Computers in Biology and Medicine 133 (2021) 104416

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography

Alberto Montolío [a,b], Alejandro Martín-Gallego [a,b], José Cegoñino [a,b], Elvira Orduna [c,d], Elisa Vilades [c,d], Elena Garcia-Martin [c,d], Amaya Pérez del Palomar [a,b,*]

[a] Group of Biomaterials, Aragon Institute of Engineering Research (I3A), University of Zaragoza, Zaragoza, Spain
[b] Department of Mechanical Engineering, University of Zaragoza, Zaragoza, Spain
[c] Ophthalmology Department, Miguel Servet University Hospital, Zaragoza, Spain
[d] GIMSO Research and Innovative Group, Aragon Institute for Health Research (IIS Aragon), Zaragoza, Spain

ARTICLE INFO

Keywords:
Multiple sclerosis
Machine learning
Optical coherence tomography
Retinal nerve fiber layer
Expanded disability status scale

ABSTRACT

Background: Multiple sclerosis (MS) is a neurodegenerative disease that affects the central nervous system, especially the brain, spinal cord, and optic nerve. Diagnosis of this disease is a very complex process and generally requires a lot of time. In addition, treatments are applied without any information on the disability course in each MS patient. For these two reasons, the objective of this study was to improve the MS diagnosis and predict the long-term course of disability in MS patients based on clinical data and retinal nerve fiber layer (RNFL) thickness, measured by optical coherence tomography (OCT).

Material and methods: A total of 104 healthy controls and 108 MS patients, 82 of whom had a 10-year follow-up, were enrolled. Classification algorithms such as multiple linear regression (MLR), support vector machines (SVM), decision tree (DT), k-nearest neighbours (k-NN), Naïve Bayes (NB), ensemble classifier (EC) and long short-term memory (LSTM) recurrent neural network were tested to develop two predictive models: MS diagnosis model and MS disability course prediction model.

Results: For MS diagnosis, the best result was obtained using EC (accuracy: 87.7%; sensitivity: 87.0%; specificity: 88.5%; precision: 88.7%; AUC: 0.8775). In line with this good performance, the accuracy was 85.4% using k-NN and 84.4% using SVM. And, for long-term prediction of MS disability course, LSTM recurrent neural network was the most appropriate classifier (accuracy: 81.7%; sensitivity: 81.1%; specificity: 82.2%; precision: 78.9%; AUC: 0.8165). The use of MLR, SVM and k-NN also showed a good performance (AUC $\geq$ 0.8).

Conclusions: This study demonstrated that machine learning techniques, using clinical and OCT data, can help establish an early diagnosis and predict the course of MS. This advance could help clinicians select more specific treatments for each MS patient. Therefore, our findings underscore the potential of RNFL thickness as a reliable MS biomarker.

1. Introduction

Multiple sclerosis (MS) is a neurodegenerative disease characterized by inflammation, demyelination, and axonal degeneration in the central nervous system (CNS) [1,2]. Diagnosis of MS is a slow and complicated process, since the most commonly used diagnostic approach is mainly based on excluding other diseases using paraclinical methods that are often prolonged, costly, and invasive [3].

While axonal loss is considered the main cause of functional disability in MS patients [4], detection of axonal damage is challenging. Magnetic resonance imaging (MRI) is used to detect inflammation and lesions [5], but is not sufficiently sensitive or specific to reveal the extent and severity of axonal damage [6]. The incorporation of cerebrospinal fluid (CSF) analysis into the McDonald criteria [3] for MS diagnosis increased diagnostic sensitivity but also decreased specificity and accuracy. Moreover, CSF analysis is highly invasive owing to the need for lumbar puncture to collect CSF samples. Another complementary test is evoked potential (EP) monitoring, which measures electrical activity induced in parts of the brain in response to light, sound and touch. The main disadvantages of EP are the length of time taken (about 2 h) and

* Corresponding author. Escuela de Ingeniería y Arquitectura. Campus Río Ebro, Edificio Betancourt. C/Maria de Luna s/n, 50018, Zaragoza, Spain.
E-mail address: amaya@unizar.es (A.P. Palomar).

https://doi.org/10.1016/j.compbiomed.2021.104416
Received 27 December 2020; Received in revised form 25 March 2021; Accepted 16 April 2021
Available online 26 April 2021

remaining doubts among clinicians as to whether the observed CNS alterations are specifically due to MS or to other conditions [3].

A faster alternative method for the diagnosis and monitoring of MS based on the assessment of axonal damage in the neuroretina has recently emerged [7–9]. The most recent optical coherence tomography (OCT) techniques allow specific measurement of the thickness of the retinal nerve fiber layer (RNFL) and even the ganglion cell layer (GCL) [10–12]. Thinning of these layers, which are not covered by myelin, constitutes a direct readout of axonal damage [7,13,14].

Studies have shown that RNFL thickness, measured by OCT, is a useful parameter to distinguish MS patients from healthy controls [4,15]. Garcia-Martin et al. [4,16] used an artificial neural network (ANN) in combination with OCT data to diagnose MS. Pérez del Palomar et al. [17] also demonstrated that machine learning techniques, specially decision tree (DT), combined with OCT data could be used to diagnose this disease. In the same vein, Cavaliere et al. [18] test the usefulness of support vector machine (SVM) in the MS diagnosis. Recently, another study compared the ability of SVM and feed-forward neural networks to detect MS using OCT recordings, and reported satisfactory results [19].

Next, the authors assessed the potential of machine learning techniques to predict the disease course in MS patients, which would help clinicians select patient-specific treatments [20]. Zhao et al. [21] evaluated the utility of SVM and logistic regression to predict the progression of MS-associated disability using brain MRI data acquired over 5 years. In another study, MRI data were also used to predict short-term disease progression using parallel convolutional pathways [22]. The work performed by Yperman et al. [23] analysed random forests and logistic regression to predict disability progression after 2 years using EP time series. Pinto et al. [24] trained classifiers such as linear regression, k-nearest neighbours (k-NN), DT and SVM to predict the progression of the disease severity after 6 and 10 years. To do that, these authors used data obtained by MRI, CSF analysis and EP. The study conducted by Seccia et al. [25] incorporated the Liquor analysis in testing different machine learning approaches to predict the course of MS in the short term.

It can be seen that most of the works limited the data to those obtained by means of the common tests included in McDonald criteria. Only a few authors have used the RNFL thickness to predict the disability course in MS patients, and they approached it from traditional statistical analysis. Rothman et al. [10] evaluated the capability of OCT data to predict the disability status 10 years later in 172 MS patients, applying linear regression models. Bsteh et al. [26,27] conducted a 3-year longitudinal study on 151 MS patients with RNFL measurements and tested also a linear regression model to predict the physical and cognitive disability progression. Following this work, these authors demonstrated that retinal thinning is associated with disability progression independent of relapse activity [28].

However, to date no studies have examined the use of machine learning techniques and RNFL data for long-term prediction of MS disease course. In this study, we tested several classification algorithms to develop a predictive model using OCT data obtained from 82 MS patients over 10 years of follow-up. Our goal was to verify RNFL thickness as MS biomarker and know in advance the progression of MS-associated disability using machine learning techniques instead of traditional statistical analysis.

2. Material and methods

2.1. Study population

Our cohort consisted of 108 MS patients and 104 healthy controls from the Miguel Servet University Hospital (Zaragoza, Spain) who underwent evaluations by ophthalmologists to perform OCT measurements and, in case of MS patients, by neurologists to determine their expanded disability status cale (EDSS) score. Moreover, a longitudinal study was carried out in 82 of these 108 MS patients who completed a total of seven visits: a baseline visit, 5 annual follow-up visits and a final visit at 10 years. From these 212 subjects of white European origin, one eye from each subject was randomly selected [29]. The diagnosis of MS was based on standard clinical and neuroimaging criteria (i.e. objective demonstration of lesions in both time and space) [30]. This study was performed in accordance with the Declaration of Helsinki and the protocol was approved by the Aragon Ethics Committee for Clinical Research (CEICA). All subjects provided written informed consent before inclusion in the study.

Table 1
Clinical characteristics of healthy controls and multiple sclerosis (MS) patients, and retinal nerve fiber layer (RNFL) thickness measured by Cirrus high definition optical coherence tomography (HD-OCT) at baseline visit. P-value, based on Student's *t*-test, shows the significance of the comparison between healthy controls and MS patients. Statistically significant differences ($p < 0.05$) are represented 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).

	Healthy controls (n = 104)	MS patients (n = 108)	P-value
Visit 0			
General parameters			
Age [years]	40.74 ± 13.13	43.16 ± 10.95	0.146
Sex (M – F)	35–69	34–74	
BCVA [Snellen]	0.84 ± 0.38	0.91 ± 0.28	0.127
OCT parameters – RNFL thickness			
Peripapillary thickness [μm]	100.16 ± 10.09	86.08 ± 16.10	**<0.001**
Superior thickness [μm]	125.63 ± 16.70	108.67 ± 20.84	**<0.001**
Nasal thickness [μm]	74.70 ± 15.95	71.69 ± 17.04	0.186
Inferior thickness [μm]	130.45 ± 17.69	109.99 ± 26.90	**<0.001**
Temporal thickness [μm]	69.51 ± 13.69	55.60 ± 14.82	**<0.001**
Foveal thickness [μm]	235.46 ± 42.26	271.94 ± 28.52	**<0.001**
MS parameters			
MS duration [years]	-	10.15 ± 8.30	
MS subtype (RRMS – SPMS – PPMS)	-	85–17–6	
Optic neuritis antecedent (Yes – No)	-	34–74	
Relapse in preceding year (Yes – No)	-	32–76	
EDSS score	-	2.65 ± 1.97	

The following inclusion criteria were applied: best-corrected visual acuity (BCVA) of 20/40 or higher; refractive error within ±5.00 diopters equivalent sphere and ±2.00 diopters astigmatism; and transparent ocular media (nuclear color/opalescence, cortical or posterior subcapsular lens opacity <1), according to the Lens Opacities Classification System III [31]. Exclusion criteria included prior intraocular surgery, diabetes or other diseases affecting the visual field or nervous system, and ongoing use of medications that could affect visual function.

Clinical data can be divided into general parameters and MS parameters. Age, sex and BCVA were considered as general parameters. For MS patients, MS duration, MS subtype (relapsing-remitting [RRMS], secondary-progressive [SPMS] and primary-progressive [PPMS]), optic neuritis antecedent, relapses in the preceding year and EDSS score were also recorded (see Table 1). These features, in combination with RNFL thicknesses, were the inputs of our predictive models, since they may significantly influence the classification decision [15,32].

