Interpretable classification using occlusion sensitivity on multilayer segmented OCT from patients with Multiple Sclerosis and healthy controls

Zahra Khodabandeh
Isfahan University of Medical Sciences
Hossein Rabbani
Isfahan University of Medical Sciences
Fereshteh Ashtari
Isfahan University of Medical Sciences
Hanna G. Zimmermann
Charité–Charité– Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health
Seyedamirhosein Motamedi
Charité–Charité– Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health
Alexander U. Brandt
Charité–Charité– Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health
Friedemann Paul
Charité–Charité– Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin Institute of Health
Rahele Kafieh (rkafieh@gmail.com)
Isfahan University of Medical Sciences

Article

Keywords: multiple sclerosis, optical coherence tomography, interpretable Artificial Intelligence, generalizable, patient-wise cross-validation

Posted Date: April 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1547669/v1

Interpretable classification using occlusion sensitivity on multilayer segmented OCT from patients with Multiple Sclerosis and healthy controls

Zahra Khodabandeh[1], Hossein Rabbani[1], Fereshteh Ashtari[2], Hanna G. Zimmermann[3,4], Seyedamirhosein Motamedi[3,4], Alexander U. Brandt[3,4,6], Friedemann Paul[3,4,7], Rahele Kafieh[1,4,5$]

[1]School of Advanced Technologies in Medicine, Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[2]Isfahan Neurosciences Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[3]Experimental and Clinical Research Center, Max Delbrück Center for Molecular Medicine and Charité – Universitätsmedizin Berlin, Berlin, Germany
[4]NeuroCure Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
[5]Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK
[6]Department of Neurology, University of California, Irvine, CA, USA
[7]Department of Neurology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
$) Corresponding author: Rahele Kafieh

Abstract

Multiple sclerosis (MS) is one of the most prevalent chronic inflammatory diseases caused by demyelination and axonal damage in the central nervous system. Structural retinal imaging via optical coherence tomography (OCT) shows promise as a biomarker for monitoring of MS. There are successful reports regarding application of Artificial Intelligence (AI) in analysis of cross-sectional OCTs in ophthalmologic diseases. However, the alteration of sub-retinal thicknesses in MS are noticeably subtle compared to other ophthalmologic diseases. Therefore, raw cross-sectional OCTs are replaced with multilayer segmented OCTs for discrimination of MS and heathy controls (HCs). To conform to the principles of trustworthy AI, interpretability is provided by visualizing regional layer contribution to classification performance with proposed occlusion sensitivity approach. The robustness of the classification is also guaranteed by showing the effectiveness of the algorithm while being tested on the new independent (but similar) dataset. The most discriminative features from different topologies of the multilayer segmented OCTs are selected by dimension reduction methods. Support vector machine (SVM), random forest (RF), and artificial neural network (ANN) are used for classification. Patient-wise cross-validation (CV) is utilized to evaluate the performance of the algorithm, where the training and test folds contain records from different subjects. The most discriminative topology is determined to be squares with side of 40 pixels and the most influential sub-retinal layers are ganglion cell and inner plexiform layer (GCIPL) and inner nuclear layer (INL). Linear SVM resulted in 88% Accuracy, 78% precision and 63% recall in discrimination of MS and HCs using macular multilayer segmented OCTs.

Keywords: multiple sclerosis, optical coherence tomography, interpretable Artificial Intelligence, generalizable, patient-wise cross-validation

Introduction

Multiple sclerosis (MS) is a chronic inflammatory and neurodegenerative disease of the central nervous system (CNS) that causes progressive neurological disability over time. MS is determined by demyelination and neuro-axonal damage that results in tissue loss and progressive neurologic deficits [1]. While the most established method to monitor the degree of CNS damage in MS is magnetic resonance imaging (MRI) [2], MS leads to widespread changes in the retina and optic nerve, which may be assessed with optical coherence tomography (OCT) and useful as disease biomarker [3,4]. OCT-derived imaging markers peripapillary retinal nerve fiber layer thickness (pRNFL) and composite thickness of macular ganglion cell layer (GCL) and Inner plexiform layer (IPL) named (GCIPL) are reported suggested biomarkers for neurodegeneration [5]. Inflammatory disease activity leads to changes in inner nuclear layer thickness (INL) [6]. Focal GCIPL thinning can be measured by aligning and subtracting focal retinal layer thicknesses from a normal healthy population [7,8].

Artificial intelligence (AI) is a promising area of health innovation [9,10]. Its application in ophthalmology is also evident in analysis of different ocular images [11,12], with purpose of segmenting the retinal boundaries [13], discriminating different diseases [14,15] or interpretation of neurological diseases using quality control (QC) criteria [16]. Cross-sectional OCTs are successfully employed in AI for detection of ophthalmologic diseases. However, the alteration of sub-retinal thicknesses in MS are noticeably subtle to be diagnosed with raw cross-sectional OCTs. The other limitation of AI in medical applications is its black box nature which contradicts with interpretability in trustworthy AI. Furthermore, limiting the training and testing datasets to single clinical centers leads to less generalizable algorithms. Finally, cross-validation (CV) in most of medical AI works is performed instance-wise, which overestimates algorithm prediction accuracy [17].

