Evaluation of machine learning algorithms and relevant biomarkers for the diagnosis of multiple sclerosis based on optical coherence tomography

Pablo García Mesa[1], Pilar Rojas Lozano[2], Nuria Díaz Gutiérrez[2], Manuel Cadena Santoyo[2], Alberto J. Beltrán Carrero[1], Juan J. Gómez-Valverde[1,3]

[1]Biomedical Image Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain.
[2]Instituto Provincial de Oftalmología, Hospital General Universitario Gregorio Marañón, Madrid, Spain.
[3]Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain.

Abstract

Multiple sclerosis (MS) is a prevalent neurodegenerative disease with significant visual pathway-related symptoms. Optical coherence tomography (OCT) has emerged as a valuable tool, and machine learning (ML) techniques hold promise for MS diagnosis. However, existing studies often lack comprehensive feature exploitation and require interpretable model analysis to improve clinical insights and diagnostic criteria. This study evaluates machine learning models for classification of healthy controls and MS patients using a comprehensive set of macular and optic-disc parameters from OCT imaging. The study included a dataset of 77 MS eyes and 54 control eyes, obtained by ophthalmic examination and OCT measurements from Optic Disc and Macular Cube scan protocols of a Cirrus HD-OCT 5000 (Carl Zeiss, Meditec, Dublin, CA, USA). Our results identified 19 features, validated by p-values ($p < 0.001$), as effective discriminators between MS patients and healthy controls. Patient-wise cross-validation is used to evaluate the performance of five ML algorithms. Gaussian Naive Bayes achieved the best AUC (87.9% ± 7.7%), while SHAP analysis reinforced the alignment with clinical observations of MS-related visual pathway changes and ganglion cell layer degeneration, with minimum ganglion cell thickness being the feature with the highest impact on classification. These findings underscore the potential of OCT-ML for early diagnosis and personalized treatment of MS.

1. Introduction

Multiple sclerosis (MS) is an autoimmune and neurodegenerative disease in which the myelin sheath surrounding nerve cells in the brain and spinal cord is damaged. This condition causes the improper transmission of nerve impulses, leading to various potential disabilities, with partial or total blindness, sensory loss, and motor disorders being the most common [1]. It is estimated that 2.8 million people worldwide are living with MS, with a higher prevalence of 140 cases per 100,000 population in Europe and the Americas [2].

Diagnosis and monitoring of MS rely on the integration of clinical, imaging and laboratory evidence, which may involve invasive procedures such as contrast-enhanced magnetic resonance imaging (MRI) and lumbar puncture [3]. Several important symptoms of MS are related to visual pathway disorders, such as: optic neuritis, diplopia and oscillopsia. The assessment of ophthalmic patients has been revolutionized by optical coherence tomography (OCT), a rapid and reproducible imaging technique that employs low-coherence interferometry to generate cross-sectional images of the retina and optic nerve head (ONH) [4]. In recent years, several studies have revealed the presence of biomarkers associated with MS as well as other neurodegenerative diseases such as Parkinson's and Alzheimer's in the retina and optic disc. These biomarkers include specific structures like the ganglion cell layer (GCL), the retinal nerve fiber layer (RNFL) of the optic disc, and the optic nerve head (ONH). [5, 6, 7]. Furthermore, with the increasing popularity of artificial intelligence (AI) techniques, several studies have employed parameters from OCT to train machine learning (ML) algorithms for the diagnosis of MS [8, 9, 10]. However, most studies do not use a wide range of features extracted from both the macular and optic-disc regions. Furthermore, an analysis of the interpretability of the resulting models is crucial to properly understand the importance of each feature. In addition, this type of in-depth analysis can help clinicians to gain insights into the disease and improve diagnostic criteria [11].