2.2. OCT protocol

Cirrus HD-OCT (model 4000; Carl Zeiss Meditec, Dublin, CA, USA) is a digital retinal imaging technology used to monitor axonal loss in MS patients. Depending on the protocol used, RNFL thickness can be measured in different areas of the retina: the peripapillary area (around the optic nerve) or the macular area (around the fovea). Specifically, we used the Optic Disc Cube 200 × 200 study protocol, which generates

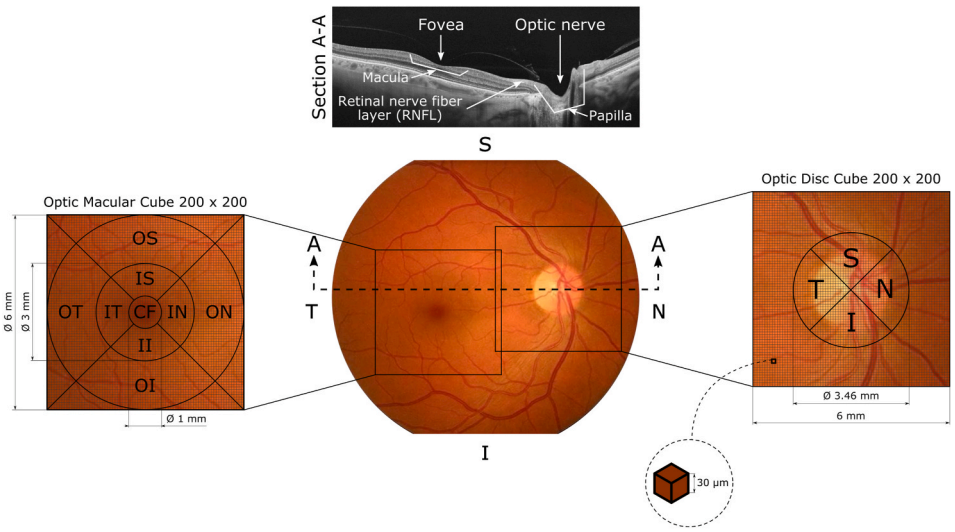

Fig. 1. Schematic representation of the protocols used to measure the retinal nerve fiber layer (RNFL) thickness. Middle: cross section of a right eye retina to indicate the location of the optical coherence tomography (OCT) protocols in the macula and in the peripapillary area. Left: Optic Macular cube 200 × 200 based on early treatment diabetic retinopathy study (ETDRS) grid to measure 9 macular areas (CF: central fovea; IS: inner superior; IN: inner nasal; II: inner inferior; IT: inner temporal; OS: outer superior; ON: outer nasal; OI: outer inferior; OT: outer temporal). Right: Optic Disc Cube 200 × 200 to analyse peripapillary area and 4 quadrants into which this area is divided (S: superior; N: nasal; I: inferior; T: temporal).

200 × 200 cube images enabling analysis of a 6 mm^3 volume around the optic nerve (Fig. 1). The OCT system automatically identifies the centre of the papilla and creates a circle-shaped sweep of 3.46 mm in diameter. This protocol calculates mean RNFL thickness by first determining the values corresponding to each of 4 quadrants into which the peripapillary area is divided (superior, nasal, inferior and temporal). In addition, we used the Macular Cube 200 × 200 study protocol, which scans an area of 6 × 6 mm covering the macula. As can be seen in Fig. 1, early treatment diabetic retinopathy study (ETDRS) grid is automatically centered on the fovea with fovea finder. With this grid, 9 macular areas can be measured (central fovea, inner superior, inner nasal, inner inferior, inner temporal, outer superior, outer nasal, outer inferior and outer temporal). However, for this study only foveal thickness was collected.

The image quality of the OCT device is based on the signal strength measurement, which combines the signal-to-noise ratio with the uniformity of the signal within a scan. The quality score ranges from 0 (poor) to 10 (excellent). Only images with a quality score of ≥7 were included in our analysis.

2.3. Machine learning

The performance of different classification algorithms for classifying between healthy controls and MS patients, and for predicting disability course of MS patients was tested. These algorithms were implemented using the Statistics and Machine Learning Toolbox and Deep Learning Toolbox in Matlab (version R2020b, Mathworks Inc., Natick, MA).

2.3.1. Data preprocessing

First, the dataset was divided into two subgroups: one to build the algorithm (training set) and another to validate it (validation set). The normalization used for numerical features begins with the training set normalization (mean of 0 and standard deviation of 1) and continues with the validation set normalization using mean and standard deviation from the training set. This procedure implies that algorithms do not have access to future information. Since machine learning algorithms need to work with numerical features, the categorical features (sex, MS subtype, optic neuritis antecedent and relapse in preceding year) were encoded into numerical values using one-hot encoding [33].

2.3.2. Feature selection

In machine learning, a rule of the thumb is to have a number of cases per class of at least ten times the number of features. Therefore, feature selection was used in order to reduce the risk of overfitting and increase the model interpretability [34]. To do this selection, two method were tested: sequential forward selection (SFS) and least absolute shrinkage and selection operator (LASSO).

SFS method attempts to minimize an objective function, the misclassification rate for classification models, over all feasible subsets of features. To do this, this sequential search algorithm adds features while evaluating the criterion until the addition of more features does not decrease the criterion. SFS is widely used for its simplicity and speed [35].

LASSO regression imposes a constraint on the model features that produces regression coefficients for some of these features to shrink toward zero taking into account the output of the model. Features with a regression coefficient equal to zero after the shrinkage process were removed from the dataset. By contrast, features with non-zero regression coefficients are strongly associated with the output [36,37].

2.3.3. Classifiers

We tested several algorithms to analyse which one works best with our data. For each model, the optimal hyperparameters of the different classifiers were determined by hyperparameter optimization. This optimization attempts to minimize the cross-validation loss by varying the hyperparameters.

2.3.3.1. Multiple linear regression. Multiple linear regression (MLR) is used to model the linear relationship between a dependent variable (response) and one or more independent variables (predictors). It is necessary that the independent variables are not highly correlated with each other. In case of two independent variables are highly correlated, only one of them should be used [38].

2.3.3.2. Support vector machine. Support vector machine is a robust binary classifier that calculates the optimal hyperplane separating different data classes [39]. This algorithm uses kernel functions to increase the dimension of original data, increasing the possibility of hyperplane separability for non-separable problems [40]. For separable classes, the optimal hyperplane maximizes the space that does not contain any observations. For non-separables classes, the objective is the same, but the algorithm imposes a penalty for each observation that is on the wrong side, which helps to prevent overfitting. This penalty factor is represented by C and is called box constraint. In a dual optimization problem, the Lagrange multipliers are bounded to be within the range [0, C], so C poses a box constraint. The kernel function was linear with a fixed value of kernel scale, the classifier divides all inputs by this value. Then, the algorithm applies the appropriate kernel norm to compute the Gram matrix. Using this procedure, SVM finds the

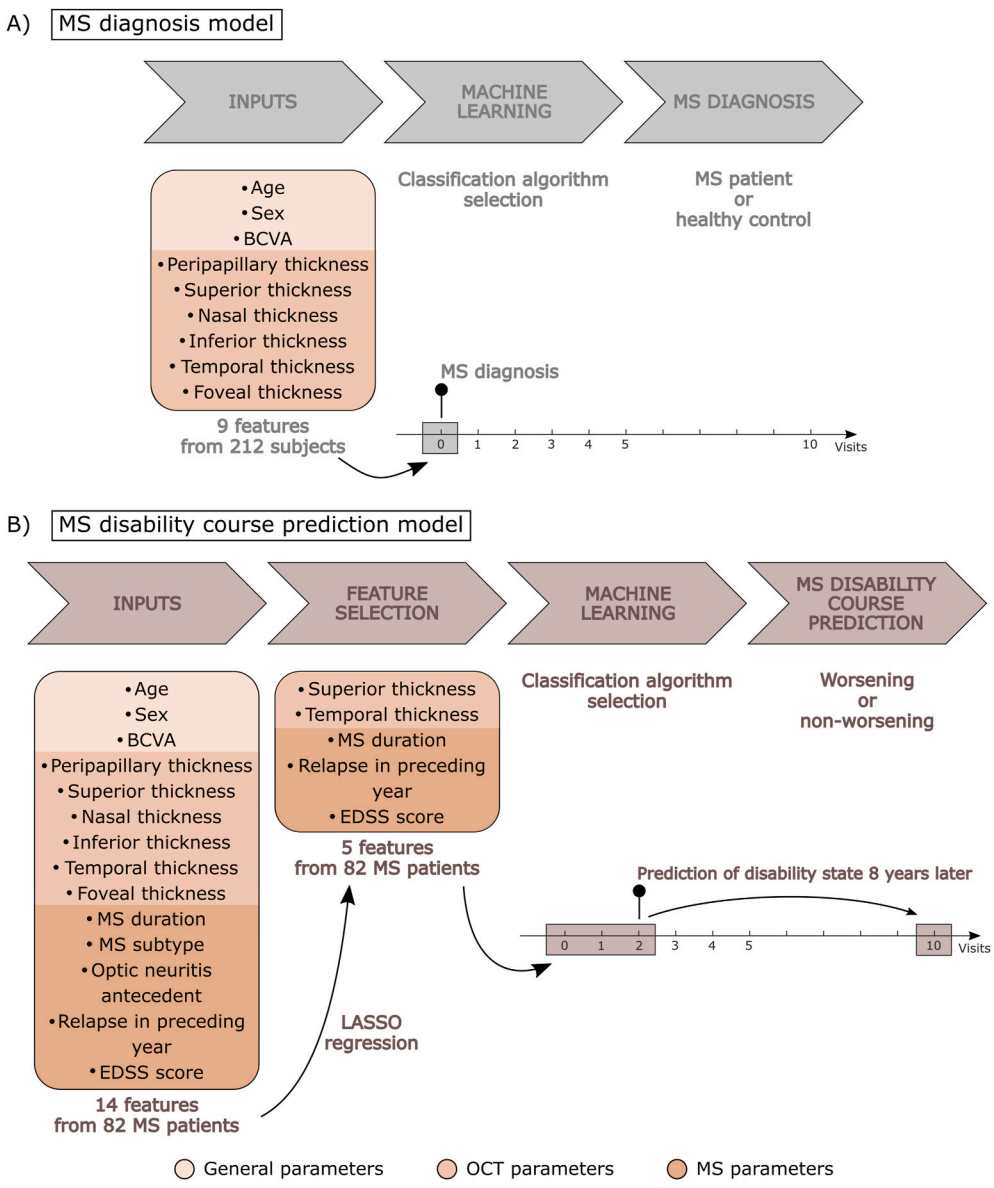

Fig. 2. Schematic diagram of the inputs and outputs of the predictive models. A) Model for diagnosis of multiple sclerosis (MS). The input data consists of 9 features from 108 MS patients and 104 healthy controls at the baseline visit. B) Model to predict the evolution of disability state. The input data consists of 5 features from 82 MS patients at the baseline visit and the first two annual follow-up visits. These 5 features were selected after applying a feature selection by least absolute shrinkage and selection operator (LASSO) regression algorithm.

optimal hyperplane in the transformed predictor space.

