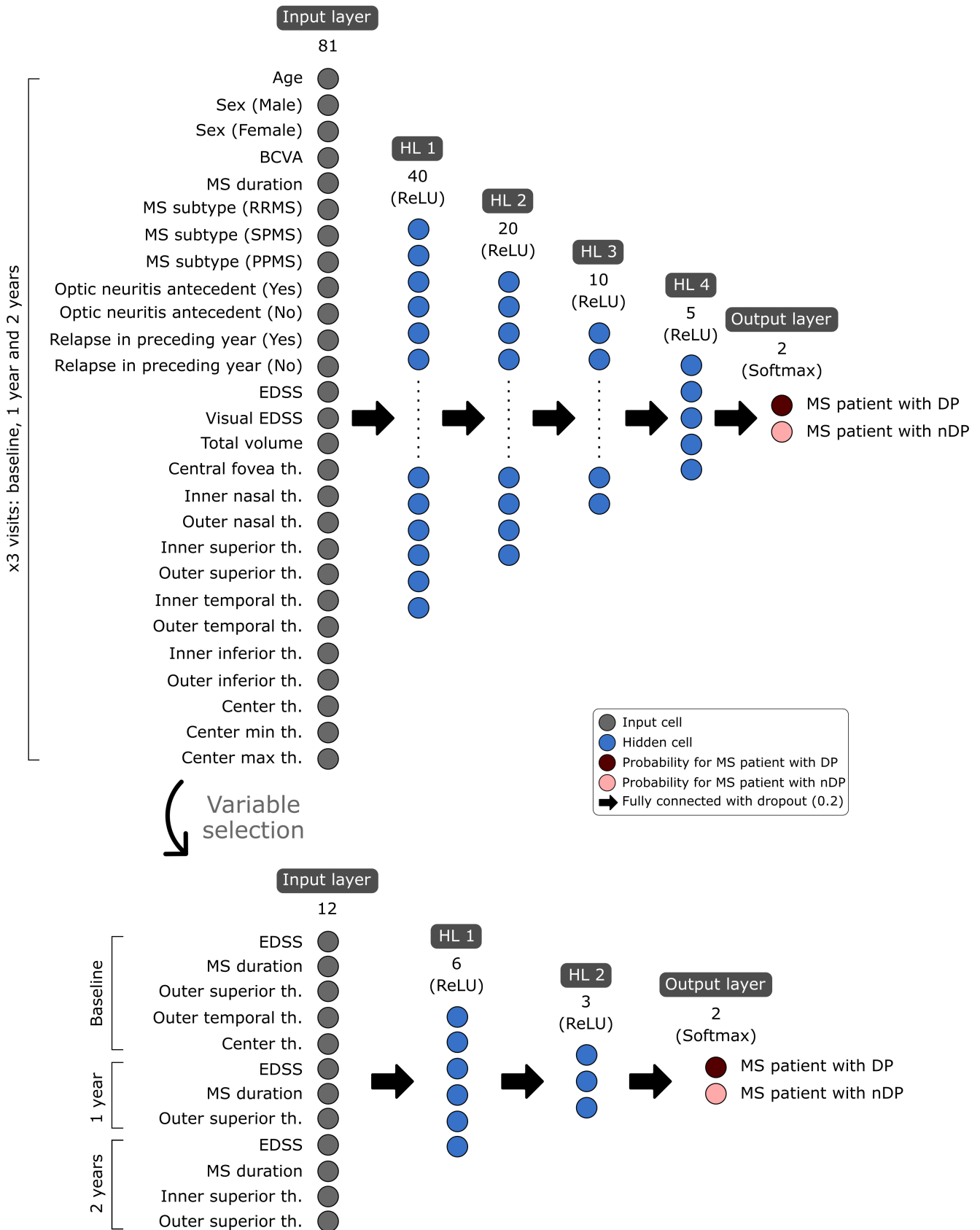


FIGURE 4 Neural network architecture for MS prognosis model. With the raw data set, a fully connected neural network (FCNN) was used with 81 input units, four hidden layers (HL) with rectified linear unit (ReLU) activation, and finally an output layer with softmax activation. After variable selection using the least absolute shrinkage and selection operator (LASSO), the architecture comprised 12 input units, an HL with 6 hidden units and ReLU activation, then an HL with 3 units with ReLU activation, and the output layer with softmax activation (BCVA, best-corrected visual acuity; DP, disability progression; EDSS, expanded disability status scale; MS, multiple sclerosis; nDP, no disability progression; PPMS, primary-progressive multiple sclerosis; RRMS, relapsing–remitting multiple sclerosis; SPMS, secondary-progressive multiple sclerosis; th, thickness).

(a) MS diagnosis model

Confusion Matrix

Target Class

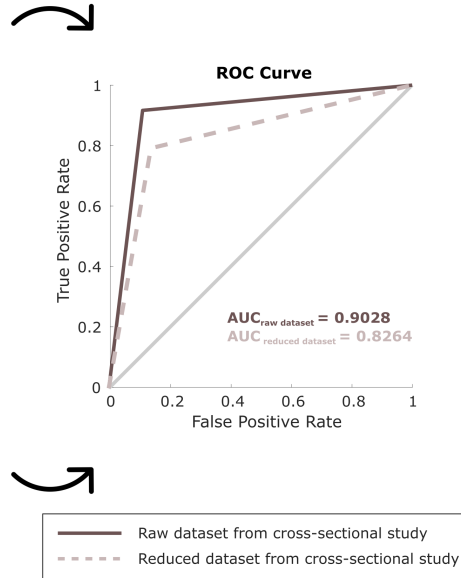
		MS patient	Healthy control	
Output Class	MS patient	64 44.4%	6 4.2%	91.4% 8.6%
	Healthy control	8 5.6%	66 45.8%	89.2% 10.8%
		88.9% 11.1%	91.7% 8.3%	90.3% 9.7%