This study focuses on the evaluation of a set of ML models for the classification of healthy controls and patients with MS based on a comprehensive collection of macular and optic-disc parameters obtained through an OCT imaging device. The study is composed of several key steps. First, a feature selection process is performed using a ML approach to identify the most crucial features for discerning between the two groups. This feature selection process is further supported by a statistical analysis to ensure the selection of relevant attributes. Next, a set of ML models is trained using the selected features and their performances are compared. This step aims to determine the most effective model for accurately classifying healthy controls and MS patients based on the OCT data. Finally, to improve the interpretability of the models' predictions, an explainable AI technique known as Shapely Additive Explanations (SHAP) is used, which allows for a detailed analysis of the importance of each individual feature during the in-

ISBN: 978-84-17853-76-1

ference process. This step provides a rich breakdown of the models' decision criteria, which might be of great interest in clinical practice.

2. Materials

2.1. Dataset

This study included 40 patients with MS and 27 healthy controls recruited from the Neuro-ophthalmology department from Gregorio Marañón University Hospital in Madrid, Spain. The protocol was approved by the Ethics Committee of the hospital and all subjects were informed of the purpose and procedures of the study, and provided written consent to be included. The following inclusion criteria were applied: a refractive error within ± 5.00 dioptres for the equivalent sphere, and ± 2.00 dioptres for astigmatism. Exclusion criteria included prior intraocular surgery or any other conditions that could impact the visual field or nervous system, and the presence of pharmacological treatment that could affect visual function.

The participants received a complete ophthalmological examination and OCT measurements were obtained. The dataset included OCT volumes acquired using a Cirrus HD-OCT 5000 (Carl Zeiss, Meditec, Dublin, CA, USA), employing the Optic Disc Cube 200x200 and Macular Cube 512x128 scan protocols (6x6 mm). Only images with a signal-to-noise ratio of 7 or higher were included to ensure adequate acquisition quality, resulting in a final dataset comprising 77 MS eyes and 54 control eyes.

2.2. OCT Protocols

The Cirrus HD-OCT 5000 device provides a comprehensive evaluation of both retinal and optic-disc regions through a combination of quantitative and qualitative analyses. This study employs three distinct categories of analysis: Ganglion Cell Analysis (GCA), which assesses the condition of the GCL and the Inner Plexiform Layer (IPL); Macular Thickness Analysis (MTA), which measures the thickness of the macular region; and Optic-disc RNFL and ONH analyses. Figure 1 shows all the layers considered in this study. GCL and IPL are combined (GCIPL) in the GCA report of the device.

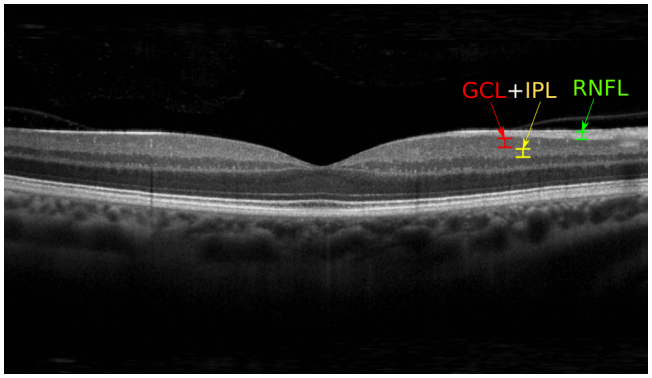


Figure 1: *Macular OCT image example of a healthy control indicating the most important retinal layers, according to the feature selection process: RNFL and GCIPL, which combines GCL and IPL.*

For GCA, the grid relies on six sector maps (Figure 2a), positioned at the fovea. The MTA analysis employs the early treatment diabetic retinopathy study (ETDRS) grid, automatically placed at the center of the fovea, effectively partitioning the macular area into nine distinct regions (Figure 2b). Lastly, the optic-disc analysis employs a clock-based grid (Figure 2c), dividing the area into twelve sectors and four regions corresponding to the anatomical locations (superior, temporal, inferior, nasal).

3. Methods

3.1. Feature selection

A feature selection process was performed to filter out the most relevant parameters and reduce overfitting. The aforementioned analyses (GCA, MTA, optic disc RNFL, and ONH) provided a total of 67 features. A tree-based method was used to remove the least relevant parameters, based on the feature importance values provided by the estimator [12]. More precisely, features that were assigned an importance value below the global mean feature importance were discarded. Significance analyses were performed using Student's t-tests to compare healthy controls and MS patients and validate the selected features.