2.3.3.3. Decision tree. A decision tree classifier is composed of a root node, several internal nodes and several leaf nodes. The root node and the internal nodes include the test conditions of the features to distinguish between subjects having different qualities. The algorithm uses entropy to evaluate the degree of homogeneity of the sample. The construction of decision tree consists of finding a feature that returns the highest information gain [41]. As hyperparameters to control the depth of the trees, it is necessary to set the minimum number of leaf node observations and the minimum number of branch observations.

2.3.3.4. K-nearest neighbours. This algorithm classifies data according to the class of their nearest neighbours. K-NN classification consists of 2 stages: determination of the nearest neighbours and determination of the class based on those neighbours [42]. In this algorithm, the number of nearest neighbours and the distance metric between neighbours were the hyperparameters to optimize the structure. The number of neighbours it is very important because a very small number of these could be too sensitive to noise, which would increase the risk of overfitting.

2.3.3.5. Naïve Bayes. Naïve Bayes (NB) classifier applies density estimation to the data according to Bayesian theory, which assumes that predictors are conditionally independent [39,43]. There are several data distributions (kernel smoothing density estimation, multinomial, multivariate multinomial or normal) to model the data. One of the most used is kernel density estimation, this distribution is defined by a smoothing parameter called bandwidth. The bandwidth selection defines the smoothness of the density plot. A small bandwidth leads to undersmoothing and a huge bandwidth lead to oversmoothing. Therefore, it is preferable to choose a bandwidth as small as the dataset allows. However, there should be a balance between the bias of the estimator and its variance.

2.3.3.6. Ensemble classifier. Ensemble classifier (EC) is a combination of multiple learning algorithms to allow a more flexible structure, obtaining a better performance [44]. The EC used here was boosting with several classification trees, so that each individual model learns from mistakes made by the previous one. For this binary classification, LogitBoost was the ensemble aggregation algorithm for training the ensemble of boosted classification trees using adaptive logistic regression [45]. The optimization searched over the number of ensemble

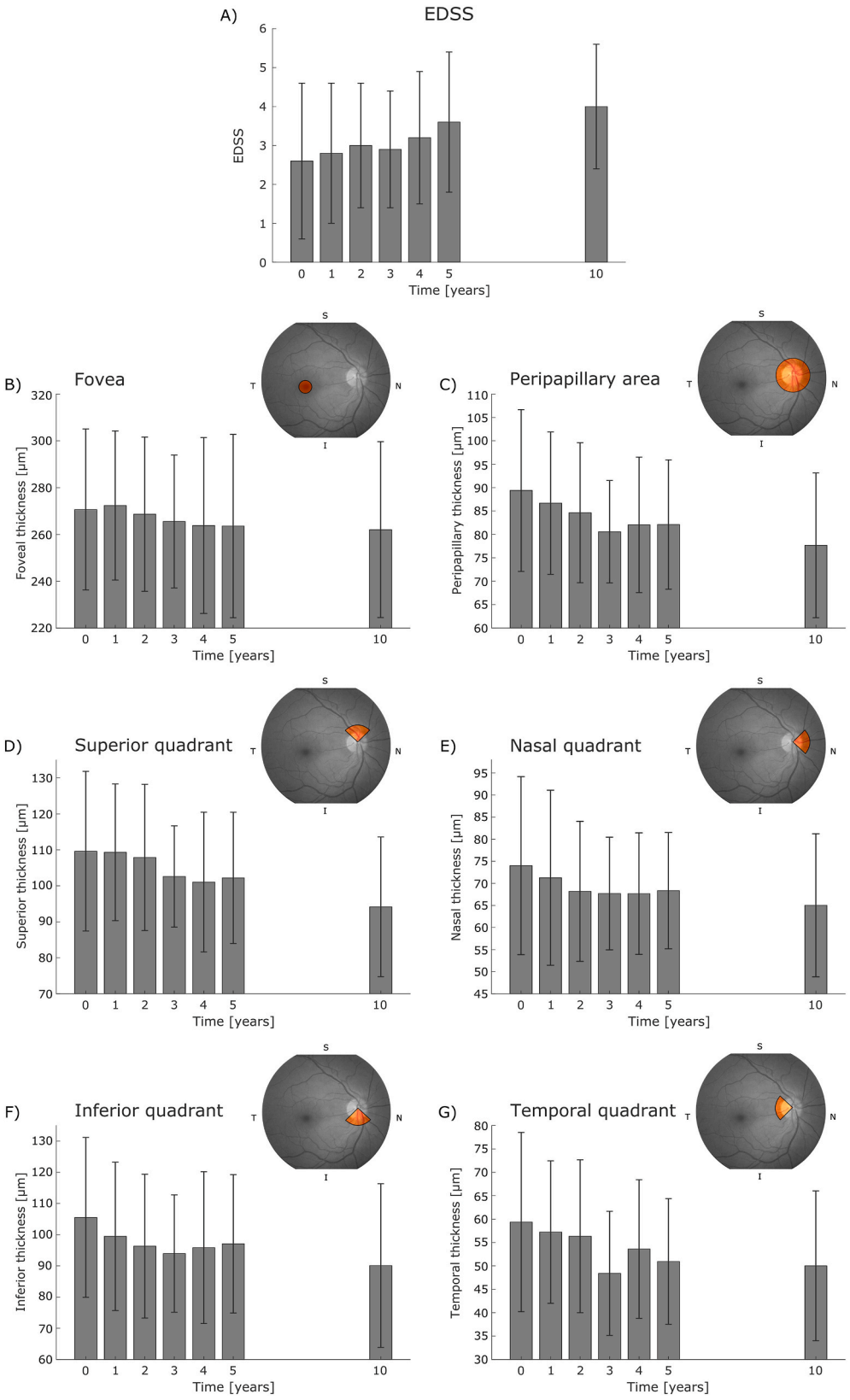

Fig. 3. A) Evolution of expanded disability status scale (EDSS) score from 82 patients with multiple sclerosis over 10 years of follow-up. B-G) Evolution of retinal nerve fiber layer (RNFL) thickness, which was analysed in the fovea and in the peripapillary area by optical coherence tomography (OCT), using Cirrus HD-OCT device, over the 10-year follow-up. Peripapillary area is divided into 4 quadrants (S: superior; N: nasal; I: inferior; T: temporal).

Table 2
Clinical characteristics and retinal nerve fiber layer (RNFL) measurements of both multiple sclerosis (MS) patient populations (non-worsening and worsening) at visits 0, 1 and 2. P-value, based on Student's *t*-test, shows the significance of the comparison between non-worsening patients and worsening patients. Statistically significant differences ($p < 0.05$) are represented in bold. (BCVA: best-corrected visual acuity; OCT: optical coherence tomography; RRMS: relapsing-remitting multiple sclerosis; SPMS: secondary-progressive multiple sclerosis; PPMS: primary-progressive multiple sclerosis; EDSS: expanded disability status scale).

	Non-worsening (n = 45)	Worsening (n = 37)	P-value	Non-worsening (n = 45)	Worsening (n = 37)	P-value	Non-worsening (n = 45)	Worsening (n = 37)	P-value
	Visit 0			Visit 1			Visit 2		
General parameters									
Age [years]	45.10 ± 11.01	39.16 ± 9.29	**0.011**	46.64 ± 10.98	40.38 ± 9.24	**0.007**	47.85 ± 11.50	41.50 ± 8.26	**0.006**
Sex (M – F)	15–30	12–25		15–30	12–25		15–30	12–25	
BCVA [Snellen]	0.91 ± 0.26	0.90 ± 0.29	0.920	0.89 ± 0.28	0.90 ± 0.30	0.944	0.89 ± 0.27	0.89 ± 0.29	0.984
OCT parameters – RNFL thickness									
Peripapillary thickness [μm]	84.88 ± 15.69	94.09 ± 16.05	**0.011**	80.81 ± 13.68	90.19 ± 14.32	**0.003**	77.46 ± 12.70	86.84 ± 12.89	**0.001**
Superior thickness [μm]	103.89 ± 22.69	116.78 ± 18.47	**0.007**	102.56 ± 18.09	114.68 ± 17.18	**0.003**	97.12 ± 18.05	112.24 ± 17.62	**<0.001**
Nasal thickness [μm]	70.24 ± 19.43	73.68 ± 17.54	0.408	65.46 ± 16.13	70.59 ± 19.20	0.192	62.91 ± 13.86	67.78 ± 14.34	0.123
Inferior thickness [μm]	113.16 ± 20.94	121.14 ± 26.99	0.136	105.96 ± 20.37	115.57 ± 23.92	0.053	101.62 ± 19.31	110.22 ± 21.22	0.059
Temporal thickness [μm]	53.71 ± 15.64	64.78 ± 16.75	**0.003**	49.36 ± 13.97	61.59 ± 12.46	**<0.001**	47.82 ± 13.85	56.41 ± 13.69	**0.006**
Foveal thickness [μm]	278.60 ± 30.49	279.86 ± 25.85	0.842	274.87 ± 23.61	280.81 ± 25.73	0.842	266.31 ± 23.88	268.84 ± 29.64	0.670
MS parameters									
MS duration [years]	12.98 ± 9.06	6.70 ± 5.72	**<0.001**	14.40 ± 9.14	7.82 ± 5.68	**<0.001**	15.62 ± 9.33	9.05 ± 5.71	**<0.001**
MS subtype (RRMS – SPMS – PPMS)	40–5–0	36–0–1		40–5–0	36–0–1		40–5–0	36–0–1	
Optic neuritis antecedent (Yes – No)	14–31	11–26		14–31	11–26		14–31	11–26	
Relapse in preceding year (Yes – No)	9–36	11–26		2–43	9–28		4–40	7–30	
EDSS score	3.52 ± 2.02	1.58 ± 1.27	**<0.001**	3.64 ± 2.03	1.65 ± 1.56	**<0.001**	4.04 ± 2.13	1.88 ± 1.68	**<0.001**

learning cycles, over the learning rate for shrinkage and over the minimum number of leaf node observations. The learning rate shrinks the contribution of each new classification tree that is added in the series. The output of this optimization was the EC with the minimum estimated cross-validation loss.

2.3.3.7. Long short-term memory. Long short-term memory (LSTM) is a supervised recurrent neural network, characterized by implementation memory, which can be used to classify sequential data [46]. A bidirectional LSTM layer learns bidirectional long-term dependencies between time steps of sequence data. These dependencies help the network learn from the complete time series at each time step. In this study, LSTM models were trained using back-propagation through time to predict the disability evolution using the time series of the 10-year follow-up. The architecture of the LSTM network consists of 5 layers. The network starts with a sequence input layer, which inputs the time series data into the network, followed by a bidirectional LSTM layer with several hidden units. The number of hidden layers corresponds to the information that is remembered between time steps. The network ends with a fully connected layer, a softmax layer and a classification output layer. The size of the input layer corresponds to the number of features of the input data. The size of the fully connected layer is the number of classes. The softmax layer is a function that converts a vector of real values into a vector of real values that sum to 1, so that they can be interpreted as probabilities. In this way, the scores are converted into a normalized probability distribution, which can be displayed to a user.