Variable selection

Confusion Matrix

Target Class

		MS patient	Healthy control	
Output Class	MS patient	62 43.1%	15 10.4%	80.5% 19.5%
	Healthy control	10 6.9%	57 39.6%	85.1% 14.9%
		86.1% 13.9%	79.2% 20.8%	82.6% 17.4%



(b) MS prognosis model

Confusion Matrix

Target Class

		MS patient with DP	MS patient with nDP	
Output Class	MS patient with DP	24 33.3%	14 19.4%	63.2% 36.8%
	MS patient with nDP	12 16.7%	22 30.6%	64.7% 35.3%
		66.7% 33.3%	61.1% 38.9%	63.9% 36.1%

Variable selection

Confusion Matrix

Target Class

		MS patient with DP	MS patient with nDP	
Output Class	MS patient with DP	30 41.7%	7 9.7%	81.1% 18.9%
	MS patient with nDP	6 8.3%	29 40.3%	82.9% 17.1%
		83.3% 16.7%	80.6% 19.4%	81.9% 18.1%

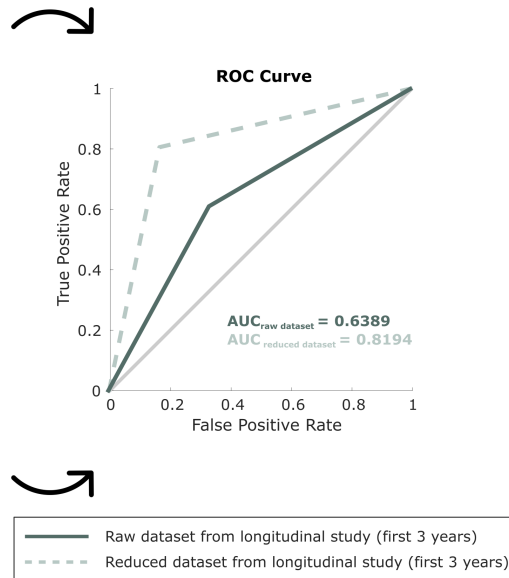


FIGURE 5 Confusion matrices and receiver operating characteristic (ROC) curves for the best results. (a) Multiple sclerosis (MS) model with raw and reduced data from our cross-sectional study. (b) MS prognosis model with raw and reduced data from the first 3 years of our 10-year longitudinal study (AUC, area under the curve; DP, disability progression; nDP, no disability progression).