3.2. Machine learning models and performance metrics

A comprehensive performance comparison of five different ML classifiers was conducted for this study. The selected models included Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Bagging Classifier, Random Forest, and Gaussian Naive Bayes (NB), similarly to other reference studies [8, 9, 10]. For each model, the default combination of hyperparameters given by [12] was employed. The most relevant ones are summarized below:

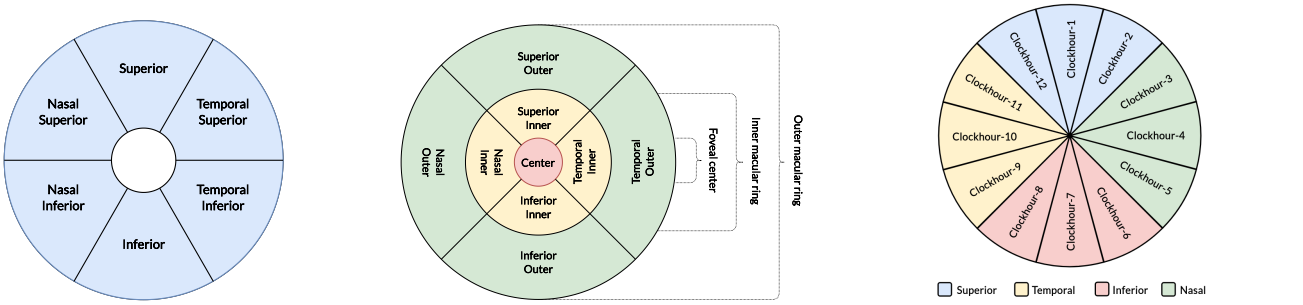
- **SVM**: $C = 1$, *kernel* = RBF, $\gamma = 1/(N_{feat} \cdot \sigma_X^2)$
- **k-NN**: *num. neighbours* = 5.
- **Random Forest**: *num. trees* = 100, *max. features* = $\sqrt{N_{feat}}$.
- **Bagging Classifier**: *estimator* = Decision Tree Classifier, *num. estimators* = 10.

Several metrics were calculated to evaluate their effectiveness, including accuracy, F1 score, and area under the curve (AUC) [13]. We divided the dataset into a 70/30 percent split, ensuring that both eyes from the same patient were grouped together within each split to maintain inter-eye correlations during model training. To assess the model's performance, we employed 5-fold cross-validation.

3.3. SHAP analysis

We used the SHAP method [14] to improve our understanding of the impact of each feature on the model. SHAP values, which are rooted in cooperative game theory, provide insight into the contribution of each

ISBN: 978-84-17853-76-1



(a) GCA. *(b) ETDRS for MTA* *(c) RNFL and ONH optic-disc analysis.*

Figure 2: *Sectors employed in the analysis of the Cirrus HD-OCT 5000 protocols.*

feature to the model's predictions. For the validation set of each cross-validation iteration, we computed the SHAP values for all features in the best model, and then averaged the results from the five iterations to derive an estimate of feature importance.

4. Results and discussion

Macula features
GCA analysis
Ganglion cells Nasal Superior thickness
Ganglion cells Nasal Inferior thickness
Ganglion cells Superior thickness
Ganglion cells Temporal Superior thickness
Ganglion cells Temporal Inferior thickness
Ganglion cells minimum thickness
Ganglion cells average thickness
Macular RNFL Nasal Inferior thickness
Macular RNFL Nasal Superior thickness
Macular RNFL Inferior thickness
Macular RNFL Superior thickness
Macular RNFL average thickness
MTA analysis
Macular cube average thickness - ILM to RPE

Table 1: *Selected macular features from GCA and MTA analysis. Ganglion cells refer to the combined layers GCL and IPL. All the presented features exhibit a high level of significance $p < 0.001$ in differentiating between MS patients and healthy controls.*

The feature selection step resulted in a final set of 19 features, which are presented in Tables 1 and 2. Notably, all of the p-values obtained from Student's t-tests of the selected features were below the standard significance threshold ($p < 0.05$), confirming that all of them are effective discriminators between the MS patients and healthy controls. In addition, previous studies [5, 6, 9] have consistently identified most of the selected parameters as important MS biomarkers, specifically the measurements related to the nasal, superior and inferior sectors of the macula and the temporal, superior and inferior sectors of the optic-disc.