The stochastic gradient descent algorithm evaluates the gradient and updates the parameters at each iteration using a subset of the training data, called mini-batch. The complete passage of the training algorithm over the entire training set using mini-batches is an epoch. A sensitivity analysis was performed to determine the value of the adjustable parameters (hidden layers, epochs and mini-batch size). This sensitivity analysis consisted of performing several experiments to know whether the results change when something changes in the way the data analysis is approached. To determine the optimal value of each parameter, it was varied while the rest of the parameters were left fixed.

2.3.4. Cross-validation

Since our dataset was not very large, the use of hold-out validation would not be appropriate. As is well known in machine learning, the smaller the data, the higher the probability of overfitting. In this respect, cross-validation was used to minimize the risk of overfitting. Furthermore, k-fold cross-validation ensures that the final results were independent of the initial division [45]. This method consists of dividing the initial dataset into k groups, using k-1 groups to train the algorithm and the remaining group to validate it. The validation group changes k times and the final precision is determined by computing the mean precision of these k validation groups. Following the general consensus in the data mining community, 10-fold cross-validation was performed since this method constitutes an effective compromise between accuracy and computational cost [47].

2.3.5. Model performance

Confusion matrix and receiver operating characteristic (ROC) curve were generated in order to quantify the predictive performance of each model. In this way, accuracy, sensitivity, specificity and precision were determined. The area under the curve (AUC) was used to select the best machine learning algorithm. Classifier performance was quantified based on the AUC as follows: excellent (0.9–1), good (0.8–0.9), moderate (0.7–0.8) and fair (<0.7). Moreover, false positives (FP) and false negatives (FN) were analysed to identify which input data were more critical for a misclassification.

2.4. MS diagnosis model

This first model was developed to validate the use of RNFL thickness as MS biomarker. It can be said that our dataset is balanced since there are almost the same number of healthy controls and MS patients. Data from 108 MS patients and 104 healthy controls collected at the baseline visit (visit 0) were used. These data were general parameters and OCT

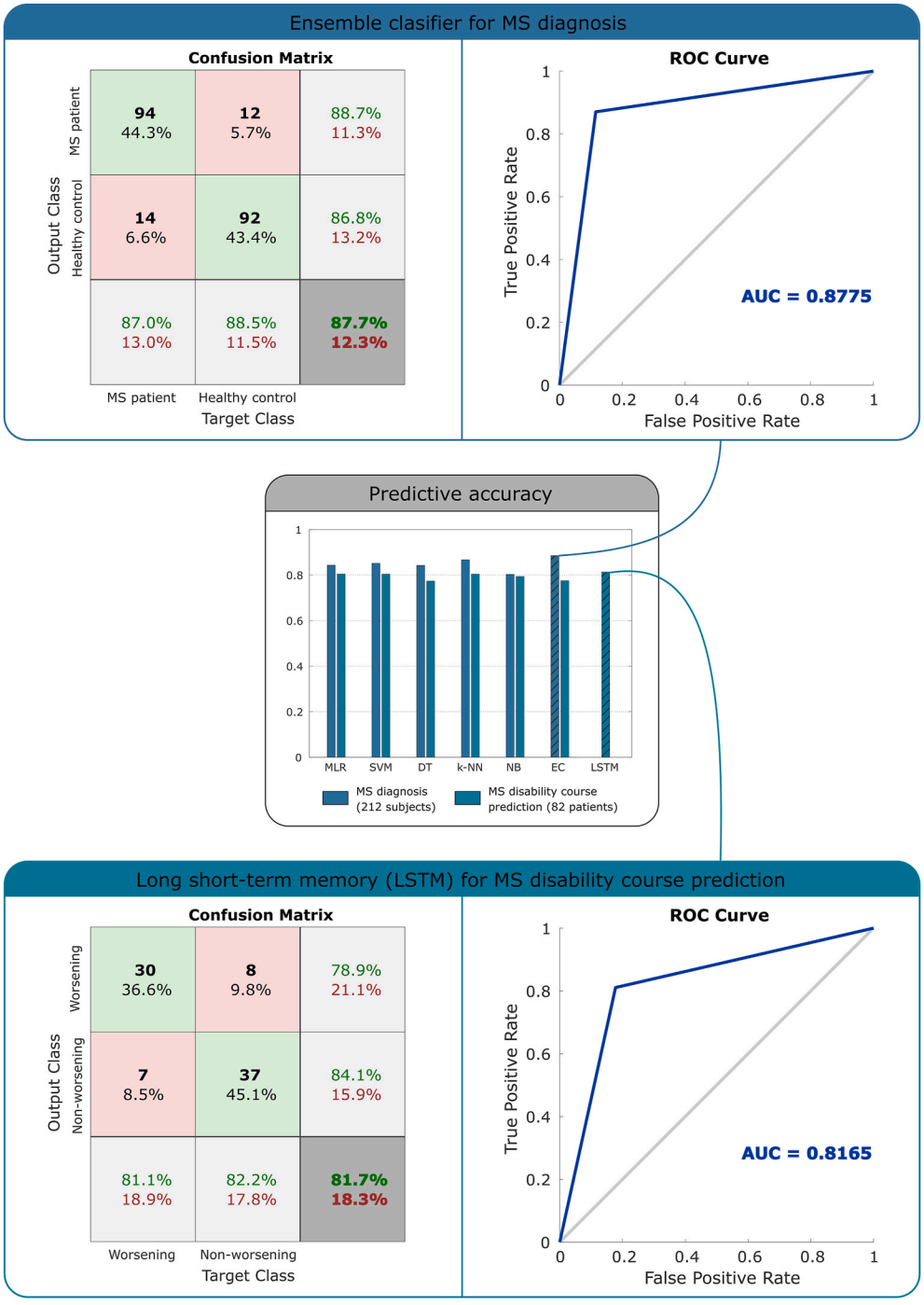

Fig. 4. Confusion matrix and receiver operating characteristic (ROC) curve, with area under curve (AUC), of the best algorithm for each predictive model. Middle: predictive accuracy for multiple sclerosis (MS) diagnosis model and MS disability course prediction model using 7 machine learning techniques: multiple linear regression (MLR), support vector machine (SVM), decision tree (DT), k-nearest neighbours (k-NN), Naïve Bayes (NB), ensemble classifier (EC) and long short-term memory (LSTM). The best algorithm for each model is highlighted by striped bars. LSTM was only tested for MS disability course prediction. Top: results using EC for MS diagnosis (accuracy: 87.7%; sensitivity: 87.0%; specificity: 88.5%; AUC: 0.8775). Bottom: results using LSTM for MS disability course prediction (accuracy: 81.7%; sensitivity: 81.1%; specificity: 82.2%; AUC: 0.8165).

parameters: age, sex, BCVA and the 6 RNFL thicknesses measured by OCT (Fig. 2A). Therefore, we characterized each subject using a total of 9 features. Comparison between data from MS patients and healthy controls was performed using Student's *t*-test (see Table 1). In this case, we had more than 90 subjects in each class, so feature selection was not necessary.

2.5. MS disability course prediction model

The next step was to develop a predictive model to assess the ability of RNFL thickness to accurately predict the progression of MS-associated disability in the long term. To do that, a 10-year longitudinal study was performed in 82 MS patients. Fig. 3 shows the evolution of EDDS score and RNFL data. An increase in EDSS score (Fig. 3A) and a decrease in RNFL thickness (Fig. 3B–G) were observed during follow-up. This model used data collected at the baseline visit (visit 0) and the first two annual follow-up visits (visits 1 and 2), and was designed to predict the disability course of MS patients at the final visit of the 10-year follow-up (i.e. 8 years later) (Fig. 2B). The features for this predictive model were general parameters, OCT parameters and MS parameters: age, sex, BCVA, the 6 RNFL thicknesses, MS duration, MS subtype, optic neuritis antecedent, relapse in preceding year and EDSS. Whereas age, sex, MS duration, MS subtype, optic neuritis antecedent were introduced only at baseline visit; BCVA, relapse in preceding year, EDSS score and the 6 RNFL thickness were taken into account at visits 0, 1 and 2.

In this prediction, the output was "worsening" or "non-worsening" in

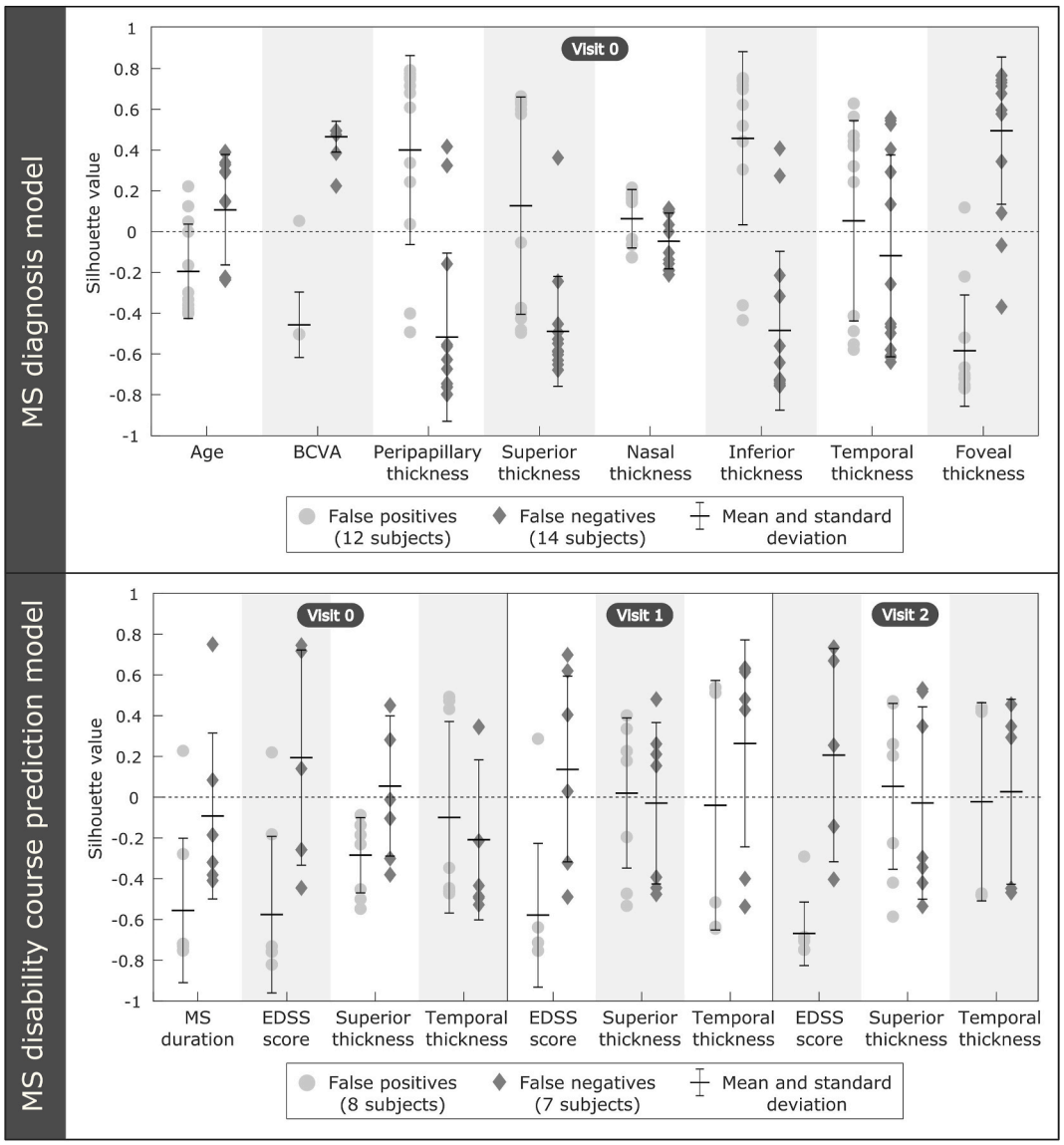

Fig. 5. Numerical feature analysis of false positives and false negatives for multiple sclerosis (MS) diagnosis model and MS disability course prediction model. The Silhouette value of each feature was calculated to compare false positives and false negatives with healthy controls and MS patients for MS diagnosis model, and with worsening and non-worsening for MS disability course prediction model (see Supplementary Material Fig. S1). Values close to +1 suggest that the subject is well matched to the assigned cluster, values close to 0 indicate that the subject coincides with the boundary between the two clusters and values close to −1 suggest that the subject may be assigned to the wrong cluster.