which OCT scans were obtained by the ophthalmology service at Miguel Servet Hospital as part of clinical routine. A decrease in mGCL thickness could be one of the most promising biomarkers of MS-associated neurodegeneration. Currently, disability progression is mostly measured by the EDSS. However, this scale is limited as it reflects neurodegenerative damage incompletely, has low sensitivity and does not take into account neuropsychological disability (Hegen et al., 2018). OCT is non-invasive, cost-effective, easy to perform, fast and provides standardized and reliable quantitative measurements (London et al., 2019; Ontaneda & Fox, 2017). Furthermore, unlike brain volume measured by MRI, GCL thickness is not directly affected by inflammation and thus represents a valuable measure of neurodegeneration, an aspect especially relevant for MS prognosis (Petzold et al., 2017).

This study has several limitations. As in our previous research (Montolío et al., 2021, 2022), the sample size remains low, which limits the reproducibility of the study. Measurement errors present in the real-world cohort were minimized by controlling the quality of the OCT scans and following the inclusion/exclusion criteria. Another important aspect is that the antecedent and timing of optic neuritis were considered in our model (see Table 2). The data collected in the clinical routine pertain to MS patients with a disease duration of 12.75 ± 7.53 years at baseline. Therefore, the disability status at the beginning of our longitudinal study could be higher than in the early-stage MS population. In addition, the variation in disease duration and the EDSS at baseline may cause bias. However, with our inclusion criteria we aimed to represent as broad an MS population as possible. In order to reduce this bias, the standard criteria for disability progression depend on the reference EDSS in each MS patient. Thus, although the disability status of each patient is different, the prognosis of their disability progression depends on their reference EDSS value. Disability progression could be affected by the treatment selected for each MS patient. However, this is not easy to capture as patients may change treatment more than once during follow-up. The distribution of the first treatment received by MS patients during follow-up was as follows: 15 patients (8 DP and 7 NDP) received Avonex® (intramuscular interferon beta-1a), 19 patients (9 DP and 10 NDP) received Betaseron® (interferon beta-1b), 17 patients (8 DP and 9 NDP) received Rebif® (subcutaneous interferon beta-1a), 16 patients (7 DP and 9 NDP) received Copaxone® (glatiramer acetate) and 5 patients (3 DP and 2 NDP) received no treatment. Moreover, 57 MS patients (30 DP and 27 NDP) changed treatment at least once during the longitudinal study.

Another limitation is the class-imbalanced data in our cross-sectional study for MS diagnosis. As there were 72 MS patients and 30 healthy controls, the SMOTE was used to oversample the minority class and reduce the risk of overfitting to the majority class. However, the use of the SMOTE results in the variability of the minority class being underestimated and the training set samples are no longer independent (Blagus & Lusa, 2013). An additional analysis was conducted following another approach with the

30 available healthy controls and 30 randomly selected MS patients for MS diagnosis. The result with this new class-balanced data was an accuracy of 87.6%. This outcome, very similar to that obtained after applying the SMOTE, showed the robustness of our model.