Optic-disc features
ONH analysis
ONH RNFL Inferior Temporal thickness (Clockhour-7)
ONH RNFL Temporal central thickness (Clockhour-9)
ONH RNFL Temporal Superior thickness (Clockhour-10)
ONH RNFL Superior Temporal thickness (Clockhour-11)
ONH RNFL Temporal Quadrant thickness
ONH RNFL average thickness

Table 2: *Selected optic-disc features from RNFL and ONH analysis. All features have $p < 0.001$ significance between MS patients and healthy controls.*

The results corresponding to the performance metrics comparison of the ML models are shown in Table 3. The mean and the standard deviation for each metric are presented and the best metric is highlighted in bold. The Gaussian NB obtained the best AUC with $87.9\% \pm 7.7\%$, and SVM obtained the best accuracy with $87.3\% \pm 10.5\%$ and F1 score with $87.3\% \pm 10.5\%$.

Classifier	Accuracy %	F1 score %	AUC %
SVM	$\mathbf{87.3 \pm 10.5}$	$\mathbf{87.3 \pm 10.5}$	87.6 ± 10.2
k-NN	83.4 ± 9.4	83.4 ± 9.4	84.3 ± 9.7
Gaussian NB	87.2 ± 8.1	87.2 ± 8.0	$\mathbf{87.9 \pm 7.7}$
Random Forest	86.5 ± 10.2	86.5 ± 10.2	86.4 ± 10.4
Bagging Classifier	84.9 ± 5.8	84.8 ± 5.7	84.3 ± 5.4

Table 3: *Computed metrics for the ML classifiers after training using 5-fold cross validation. Mean and standard deviation are shown.*

Figure 3 illustrates the results of the SHAP analysis applied to the Gaussian NB model. The SHAP values confirm their correspondence with clinically observed variations in the visual pathway in individuals with MS. This

ISBN: 978-84-17853-76-1

alignment is particularly pronounced in the nasal sector of the macular region and the temporal sector of the optic disc, consistent with previous studies[5, 15].

In addition, our results are consistent with previous research showing progressive degeneration of the GCIPL in MS[15]. This progressive thinning of the GCIPL is underscored by the classifier's emphasis on five specific features closely associated with the nasal sectors of the GCIPL. Notably, the model places primary importance on the feature indicating minimum GCIPL thickness.

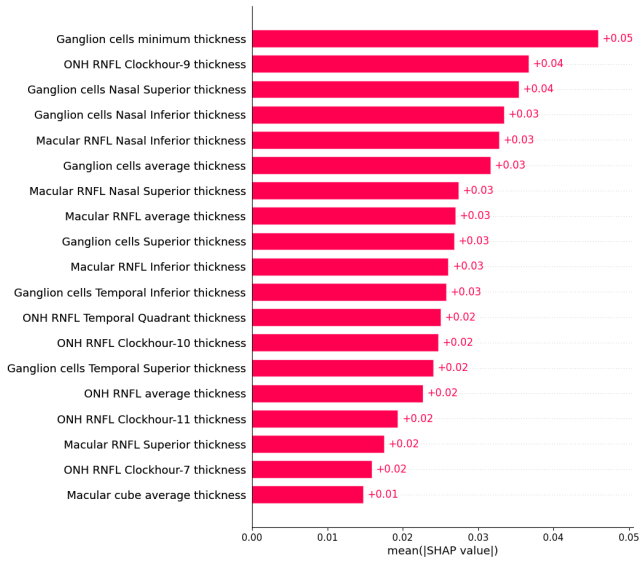


Figure 3: *Average SHAP values calculated across the 5-fold validation set for the Gaussian NB model.*