EDSS score at the target visit. We defined "worsening" as an increase of 1 or more points in EDSS between visit 2 and the final visit of the 10-year follow-up ($\Delta EDSS \geq 1$). And "non-worsening" represented patients whose EDSS value increased less than 1 point or decreased ($\Delta EDSS < 1$). In our dataset, 37 patients were worsening and 45 patients were non-worsening, so these classes were balanced. The statistical analysis of the input features of these two classes can be seen in Table 2, where significant differences ($p < 0.05$) are represented in bold.

As mentioned before, there were 14 inputs and we had less than 140 MS patients per class, so feature selection was necessary. Both SFS and LASSO were used to perform the feature selection and the resulting dataset after applying LASSO showed the best model performance (see Fig. S1). In this way, 5 features were chosen: MS duration, relapse in preceding year, EDSS score, temporal RNFL thickness and superior RNFL thickness (Fig. 2B). In this way, the reduced dataset was MS duration at baseline visit; relapse in preceding year, EDSS score, temporal RNFL thickness and superior RNFL thickness at visits 0, 1 and 2.

3. Results

As described above, 7 different classification algorithms were used to analyse the predictive accuracy of these two models. LSTM was only used for MS disability course prediction. While only the most relevant findings are discussed here, the complete results obtained for each classification algorithm are summarized in Fig. 4. The optimal value of the hyperparameters of each algorithm is shown in Table S1.

3.1. MS diagnosis model

The model for MS diagnosis was designed to classify subjects as MS patients or healthy controls based on age, sex, BCVA and RNFL measurements (9 features) acquired at the baseline visit. Whereas age, sex, BCVA and nasal thickness did not differ significantly between classes, the differences in peripapillary, superior, inferior, temporal and foveal thickness were significant (Table 1). The best result was obtained using EC, this algorithm correctly classified 186 out of 212 subjects (12 FP and

Table 3
Quantitative performance comparison of the best diagnostic methods for multiple sclerosis (MS) disease. Results of MS diagnosis using optical coherence tomography (OCT) from related works are compared. The following data are shown: OCT device, study population, classifier and model assessment. We indicated n/a when data were not available in the study. (ANN: artificial neural network; DT: decision tree; SVM: support vector machine; EC: ensemble classifier).

Study	Data type	MS patients	Healthy controls	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC (%)
MS diagnosis model									
Garcia-Martin et al. (2013) [4]	Spectralis OCT	106	115	ANN	n/a	n/a	n/a	n/a	94.5
Garcia-Martin et al. (2015) [16]	Spectralis OCT	112	105	ANN	88.5	89.3	87.6	88.5	n/a
Pérez del Palomar et al. (2019) [17]	SS-OCT Triton	80	180	DT	97.2	95.5	97.8	94.1	99.5
Cavaliere et al. (2019) [18]	SS-OCT Triton	48	48	SVM	90.6	89.8	91.5	91.7	97.0
Garcia-Martin et al. (2021) [19]	SS-OCT Triton	48	48	ANN	97.9	97.9	97.9	97.9	n/a
Our work	Cirrus HD-OCT	108	104	EC	87.7	87.0	88.5	88.7	87.8

14 FN, see confusion matrix in Fig. 4). Thus, the best accuracy for MS diagnosis was 87.7% (sensitivity: 87.0%; specificity: 88.5%; precision: 88.7%; AUC: 0.8775). The optimal hyperparameters of EC were: 100 classification trees, 0.196 learning rate and 1 minimum observation for leaf nodes.

The k-NN algorithm showed an accuracy of 85.4% with the use of 7 nearest neighbours and the Euclidean distance as distance metric between them (sensitivity: 83.3%; specificity: 87.5%; precision: 87.4%; AUC: 0.8542). With SVM the accuracy was 84.4% (sensitivity: 85.2%; specificity: 83.7%; precision: 84.4%; AUC: 0.8442). After optimizing the hyperparameters, box constraint was set to 1.752 and kernel scale to 0.055.

3.2. MS disability course prediction model

In this model, we evaluated the ability of the model to predict whether the degree of disability of a given MS patient would worsen or not. This predictive model used data from 82 MS patients acquired at 3 consecutive visits (baseline and annual follow-up visits 1 and 2) to predict the disability course of each patient at year 10. It can be seen that the reduced dataset, after the application of LASSO regression, corresponds to the input features whose difference was the most significant ($p < 0.001$) (see Table 2). The predictions generated by LSTM neural network were correct in 67 of 82 cases (8 FP and 7 FN, see confusion matrix in Fig. 4) giving an accuracy of 81.7% (sensitivity: 81.1%; specificity: 82.2%; precision: 78.9%; AUC: 0.8165). This result was obtained with the following structure: 30 hidden layers, 30 epochs and a mini-batch size of 20. Regarding hidden layers, the performance increased as the number of hidden layers increased up to 30 layers, from that point the performance was constant. For epochs and mini-batch size, the performance increased until reaching 30 epochs and a size of 20, and then the performance started to decrease.

SVM, MLR and k-NN also showed a good performance, correctly classifying 66 out of 82 MS patients. The optimal SVM structure was a box constraint of 2.120 and a kernel scale of 2.610 (accuracy: 80.5%; sensitivity: 83.8%; specificity: 77.8%; precision: 75.6%; AUC: 0.8078). In the case of k-NN, the hyperparameter optimization showed 13 neighbours as the optimal number of nearest neighbours and standardized Euclidean distance as the distance metric between neighbours (accuracy: 80.5%; sensitivity: 75.7%; specificity: 84.4%; precision: 80.0%; AUC: 0.8006). As can be seen in Fig. 4, the same accuracy was obtained using MLR (accuracy: 80.5%; sensitivity: 78.4%; specificity: 82.2%; precision: 78.4%; AUC: 0.8030). The rest of the classifiers showed moderate performance (AUC < 0.8).

3.3. Misclassified subjects

In order to understand why some subjects were misclassified, the features of each individual subject were analysed. Each input feature of a specific subject was studied to check if its value was more related to one class or the other, i.e. healthy control or MS patient for the first model and non-worsening or worsening for the second one. To do that, Silhouette value of each numerical input variable was calculated to compare the 12 FP and 14 FN with healthy controls and MS patients in case of MS diagnosis model, and the 8 FP and the 7 FN with worsening and non-worsening for MS disability course prediction. The Silhouette method measures how close each data point is to its own cluster compared to other clusters, this measure ranges from -1 to $+1$ [48]. A Silhouette value near $+1$ indicates that the data point is well matched to its own cluster and poorly matched to neighbouring clusters, a value of 0 suggests that the data point is on the decision boundary between two neighbouring clusters and negative values indicate that those data points might have been assigned to the wrong cluster. The Silhouette value of each numerical input variable, for each visit used in our predictive models, of FP and FN is available in the Supplementary Material. These data are summarized in Fig. 5.

The features whose mean Silhouette value was negative could cause a misclassification in our models because their values would be more related to the opposite group. First, paying attention to MS diagnosis model, the features with lowest Silhouettes values were: BCVA and foveal thickness for FP, and peripapillary, superior and inferior thickness for FN. Second, in the MS course prediction model, MS duration and EDSS score had values near -1 for FP. However, there are no features with values close to -1 for FN.

4. Discussion

Machine learning is based on the use of mathematical algorithms that seek to match input and output data, and is emerging as a very useful tool in many scientific fields. These algorithms can be used to improve the diagnosis of certain diseases and to aid clinicians in selecting the most appropriate therapy for a given patient [18,41,49]. More specifically, classification algorithms can use clinical data to improve diagnosis and predict the disease progression in MS patients [17,23]. In the present study, we evaluate the utility of machine learning techniques to diagnose MS disease and provide long-term predictions of the degree of disability in MS patients based on clinical data and RNFL thickness measurements acquired by OCT. This imaging technique offers certain advantages over other techniques, such as MRI, since it provides a rapid, cost-effective, and non-invasive means to evaluate RNFL thickness [17].

It is known that RNFL thickness is a very useful parameter to diagnose this disease [50] and, with our results, this fact has been corroborated, establishing this layer of the retina as a MS biomarker [51]. In addition, the significance of each RNFL thickness was analysed to find out which are the most important areas for MS diagnosis. The difference of RNFL thickness between MS patients and healthy controls was statistically significant in peripapillary area where superior, inferior

Table 4
Quantitative performance comparison of the best methods for disability course prediction in multiple sclerosis (MS) patients. Data type indicates the source of the data used. Data from represents the data points used to train the predictive model and prediction time is the time from the last data point to the output time. The studies had different output types, these are based on the variation of the expanded disability status scale (ΔEDSS) or on the evolution from relapse-remitting (RR) to secondary-progressive (SP) form of the disease. Study population, classifier and model assessment are also shown. We indicated n/a when data were not available in the study. (MRI: magnetic resonance imaging; EP: evoked potential; CSF: cerebrospinal fluid; OCT: optical coherence tomography; SVM: support vector machine; CNN: convolutional neural network; RF: random forest; LSTM: long short-term memory).