Adopting the idea of harnessing the benefits of AI applied to MS (Afzal et al., 2020; Bonacchi et al., 2022), we contributed to the research field that uses OCT data in combination with AI models for MS diagnosis and prognosis (Kenney et al., 2022). In addition to all the advantages already mentioned, OCT is a test that can be performed by any clinician in a few minutes without discomfort for the patient. In conclusion, both cross-sectional and longitudinal measurements of mGCL thickness are reliable biomarkers with which to diagnose MS and predict disability progression in each patient. GCL thinning therefore represents a quantitative measure of MS-associated neuroaxonal damage. Including the use of OCT technology in the treatment of MS would bring great benefits to clinicians, who would make early and comprehensive diagnosis and would select the most specific treatments with which to improve patients' lives.

Future directions for this research could be validation in other cohorts of subjects with suspected MS and, in addition, prediction of its development, evaluating the specificity of the algorithms. It would also be of clinical interest to develop models with which to identify diseases with MS-like symptoms.

ACKNOWLEDGEMENTS


This study was supported by the Spanish Ministry of Economy and Competitiveness (project DPI 2016-79302-R), the Spanish Ministry of Science, Innovation and Universities (grant BES-2017-080384) and the Carlos III Health Institute (PI17/01726).

ORCID

Alberto Montolío  <https://orcid.org/0000-0001-7248-4399>

José Cegoñino  <https://orcid.org/0000-0002-2967-6747>

Elena Garcia-Martin  <https://orcid.org/0000-0001-6258-2489>

Amaya Pérez del Palomar  <https://orcid.org/0000-0003-0669-777X>

REFERENCES

- Afzal, H.M.R., Luo, S., Ramadan, S. & Lechner-Scott, J. (2020) The emerging role of artificial intelligence in multiple sclerosis imaging. *Multiple Sclerosis Journal*, 28, 849–858.
- Alonso, R., Gonzalez-Moron, D. & Garcea, O. (2018) Optical coherence tomography as a biomarker of neurodegeneration in multiple sclerosis: a review. *Multiple Sclerosis and Related Disorders*, 22, 77–82.
- Betzler, B.K., Rim, T.H., Sabanayagam, C. & Cheng, C.-Y. (2022) Artificial intelligence in predicting systemic parameters and diseases from ophthalmic imaging. *Frontiers in Digital Health*, 4, 889445.
- Blagus, R. & Lusa, L. (2013) SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14, 106.
- Bonacchi, R., Filippi, M. & Rocca, M.A. (2022) Role of artificial intelligence in MS clinical practice. *NeuroImage Clin*, 35, 103065.
- Britze, J., Pihl-Jensen, G. & Frederiksen, J.L. (2017) Retinal ganglion cell analysis in multiple sclerosis and optic neuritis: a systematic review and meta-analysis. *Journal of Neurology*, 264, 1837–1853.
- Bsteh, G., Berek, K., Hegen, H., Altmann, P., Wurth, S., Auer, M. et al. (2020) Macular ganglion cell–inner plexiform layer thinning