5. Conclusions

This study highlights OCT and ML's potential for enhancing MS diagnosis by evaluating comprehensive features from routine ophthalmologic exams. Our results reveal key features from GCA, MTA, and ONH analyses that significantly contribute to the highest-performing classification model (Gaussian NB) with an AUC of 87.9%. However, limitations include a single-center dataset with one OCT device, potentially affecting result generalizability. Future research should explore longitudinal studies to demonstrate OCT's value in patient follow-up and its potential integration into standard MS care criteria. Additionally, expanding the dataset to multiple centers and considering alternative modalities, such as OCT angiography images, are promising avenues for future investigation.

Acknowledgments

This research is part of the grant TED2021-131951B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the the "European Union NextGenerationEU/PRTR" and the grant "Primeros Proyetos" funded by the ETSI Telecomunicación (Universidad Politécnica de Madrid).

References

[1] Daniel S. Reich, Claudia F. Lucchinetti, and Peter A. Calabresi. "Multiple Sclerosis". In: *The New England Journal of Medicine* 378.2 (Jan. 2018), pp. 169–180. DOI: 10.1056/nejmra1401483.

[2] Clare Walton et al. "Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition". In: *Multiple Sclerosis Journal* 26.14 (Nov. 2020), pp. 1816–1821. DOI: 10.1177/1352458520970841.

[3] Marisa P McGinley, Carolyn H Goldschmidt, and Alexander D Rae-Grant. "Diagnosis and treatment of multiple sclerosis: a review". In: *Jama* 325.8 (2021), pp. 765–779.

[4] Neda Minakaran et al. "Optical coherence tomography (OCT) in neuro-ophthalmology". In: *Eye* 35.1 (Nov. 2020), pp. 17–32. DOI: 10.1038/s41433-020-01288-x.

[5] Friedemann Paul et al. "Optical coherence tomography in multiple sclerosis: A 3-year prospective multicenter study". In: *Annals of clinical and translational neurology* 8.12 (Nov. 2021), pp. 2235–2251. DOI: 10.1002/acn3.51473.

[6] Ricardo N. Alonso, Dolores González-Morón, and Orlando Garcea. "Optical coherence tomography as a biomarker of neurodegeneration in multiple sclerosis: A review". In: *Multiple sclerosis and related disorders* 22 (May 2018), pp. 77–82. DOI: 10.1016/j.msard.2018.03.007.

[7] Maria Satue et al. "Optical coherence tomography as a biomarker for diagnosis, progression, and prognosis of neurodegenerative diseases". In: *Journal of Ophthalmology* 2016 (2016).

[8] Elena Garcia-Martin et al. "Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier". In: *Computers in Biology and Medicine* 129 (2021), p. 104165.

[9] Alberto Montolío et al. "Comparison of machine learning methods using Spectralis OCT for diagnosis and disability progression prognosis in multiple sclerosis". In: *Annals of Biomedical Engineering* 50.5 (Feb. 2022), pp. 507–528. DOI: 10.1007/s10439-022-02930-3.

[10] Zahra Khodabandeh et al. "Discrimination of Multiple Sclerosis using multicenter OCT images". In: *Multiple Sclerosis and Related Disorders* (2023), p. 104846.

[11] Cristina González-Gonzalo et al. "Trustworthy AI: closing the gap between development and integration of AI systems in ophthalmic practice". In: *Progress in retinal and eye research* 90 (2022), p. 101034.

[12] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[13] Yasen Jiao and Pufeng Du. "Performance measures in evaluating machine learning based bioinformatics predictors for classifications". In: *Quantitative Biology* 4 (2016), pp. 320–330.

[14] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017.

[15] Ce Shi et al. "Visual function and disability are associated with focal thickness reduction of the ganglion cell-inner plexiform layer in patients with multiple sclerosis". In: *Investigative ophthalmology & visual science* 60.4 (2019), pp. 1213–1223.