Study	Data type	Data from	Pred. time	Output type (positives vs negatives)	Positive patients	Negative patients	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC (%)
MS disability course prediction model												
Zhao et al. (2017) [21]	MRI	3 years	2 years	ΔEDSS $\geq$ 1.5 vs ΔEDSS $<$ 1.5	212	362	SVM	71.0	65.0	74.0	n/a	n/a
Tousignant et al. (2019) [22]	MRI	1 year	1 year	ΔEDSS $\geq$ 1.5/1/0.5 vs ΔEDSS $<$ 1.5/1/0.5	103	980	CNN	n/a	n/a	n/a	n/a	70.1
Yperman et al. (2020) [23]	EP	2 years	2 years	ΔEDSS $\geq$ 1/0.5 vs ΔEDSS $<$ 1/0.5	46	373	RF	n/a	n/a	n/a	n/a	75.0
Pinto et al. (2020) [24]	MRI CSF EP	2 years	4 years	RR to SP vs otherwise	21	166	SVM	n/a	76.0	0.77	n/a	86.0
	MRI CSF EP	2 years	4 years	EDSS $\geq$ 3 vs EDSS $<$ 3	38	107	SVM	n/a	84.0	81.0	n/a	89.0
Seccia et al. (2020) [25]	MRI Liquor	1 year	2 years	RR to SP vs otherwise	1168	207	SVM	87.8	77.3	87.9	9.3	n/a
	MRI Liquor	6 years	2 years	RR to SP vs otherwise	1168	207	LSTM	98.0	67.3	98.5	42.7	n/a
Our work	OCT	3 years	8 years	ΔEDSS $\geq$ 1 vs ΔEDSS $<$ 1	37	45	LSTM	81.7	81.1	82.2	78.9	81.7

and temporal quadrants were also significant, highlighting the temporal quadrant. This difference turned out to be significant in the fovea, corroborating the importance of macular area [17].

Our results are totally in accordance with other works [52–54]. These studies establish that, although axonal loss diffusely affects RNFL thickness throughout the peripapillary area, the temporal quadrant is the most affected area in MS patients from early stages of the disease, even without previous optic neuritis episodes. Some authors suggest that, because the temporal quadrant of the optic disc is relatively thinner than other quadrants, atrophy is noted from earlier stages [52]. Other authors, including our group, suggest that this temporal thinning reflects a predominant involvement of the fovea and the papillomacular bundle, which is the structure that transmits information from the fovea. The fovea is the central macular structure and is primarily responsible for detailed visual and color functions, which are affected in MS. Therefore, a reduction in macular volume is also observed in these patients, especially in the nasal sectors of the macula, which correspond to the papillomacular bundle [55].

As can be seen in Table 3, RNFL thickness, measured by different OCT devices, has been evaluated to train classifiers for MS diagnosis. Garcia-Martin et al. used Spectralis OCT in combination with ANN, obtaining an AUC of 94.5% [4] and an accuracy of 88.5% [16]. Other studies employed SS-OCT Triton to measure the RNFL thickness and the results were: an accuracy of 97.2% with DT [17], 90.6% with SVM [18] and 97.9% with ANN [19]. Whereas we have used RNFL thickness measured by Cirrus HD-OCT and the best classifier was EC with an accuracy of 87.7%. The results of these studies could indicate that the best OCT device for MS diagnosis, using machine learning techniques, would be SS-OCT Triton. However, the study populations with this OCT device were either too small or imbalanced. Therefore, a more significant result would be an accuracy of 88.5%, obtained by Garcia-Martin et al. [16] with ANN, very similar to ours.

Previous studies [7,26,56] have reported associations between RNFL thickness and MS disability course, suggesting that the retina may also be a reliable biomarker of the progression of this neurodegenerative disease. However, it was not possible to establish a direct relationship between disability and RNFL thickness. First, RNFL thinning is not caused exclusively by MS, but is also affected by aging [57–59]. Second, there is marked variation in RNFL thickness among MS patients [7,27,60]. For these reasons, we examined the potential utility of machine learning techniques to predict the progression of disability in MS patients.

In this study, for MS disability course prediction, it has been distinguished between two classes (worsening versus non-worsening). As can be seen in Table 2, input features such as age, MS duration, EDSS, peripapillary, superior and temporal thicknesses were significant ($p < 0.05$). However, after feature selection by LASSO, only the most significant features were used ($p < 0.001$). These results are in line with the findings of the work by Birkeldh et al. [54] in which it was shown that the temporal quadrant was highly associated with the EDSS score. Higher EDSS values correspond to a significant RNFL reduction in peripapillary, superior and temporal thicknesses for non-worsening patients. The reason for that was well explained in our previous study [7], MS patients with greater disability have a lower RNFL thickness because the most part of this thinning occurs during the initial stages of the disease (before the onset of significant disability).

Several studies have assessed the ability of machine learning techniques to predict the short-term progression of MS. These studies used datasets with information from MRI, EP or CSF analysis and tested different classifiers to predict the disability progression in MS patients (see Table 4). However, none of these studies has evaluated the use of RNFL thickness to predict the long-term progression of MS-associated disability. Therefore, our study is the first to apply machine learning techniques to OCT data obtained from a 10-year longitudinal study. The progression was based on the variation of EDSS score in the majority of the works [21–24], although some studies also predicted the progression or not from RRMS to SPMS form of the disease [24,25]. The time from the last data point used to train the predictive model to the output time

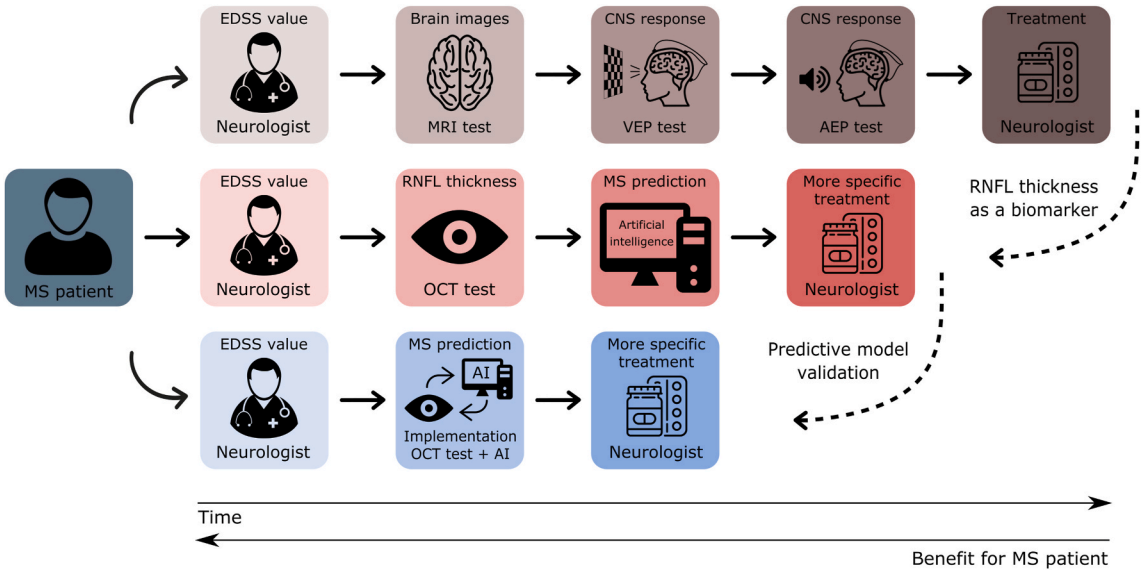

Fig. 6. Flowchart of the treatment process for MS patients. The top row shows the methodology currently implemented in most hospitals: magnetic resonance imaging (MRI) to detect changes in brain volume; visual evoked potential (VEP) and auditory evoked potential (AEP) tests to analyse central nervous system (CNS) response. The middle row shows our proposed method, based on the use of retinal nerve fiber layer (RNFL) thickness together with artificial intelligence (AI) to predict the progression of MS-associated disability and thus to be able to apply a more specific treatment. The bottom row depicts the proposed method following incorporation of classification algorithms into optical coherence tomography (OCT) devices after prior validation of our predictive models. In all cases, a neurologist must establish the disability state using the expanded disability status scale (EDSS).

was usually 1 or 2 years, but Pinto et al. predicted the disability course 4 years later with an AUC of 89% using SVM [24]. Seccia et al. tested LSTM trained with 6 data points to predict the disability course 2 year later obtaining a high accuracy and specificity but low sensitivity and precision because their data were class-imbalanced [25]. To obtain a significant result, several factors should be taken into account. First, it is necessary a balance between the data points used for training and the prediction time since it is desirable to achieve a good prediction with the lowest number of progression years. Secondly, the dataset should be balanced to obtain a good performance in all parameters of the predictive model evaluation. For these reasons, we have used 3 data points from a class-balanced dataset to predict the disability progression 8 years later with an accuracy of 81.7% using LSTM.

While our findings constitute an important step towards patient-specific prediction of the course of MS, our results should be interpreted considering several limitations. Our predictive models are largely based on OCT data. However, the use of OCT devices in combination with other techniques such as MRI, EP or CSF analysis could improve the model performance. Given that the quality of OCT data directly influences the results, the precision of the OCT device must also be considered. Moreover, these data should be used in combination with clinical data, such as the EDSS score. Although EDSS score is considered the most useful tool for measurement of MS progression and is assumed by many neurologists to directly determine worsening/non-worsening, this variable has some drawbacks in terms of reliability and sensitivity [61]. Our progression prediction was defined as a variation in EDSS score (ΔEDSS), so it is a qualitative prediction because the value of the disability state is not predicted. Besides, it should be borne in mind that the treatment regimens of MS patients, which were not considered when developing our models, could potentially influence disease course.

The heterogeneity of our sample population should also be noted. Patients began the follow-up at different stages of disease progression and, therefore, there was considerable variation among patients in MS duration at baseline and consequently in EDSS score. However, since MS progression in each patient was measured relative to their baseline EDSS score, this variation would not affect our results. Furthermore, although our study population included patients with all subtypes of this disease, RRMS clearly predominated, as shown in Table 1.

Taking into account the numerical input features used in the models, we analysed the reasons why some subjects could have been incorrectly classified. As can be seen in Fig. 5, for MS diagnosis model, BCVA and foveal thickness for FP; and peripapillary, superior and inferior thickness for FN had values closer to the wrong class than the correct one (i.e. Silhouette values near −1), and this fact could confuse algorithms. In case of MS disability course prediction, MS duration and EDSS score were the most confusing variables for FP, while there are no features with mean Silhouette value near −1 for FN. These results could indicate that subjects with values closer to those of the opposite class in some significant variable could be misclassified. Machine learning techniques have great potential to improve the diagnosis and treatment of MS, but it is important that they are used in conjunction with large datasets that appropriately represent all cases.

Although more longitudinal studies of MS patients from other parts of the world will be required to validate the predictive models presented here, our results indicate that RNFL thickness measured by OCT is a useful biomarker to establish an early diagnosis and predict the progression of MS, in line with the findings of previous studies [62,63]. For this reason, we support the proposal of several authors to review the approach for MS diagnosis [64]. Currently, MS diagnosis requires a large number of tests, including MRI to detect MS lesions in the brain tissue, the analysis of CSF samples and multiple EP tests to detect possible decreases in the rate of neurotransmission in the CNS. Visual evoked potential (VEP) and auditory evoked potential (AEP) are measures of CNS activity in response to light and acoustic stimuli, respectively. Another less commonly used test is somatosensory evoked potential (SSEP) testing, which is used to detect deficits in electrical activity in the spinal cord by applying electrical stimuli to the wrist, the back of the knee, or other locations. These all procedures are time-consuming.