- as a biomarker of disability progression in relapsing multiple sclerosis. *Multiple Sclerosis Journal*, 27, 684–694.
- Cavaliere, C., Vilades, E., Alonso-Rodríguez, M., Rodrigo, M., Pablo, L., Miguel, J. et al. (2019) Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. *Sensors*, 19, 5323.
- Chylack, L.T., Wolfe, J.K., Singer, D.M. et al. (1993) The lens opacities classification system III. The longitudinal study of cataract study group. *Archives of Ophthalmology*, 111, 831–836.
- Cordon, B., Vilades, E., Orduna, E., Satue, M., Perez-Velilla, J., Sebastian, B. et al. (2020) Angiography with optical coherence tomography as a biomarker in multiple sclerosis. *PLoS One*, 15, e0243236.
- Coric, D., Balk, L.J., Verrijp, M., Eijlers, A., Schoonheim, M.M., Killestein, J. et al. (2018) Cognitive impairment in patients with multiple sclerosis is associated with atrophy of the inner retinal layers. *Multiple Sclerosis*, 24, 158–166.
- Cruz-Herranz, A., Balk, L.J., Oberwahrenbrock, T., Saidha, S., Martinez-Lapiscina, E.H., Lagreze, W.A. et al. (2016) The APOSTEL recommendations for reporting quantitative optical coherence tomography studies. *Neurology*, 86, 2303–2309.
- Dekker, I., Eijlers, A.J.C., Popescu, V., Balk, L.J., Vrenken, H., Wattjes, M.P. et al. (2019) Predicting clinical progression in multiple sclerosis after 6 and 12 years. *European Journal of Neurology*, 26, 893–902.
- Eslami, F., Ghiasian, M., Khanlarzade, E. & Moradi, E. (2020) Retinal nerve fiber layer thickness and total macular volume in multiple sclerosis subtypes and their relationship with severity of disease, a cross-sectional study. *Eye Brain*, 12, 15–23.
- Frohman, E., Costello, F., Zivadinov, R., Stuve, O., Conger, A., Winslow, H. et al. (2006) Optical coherence tomography in multiple sclerosis. *Lancet Neurology*, 5, 853–863.
- Garcia-Martin, E., Ortiz, M., Boquete, L., Sánchez-Morla, E.M., Barea, R., Cavaliere, C. et al. (2021) Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier. *Computers in Biology and Medicine*, 129, 104165.
- Graham, E.C., You, Y., Yiannikas, C., Garrick, R., Parratt, J., Barnett, M.H. et al. (2016) Progressive loss of retinal ganglion cells and axons in nonoptic neuritis eyes in multiple sclerosis: a longitudinal optical coherence tomography study. *Investigative Ophthalmology and Visual Science*, 57, 2311–2317.
- Hashemi, S.R., Salehi, S.S.M., Erdogmus, D., Prabhu, S.P., Warfield, S.K. & Gholipour, A. (2019) Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: application to multiple sclerosis lesion detection. *IEEE Access*, 7, 1721–1735.
- Hegen, H., Bsteh, G. & Berger, T. (2018) 'No evidence of disease activity' – is it an appropriate surrogate in multiple sclerosis? *European Journal of Neurology*, 25, 1107–e101.
- Kalincik, T., Cutter, G., Spelman, T., Jokubaitis, V., Havrdova, E., Horakova, D. et al. (2015) Defining reliable disability outcomes in multiple sclerosis. *Brain*, 138, 3287–3298.
- Kenney, R.C., Liu, M., Hasanaj, L., Joseph, B., Abu al-Hassan, A., Balk, L.J. et al. (2022) The role of OCT criteria and machine learning in multiple sclerosis and optic neuritis diagnosis. *Neurology*, 99, e1100–e1112. Available from: <https://doi.org/10.1212/WNL.0000000000200883>
- Ker, J., Wang, L., Rao, J. & Lim, T. (2018) Deep learning applications in medical image analysis. *IEEE Access*, 6, 9375–9389.
- Kuhn, M. & Johnson, K. (2013) *Applied predictive modeling*. New York, NY: Springer New York.
- Lambe, J., Fitzgerald, K.C., Murphy, O.C., Filippatou, A.G., Sotirchos, E.S., Kalaitzidis, G. et al. (2021) Association of Spectral-Domain OCT with long-term disability worsening in multiple sclerosis. *Neurology*, 96, e2058–e2069.
- London, F., Zéphir, H., Drumez, E., Labreuche, J., Hadhoum, N., Lannoy, J. et al. (2019) Optical coherence tomography: a window to the optic nerve in clinically isolated syndrome. *Brain*, 142, 903–915.
- López-Dorado, A., Ortiz, M., Satue, M., Rodrigo, M.J., Barea, R., Sánchez-Morla, E.M. et al. (2021) Early diagnosis of multiple sclerosis using swept-source optical coherence tomography and convolutional neural networks trained with data augmentation. *Sensors*, 22, 167.
- Marzullo, A., Kocevar, G., Stamile, C., Durand-Dubief, F., Terracina, G., Calimeri, F. et al. (2019) Classification of multiple sclerosis clinical profiles via graph convolutional neural networks. *Frontiers in Neuroscience*, 13, 1–15.
- McKinley, R., Wepfer, R., Grunder, L., Aschwanden, F., Fischer, T., Friedli, C. et al. (2020) Automatic detection of lesion load change in multiple sclerosis using convolutional neural networks with segmentation confidence. *NeuroImage. Clinical*, 25, 102104.
- Montolio, 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. *Annals of Biomedical Engineering*, 50, 507–528.
- Montolio, A., Cegoñino, J., Orduna, E., Sebastian, B., Garcia-Martin, E. & Pérez del Palomar, A. (2019) A mathematical model to predict the evolution of retinal nerve fiber layer thinning in multiple sclerosis patients. *Computers in Biology and Medicine*, 111, 103357.
- Montolio, A., Martín-Gallego, A., Cegoñino, J., Orduna, E., Vilades, E., Garcia-Martin, E. et al. (2021) Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. *Computers in Biology and Medicine*, 133, 104416.
- Ontaneda, D. & Fox, R.J. (2017) Imaging as an outcome measure in multiple sclerosis. *Neurotherapeutics*, 14, 24–34.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49, 1373–1379.
- Pérez del Palomar, A., Cegoñino, J., Montolio, A., Orduna, E., Vilades, E., Sebastián, B. et al. (2019) Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques. *PLoS One*, 14, e0216410.
- Petzold, A., Albrecht, P., Balcer, L., Bekkers, E., Brandt, A.U., Calabresi, P.A. et al. (2021) Artificial intelligence extension of the OSCAR-IB criteria. *Annals of Clinical Translational Neurology*, 8, 1528–1542.
- Petzold, A., Balcer, L.J., Calabresi, P.A., Costello, F., Frohman, T.C., Frohman, E.M. et al. (2017) Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurology*, 16, 797–812.
- Pietroboni, A.M., Dell'Arti, L., Caprioli, M., Scarioni, M., Carandini, T., Arighi, A. et al. (2019) The loss of macular ganglion cells begins from the early stages of disease and correlates with brain atrophy in multiple sclerosis patients. *Multiple Sclerosis Journal*, 25, 31–38.
- Pinto, M.F., Oliveira, H., Batista, S., Cruz, L., Pinto, M., Correia, I. et al. (2020) Prediction of disease progression and outcomes in multiple sclerosis with machine learning. *Scientific Reports*, 10, 1–13.
- Polman, C.H., Reingold, S.C., Banwell, B., Clanet, M., Cohen, J.A., Filippi, M. et al. (2011) Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology*, 69, 292–302.
- Potdar, K., Pardawala, T. & Pai, C. (2017) A comparative study of categorical variable encoding techniques for neural network classifiers. *International Journal of Computers and Applications*, 175, 7–9.
- Pueyo, V., Martin, J., Fernandez, J., Almarcegui, C., Ara, J., Egea, C. et al. (2008) Axonal loss in the retinal nerve fiber layer in patients with multiple sclerosis. *Multiple Sclerosis Journal*, 14, 609–614.
- Rocca, M.A., Anzalone, N., Storelli, L., Del Poggio, A., Cacciaguerra, L., Manfredi, A.A. et al. (2021) Deep learning on conventional magnetic resonance imaging improves the diagnosis of multiple sclerosis mimics. *Investigative Radiology*, 56, 252–260.
- Rodríguez, J.D., Perez, A. & Lozano, J.A. (2010) Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 569–575.
- Saidha, S., Al-Louzi, O., Ratchford, J.N. et al. (2015) Optical coherence tomography reflects brain atrophy in multiple sclerosis: a four-year study. *Annals of Neurology*, 78, 801–813.
- Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J. et al. (2019) Multiple sclerosis lesion synthesis in MRI using an encoder-decoder U-NET. *IEEE Access*, 7, 25171–25184.