The same happens to design an appropriate treatment for a MS patient. First, a neurologist evaluates the disability state by establishing the EDSS value. Then, MRI and EP tests are performed to detect lesions in the CNS. Thus, the neurologist prescribes a treatment without any prior information of the MS progression for that specific patient. In this work, we propose that OCT data should be used in combination with artificial intelligence to MS diagnose and predict MS course (see Fig. 6). OCT is an objective and reproducible test that can be completed in as

little as 2 min by non-specialized personnel without exposing the patient to radiation or causing any discomfort [18]. Moreover, OCT devices are available in practically any hospital and even in some outpatient centres.

This would be of significant benefit to clinicians, who could select more specific and appropriate treatments based on the predicted disease course for a given patient. Moreover, as can be seen in Fig. 6, this approach would help to reduce waiting times and improve the overall quality of life of MS patients. To make this possible, it will be necessary to validate the predictive models presented here in other populations in order to incorporate these algorithms into OCT devices, improving the diagnosis and disability prediction of MS.

Declaration of competing interest

The authors state that there are no conflicts of interest.

Acknowledgements

This work 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 Instituto de Salud Carlos III (PI17/01726).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104416.

References

[1] B. Ferguson, M.K. Matyszak, M.M. Esiri, V.H. Perry, Axonal damage in acute multiple sclerosis lesions, Brain 120 (1997) 393–399, https://doi.org/10.1093/brain/120.3.393.

[2] Y. You, M.H. Barnett, C. Yiannikas, J. Parratt, J. Matthews, S.L. Graham, A. Klistorner, Chronic demyelination exacerbates neuroaxonal loss in patients with MS with unilateral optic neuritis, Neurol. Neuroimmunol. Neuroinflammation. 7 (2020), https://doi.org/10.1212/NXI.0000000000000700.

[3] C.H. Polman, S.C. Reingold, G. Edan, M. Filippi, H.-P. Hartung, L. Kappos, F. D. Lublin, L.M. Metz, H.F. McFarland, P.W. O'Connor, M. Sandberg-Wollheim, A. J. Thompson, B.G. Weinshenker, J.S. Wolinsky, Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria," Ann, Neurol. 58 (2005) 840–846, https://doi.org/10.1002/ana.20703.

[4] E. Garcia-Martin, L.E. Pablo, R. Herrero, J.R. Ara, J. Martin, J.M. Larrosa, V. Polo, J. Garcia-Feijoo, J. Fernandez, Neural networks to identify multiple sclerosis with optical coherence tomography, Acta Ophthalmol. 91 (2013) e628–e634, https://doi.org/10.1111/aos.12156.

[5] S. Roy, D. Bhattacharyya, S.K. Bandyopadhyay, T.H. Kim, An effective method for computerized prediction and segmentation of multiple sclerosis lesions in brain MRI, Comput. Methods Progr. Biomed. 140 (2017) 307–320, https://doi.org/10.1016/j.cmpb.2017.01.003.

[6] Y. Fu, T.M. Talavage, J.-X. Cheng, New imaging techniques in the diagnosis of multiple sclerosis, Expert Opin. Med. Diagn. 2 (2008) 1055–1065, https://doi.org/10.1517/17530050802361161.

[7] A. Montolío, J. Cegoñino, E. Orduna, B. Sebastian, E. Garcia-Martin, A. Pérez del Palomar, A mathematical model to predict the evolution of retinal nerve fiber layer thinning in multiple sclerosis patients, Comput. Biol. Med. 111 (2019) 103357, https://doi.org/10.1016/j.compbiomed.2019.103357.

[8] L.J. Balk, A. Cruz-Herranz, P. Albrecht, S. Arnow, J.M. Gelfand, P. Tewarie, J. Killestein, B.M.J. Uitdehaag, A. Petzold, A.J. Green, Timing of retinal neuronal and axonal loss in MS: a longitudinal OCT study, J. Neurol. 263 (2016) 1323–1331, https://doi.org/10.1007/s00415-016-8127-y.

[9] F. Eslami, M. Ghiasian, E. Khanlarzade, E. Moradi, 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 (2020) 15–23, https://doi.org/10.2147/EB.S229814.

[10] A. Rothman, O.C. Murphy, K.C. Fitzgerald, J. Button, E. Gordon-Lipkin, J. N. Ratchford, S.D. Newsome, E.M. Mowry, E.S. Sotirchos, S.B. Syc-Mazurek, J. Nguyen, N.G. Caldito, L.J. Balcer, E.M. Frohman, T.C. Frohman, D.S. Reich, C. Crainiceanu, S. Saidha, P.A. Calabresi, Retinal measurements predict 10-year disability in multiple sclerosis, Ann. Clin. Transl. Neurol. 6 (2019) 222–232, https://doi.org/10.1002/acn3.674.

[11] A.U. Brandt, E.H. Martinez-Lapiscina, R. Nolan, S. Saidha, Monitoring the course of MS with optical coherence tomography, Curr. Treat. Options Neurol. 19 (2017) 15, https://doi.org/10.1007/s11940-017-0452-7.

[12] S. Saidha, O. Al-Louzi, J.N. Ratchford, P. Bhargava, J. Oh, S.D. Newsome, J. L. Prince, D. Pham, S. Roy, P. van Zijl, L.J. Balcer, E.M. Frohman, D.S. Reich, C. Crainiceanu, P.A. Calabresi, Optical coherence tomography reflects brain atrophy in multiple sclerosis: a four-year study, Ann. Neurol. 78 (2015) 801–813, https://doi.org/10.1002/ana.24487.

[13] G. Kitsos, E.T. Detorakis, S. Papakonstantinou, A.P. Kyritsis, S.H. Pelidou, Perimetric and peri-papillary nerve fibre layer thickness findings in multiple sclerosis, Eur. J. Neurol. 18 (2011) 719–725, https://doi.org/10.1111/j.1468-1331.2010.03256.x.

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

[15] E. Garcia-Martin, L.E. Pablo, R. Herrero, M. Satue, V. Polo, J.M. Larrosa, J. Martin, J. Fernandez, Diagnostic ability of a linear discriminant function for spectral-domain optical coherence tomography in patients with multiple sclerosis, Ophthalmology 119 (2012) 1705–1711, https://doi.org/10.1016/j.ophtha.2012.01.046.

[16] E. Garcia-Martin, R. Herrero, M.P. Bambo, J.R. Ara, J. Martin, V. Polo, J. M. Larrosa, J. Garcia-Feijoo, L.E. Pablo, Artificial neural network techniques to improve the ability of optical coherence tomography to detect optic neuritis, Semin. Ophthalmol. 30 (2015) 11–19, https://doi.org/10.3109/08820538.2013.810277.

[17] A. Pérez del Palomar, J. Cegoñino, A. Montolío, E. Orduna, E. Vilades, B. Sebastián, L.E. Pablo, E. Garcia-Martin, Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques, PloS One 14 (2019), e0216410, https://doi.org/10.1371/journal.pone.0216410.

[18] C. Cavaliere, E. Vilades, M. Alonso-Rodríguez, M. Rodrigo, L. Pablo, J. Miguel, E. López-Guillén, E. Morla, L. Boquete, E. Garcia-Martin, Computer-Aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features, Sensors 19 (2019) 5323, https://doi.org/10.3390/s19235323.

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

[20] E. Garcia-Martin, D. Rodriguez-Mena, R. Herrero, C. Almarcegui, I. Dolz, J. Martin, J.R. Ara, J.M. Larrosa, V. Polo, J. Fernández, L.E. Pablo, Neuro-ophthalmologic evaluation, quality of life, and functional disability in patients with MS, Neurology 81 (2013) 76–83, https://doi.org/10.1212/WNL.0b013e318299ccd9.

[21] Y. Zhao, B.C. Healy, D. Rotstein, C.R.G. Guttmann, R. Bakshi, H.L. Weiner, C. E. Brodley, T. Chitnis, Exploration of machine learning techniques in predicting multiple sclerosis disease course, PloS One 12 (2017), e0174866, https://doi.org/10.1371/journal.pone.0174866.

[22] A. Tousignant, M. Paul LemâıtreLemâıtre, C. Doina Precup, D.L. Arnold, Prediction of disease progression in multiple sclerosis patients using deep learning analysis of MRI data tal arbel 3, Proc. Mach. Learn. Res. 102 (2019) 483–492, in: http://proceedings.mlr.press/v102/tousignant19a.html.

[23] J. Yperman, T. Becker, D. Valkenborg, V. Popescu, N. Hellings, B. Van Wijmeersch, L.M. Peeters, Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis, BMC Neurol. 20 (2020) 1–15, https://doi.org/10.1186/s12883-020-01672-w.

[24] M.F. Pinto, H. Oliveira, S. Batista, L. Cruz, M. Pinto, I. Correia, P. Martins, C. Teixeira, Prediction of disease progression and outcomes in multiple sclerosis with machine learning, Sci. Rep. 10 (2020) 1–13, https://doi.org/10.1038/s41598-020-78212-6.

[25] R. Seccia, D. Gammelli, F. Dominici, S. Romano, A.C. Landi, M. Salvetti, A. Tacchella, A. Zaccaria, A. Crisanti, F. Grassi, L. Palagi, Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis, PloS One 15 (2020) 1–18, https://doi.org/10.1371/journal.pone.0230219.

[26] G. Bsteh, H. Hegen, B. Teuchner, M. Amprosi, K. Berek, F. Ladstätter, S. Wurth, M. Auer, F. Di Pauli, F. Deisenhammer, T. Berger, Peripapillary retinal nerve fibre layer as measured by optical coherence tomography is a prognostic biomarker not only for physical but also for cognitive disability progression in multiple sclerosis, Mult. Scler. J. 25 (2019) 196–203, https://doi.org/10.1177/1352458517740216.

[27] G. Bsteh, H. Hegen, B. Teuchner, K. Berek, S. Wurth, M. Auer, F. Di Pauli, F. Deisenhammer, T. Berger, Peripapillary retinal nerve fibre layer thinning rate as a biomarker discriminating stable and progressing relapsing–remitting multiple sclerosis, Eur. J. Neurol. 26 (2019) 865–871, https://doi.org/10.1111/ene.13897.

[28] G. Bsteh, H. Hegen, P. Altmann, M. Auer, K. Berek, F. Di Pauli, S. Wurth, A. Zinganell, P. Rommer, F. Deisenhammer, F. Leutmezer, T. Berger, Retinal layer thinning is reflecting disability progression independent of relapse activity in multiple sclerosis, Mult. Scler. J. - Exp. Transl. Clin. 6 (2020), https://doi.org/10.1177/2055217320966344.

[29] R.A. Armstrong, Statistical guidelines for the analysis of data obtained from one or both eyes, Ophthalmic Physiol. Opt. 33 (2013) 7–14, https://doi.org/10.1111/opo.12009.