- Santos, M.S., Soares, J.P., Abreu, P.H., Araujo, H. & Santos, J. (2018) Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches. *IEEE Computational Intelligence Magazine*, 13, 59–76.
- Schurz, N., Sariaslani, L., Altmann, P., Leutmezer, F., Mitsch, C., Pemp, B. et al. (2021) Evaluation of retinal layer thickness parameters as biomarkers in a real-world multiple sclerosis cohort. *Eye Brain*, 13, 59–69.
- Seccia, R., Romano, S., Salvetti, M., Crisanti, A., Palagi, L. & Grassi, F. (2021) Machine learning use for prognostic purposes in multiple sclerosis. *Life*, 11, 1–18.
- Shi, C., Jiang, H., Gameiro, G.R., Hu, H., Hernandez, J., Delgado, S. et al. (2019) Visual function and disability are associated with focal thickness reduction of the ganglion cell-inner plexiform layer in patients with multiple sclerosis. *Investigative Ophthalmology & Visual Science*, 60, 1213–1223.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Swanton, J., Fernando, K. & Miller, D. (2014) Early prognosis of multiple sclerosis. In: *Handb Clin Neurol*, Vol. 122, 1st edition. Amsterdam: Elsevier B.V.
- Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G. et al. (2018) Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurology*, 17, 162–173.
- Tibshirani, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288.
- Toledo, J., Sepulcre, J., Salinas-Alaman, A., Garcia-Layana, A., Murie-Fernandez, M., Bejarano, B. et al. (2008) Retinal nerve fiber layer atrophy is associated with physical and cognitive disability in multiple sclerosis. *Multiple Sclerosis*, 14, 906–912.
- Yap, T.E., Balendra, S.I., Almonte, M.T. & Cordeiro, M.F. (2019) Retinal correlates of neurological disorders. *Therapeutic Advances in Chronic Disease*, 10, 1–32.
- Yoo, Y., Tang, L.Y.W., Li, D.K.B., Metz, L., Kolind, S., Traboulsee, A.L. et al. (2019) Deep learning of brain lesion patterns and user-defined clinical and MRI features for predicting conversion to multiple sclerosis from clinically isolated syndrome. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 7, 250–259.
- You, Y., Barnett, M.H., Yiannikas, C., Parratt, J., Matthews, J., Graham, S.L. et al. (2020) Chronic demyelination exacerbates neuroaxonal loss in patients with MS with unilateral optic neuritis. *Neurology Neuroimmunology & Neuroinflammation*, 7, e700.
- Yperman, J., Becker, T., Valkenburg, D., Popescu, V., Hellings, N., Wijmeersch, B.V. et al. (2020) Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. *BMC Neurology*, 20, 1–15.
- Zhang, H., Alberts, E., Pongratz, V., Mühlau, M., Zimmer, C., Wiestler, B. et al. (2019) Predicting conversion from clinically isolated syndrome to multiple sclerosis—an imaging-based machine learning approach. *NeuroImage: Clinical*, 21, 101593.
- Zimmermann, H.G., Knier, B., Oberwahrenbrock, T., Behrens, J., Pfuhl, C., Aly, L. et al. (2018) Association of retinal ganglion cell layer thickness with future disease activity in patients with clinically isolated syndrome. *JAMA Neurology*, 75, 1071–1079.
- Zurita, M., Montalba, C., Labbé, T., Cruz, J.P., Dalboni da Rocha, J., Tejos, C. et al. (2018) Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. *NeuroImage: Clinical*, 20, 724–730.

How to cite this article: Montolío, A., Cegoñino, J., Garcia-Martin, E. & Pérez del Palomar, A. (2023) The macular retinal ganglion cell layer as a biomarker for diagnosis and prognosis in multiple sclerosis: A deep learning approach. *Acta Ophthalmologica*, 00, 1–13. Available from: <https://doi.org/10.1111/aos.15722>