[30] W.I. McDonald, A. Compston, G. Edan, D. Goodkin, H.P. Hartung, F.D. Lublin, H. F. McFarland, D.W. Paty, C.H. Polman, S.C. Reingold, M. Sandberg-Wollheim, W. Sibley, A. Thompson, S. van den Noort, B.Y. Weinshenker, J.S. Wolinsky, Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis, Ann. Neurol. 50 (2001) 121–127, https://doi.org/10.1002/ana.1032.

[31] L.T. Chylack, J.K. Wolfe, D.M. Singer, M.C. Leske, M.A. Bullimore, I.L. Bailey, J. Friend, D. McCarthy, S.Y. Wu, The lens Opacities classification system III. The longitudinal study of cataract study group, Arch. Ophthalmol. (Chicago, Ill. 111

(1960) 831–836, https://doi.org/10.1001/archopht.1993.01090060119035, 1993.
[32] A. Minneboo, B. Jasperse, F. Barkhof, B.M.J. Uitdehaag, D.L. Knol, V. De Groot, C. H. Polman, J.A. Castelijns, Predicting short-term disability progression in early multiple sclerosis: added value of MRI parameters, J. Neurol. Neurosurg. Psychiatry 79 (2008) 917–923, https://doi.org/10.1136/jnnp.2007.124123.
[33] K. Potdar, T. S, C. D, A comparative study of categorical variable encoding techniques for neural network classifiers, Int. J. Comput. Appl. 175 (2017) 7–9, https://doi.org/10.5120/ijca2017915495.
[34] P. Peduzzi, J. Concato, E. Kemper, T.R. Holford, A.R. Feinstein, A simulation study of the number of events per variable in logistic regression analysis, J. Clin. Epidemiol. 49 (1996) 1373–1379, https://doi.org/10.1016/S0895-4356(96)00236-3.
[35] M. Tan, J. Pu, B. Zheng, Optimization of breast mass classification using sequential forward floating selection (SFFS) and a support vector machine (SVM) model, Int. J. Comput. Assist. Radiol. Surg. 9 (2014) 1005–1020, https://doi.org/10.1007/s11548-014-0992-1.
[36] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B. 58 (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[37] E.M. Mowry, A.K. Hedström, M.A. Gianfrancesco, X. Shao, C.A. Schaefer, L. Shen, K.H. Bellesis, F.B.S. Briggs, T. Olsson, L. Alfredsson, L.F. Barcellos, Incorporating machine learning approaches to assess putative environmental risk factors for multiple sclerosis, Mult. Scler. Relat. Disord. 24 (2018) 135–141, https://doi.org/10.1016/j.msard.2018.06.009.
[38] J. Mata, Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models, Eng. Struct. 33 (2011) 903–910, https://doi.org/10.1016/j.engstruct.2010.12.011.
[39] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, New York, NY, 2009, https://doi.org/10.1007/978-0-387-84858-7.
[40] A. Mathur, G.M. Foody, Multiclass and binary SVM classification: implications for training and classification users, Geosci. Rem. Sens. Lett. IEEE 5 (2008) 241–245, https://doi.org/10.1109/LGRS.2008.915597.
[41] N.I.R. Yassin, S. Omran, E.M.F. El Houby, H. Allam, Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review, Comput. Methods Progr. Biomed. 156 (2018) 25–45, https://doi.org/10.1016/j.cmpb.2017.12.012.
[42] P. Cunningham, S.J. Delany, k-Nearest neighbour classifiers, Mult. Classif. Syst. 34 (2007) 1–17.
[43] P.C. Cheeseman, M. Self, J. Kelly, W. Taylor, D. Freeman, J.C. Stutz, Bayesian classification, in: Proc. 7th Natl. Conf. Artif. Intell. St. Paul, MN, USA, August 21-26, 1988, 1988, pp. 607–611.
[44] H. Bonab, F. Can, Less is more: a comprehensive framework for the number of components of ensemble classifiers, IEEE Trans. Neural Networks Learn. Syst. 30 (2019) 2735–2745, https://doi.org/10.1109/TNNLS.2018.2886341.
[45] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer New York, New York, NY, 2013, https://doi.org/10.1007/978-1-4614-6849-3.
[46] Z.C. Lipton, D.C. Kale, C. Elkan, R. Wetzel, Learning to diagnose with LSTM recurrent neural networks, in: 4th Int. Conf. Learn. Represent, ICLR, 2016, pp. 1–18, 2016, http://arxiv.org/abs/1511.03677.
[47] J.D. Rodriguez, A. Perez, J.A. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 569–575, https://doi.org/10.1109/TPAMI.2009.187.
[48] F. Wang, H.-H. Franco-Penya, J.D. Kelleher, J. Pugh, R. Ross, in: P. Perner (Ed.), An Analysis of the Application of Simplified Silhouette to the Evaluation of K-Means Clustering Validity, Springer International Publishing, Cham, 2017, pp. 291–305, https://doi.org/10.1007/978-3-319-62416-7_21.
[49] C.C. Wu, W.C. Yeh, W.D. Hsu, M.M. Islam, P.A. (Alex) Nguyen, T.N. Poly, Y. C. Wang, H.C. Yang, Y.C. (Jack) Li, Prediction of fatty liver disease using machine learning algorithms, Comput. Methods Progr. Biomed. 170 (2019) 23–29, https://doi.org/10.1016/j.cmpb.2018.12.032.
[50] F. London, H. Zéphir, E. Drumez, J. Labreuche, N. Hadhoum, J. Lannoy, J. Hodel, P. Vermersch, J.-P. Pruvo, X. Leclerc, O. Outteryck, Optical coherence tomography: a window to the optic nerve in clinically isolated syndrome, Brain 142 (2019) 903–915, https://doi.org/10.1093/brain/awz038.
[51] F. Costello, J. Burton, Retinal imaging with optical coherence tomography: a biomarker in multiple sclerosis? Eye Brain 10 (2018) 47–63, https://doi.org/10.2147/EB.S139417.
[52] S. Noval, I. Contreras, S. Muñoz, C. Oreja-Guevara, B. Manzano, G. Rebolleda, Optical coherence tomography in multiple sclerosis and neuromyelitis optica: an update, Mult. Scler. Int. 2011 (2011) 1–11, https://doi.org/10.1155/2011/472790.
[53] C. Fjeldstad, M. Bemben, G. Pardo, Reduced retinal nerve fiber layer and macular thickness in patients with multiple sclerosis with no history of optic neuritis identified by the use of spectral domain high-definition optical coherence tomography, J. Clin. Neurosci. 18 (2011) 1469–1472, https://doi.org/10.1016/j.jocn.2011.04.008.
[54] U. Birkeldh, A. Manouchehrinia, M.A. Hietala, J. Hillert, T. Olsson, F. Piehl, I. S. Kockum, L. Brundin, O. Zahavi, M. Wahlberg-Ramsay, R. Brautaset, M. Nilsson, The temporal retinal nerve fiber layer thickness is the most important optical coherence tomography estimate in multiple sclerosis, Front. Neurol. 8 (2017), https://doi.org/10.3389/fneur.2017.00675.
[55] A.A. Khanifar, G.J. Parlitsis, J.R. Ehrlich, G.D. Aaker, D.J. D'Amico, S.A. Gauthier, S. Kiss, Retinal nerve fiber layer evaluation in multiple sclerosis with spectral domain optical coherence tomography, Clin. Ophthalmol. 4 (2010) 1007–1013, https://doi.org/10.2147/OPTH.S13278.
[56] A. Klistorner, E.C. Graham, C. Yiannikas, M. Barnett, J. Parratt, R. Garrick, C. Wang, Y. You, S.L. Graham, Progression of retinal ganglion cell loss in multiple sclerosis is associated with new lesions in the optic radiations, Eur. J. Neurol. 24 (2017) 1392–1398, https://doi.org/10.1111/ene.13404.
[57] E. Viladés, A.P. Del Palomar, J. Cegoñino, J. Obis, M. Satue, E. Orduna, L.E. Pablo, M. Ciprés, E. GarciaMartin, Physiological changes in retinal layers thicknesses measured with swept source optical coherence tomography, PloS One 15 (2020) 1–14, https://doi.org/10.1371/journal.pone.0240441.
[58] B. Alamouti, J. Funk, Retinal thickness decreases with age: an OCT study, Br. J. Ophthalmol. 87 (2003) 899–901, https://doi.org/10.1136/bjo.87.7.899.
[59] C.K.S. Leung, M. Yu, R.N. Weinreb, C. Ye, S. Liu, G. Lai, D.S.C. Lam, Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a prospective analysis of age-related loss, Ophthalmology 119 (2012) 731–737, https://doi.org/10.1016/j.ophtha.2011.10.010.
[60] E.S. Sotirchos, N. Gonzalez Caldito, A. Filippatou, K.C. Fitzgerald, O.C. Murphy, J. Lambe, J. Nguyen, J. Button, E. Ogbuokiri, C.M. Crainiceanu, J.L. Prince, P. A. Calabresi, S. Saidha, Progressive multiple sclerosis is associated with faster and specific retinal layer atrophy, Ann. Neurol. (2020), https://doi.org/10.1002/ana.25738.
[61] S. Meyer-Moock, Y.-S. Feng, M. Maeurer, F.-W. Dippel, T. Kohlmann, Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis, BMC Neurol. 14 (2014) 58, https://doi.org/10.1186/1471-2377-14-58.
[62] M. Satue, J. Obis, M.J. Rodrigo, S. Otin, M.I. Fuertes, E. Vilades, H. Gracia, J. R. Ara, R. Alarcia, V. Polo, J.M. Larrosa, L.E. Pablo, E. Garcia-Martin, Optical coherence tomography as a biomarker for diagnosis, progression, and prognosis of neurodegenerative diseases, J. Ophthalmol. (2016) 1–9, https://doi.org/10.1155/2016/8503859, 2016.
[63] R. Alonso, D. Gonzalez-Moron, O. Garcea, Optical coherence tomography as a biomarker of neurodegeneration in multiple sclerosis: a review, Mult. Scler. Relat. Disord. 22 (2018) 77–82, https://doi.org/10.1016/j.msard.2018.03.007.
[64] A.J. Thompson, B.L. Banwell, F. Barkhof, W.M. Carroll, T. Coetzee, G. Comi, J. Correale, F. Fazekas, M. Filippi, M.S. Freedman, K. Fujihara, S.L. Galetta, H. P. Hartung, L. Kappos, F.D. Lublin, R.A. Marrie, A.E. Miller, D.H. Miller, X. Montalban, E.M. Mowry, P.S. Sorensen, M. Tintoré, A.L. Traboulsee, M. Trojano, B.M.J. Uitdehaag, S. Vukusic, E. Waubant, B.G. Weinshenker, S.C. Reingold, J. A. Cohen, Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria, Lancet Neurol. 17 (2018) 162–173, https://doi.org/10.1016/S1474-4422(17)30470-2.