Here we propose an AI method that aims to capture ultra-fine changes in retinal sub-layers by using multilayer segmented OCT. The method is interpretable, which means regional layer contribution to classification performance is visualized using proposed occlusion sensitivity approach. A novel approach is proposed to add interpretability to current machine learning approaches. We test the trained model on an independent second dataset to show robustness. The patient-wise CV is used where the training and test folds contain eyes from different subjects; therefore, in testing stage, the performance is measured on a new subject whose data from the fellow eye has not been used for training.

By considering the mentioned concepts, feature selection from different topologies of multilayer segmented OCTs is done. We compare performances of support vector machine (SVM), random forest (RF), and artificial neural network (ANN), and identify the most discriminative topology and the most influential sub-retinal layers.

Methods

Structure of the datasets

Generalizable algorithms are of interest in medical AI, but when both training and testing datasets come from single clinical centers, attaining this goal cannot be evaluated. We therefore concentrate on two independent (but similar) datasets to be used as separate training and testing datasets in measuring the robustness of the algorithm.

Charité dataset: The first OCT dataset is from the NeuroCure Clinical Research Center (NCRC) at Charité – Universitätsmedizin Berlin, Berlin, Germany. It consists of 422 HC and 106 MS OCTs from two multimodal register studies to evaluate quantitative measurements of neuroaxonal damage in MS. The OCT data in this dataset includes 40 to 51 B-scans with a size of 496 × (479 to 555) pixels for each B-scan. All OCT measurements were carried out with an SD-OCT and Heidelberg Eye Explorer (HEYEX) version 5.7.5.0 by eight individual operators and automatic real-time function for image averaging and an activated eye tracker in a dimly lit room. Retrospective inclusion criteria for the study were participants in a healthy condition aged between 18 and 70 years, Caucasian ethnicity, and high-quality macular OCT scans. Collecting this dataset was approved by the ethics committee of Charité - Universitätsmedizin Berlin and was conducted according to the Declaration of Helsinki in the applicable version. The macular OCT scans were produced from the device and stored in HEYEX vol file format and then a segmentation approach was carried out using a segmentation pipeline. All segmentation results were quality controlled and manually corrected [18].

Isfahan dataset: The second OCT dataset is from the Kashani Comprehensive MS center in Isfahan, Iran between April 2017 and March 2019 [19]. The images were obtained using HEYEX version 5.1 by one trained technician with automatic real-time (ART) of 9 frames function for image averaging. The dataset consists of 45 HC and 45 MS eyes. The automated segmentation was carried out using a graph-based method [20,21]. All segmentation results were quality controlled and manually corrected by an experienced grader using custom-developed software [19,22].

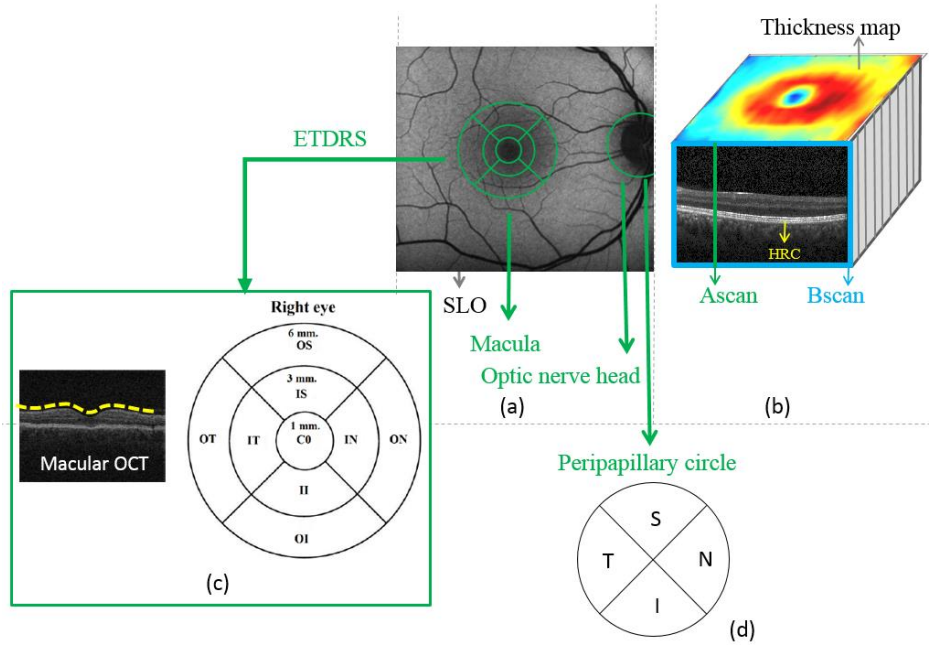

Figure 1: Retinal parameters acquired by OCT. (a) location of sectors and ring scan on SLO image. (b) A-scan, B-scan, and thickness map on OCT data. (c) quadrants in ETDRS: central fovea (CF), inner superior (IS), inner nasal (IN), inner inferior (II), inner temporal (IT), outer superior (OS), outer superior (OS), outer nasal (ON), outer inferior (OI) and outer temporal (OT). (d) quadrants in the peripapillary circle: superior (S), inferior (I), temporal (T), and nasal(N) [19].

Preprocessing and feature extraction

Intraretinal thickness changes in MS are noticeably subtle compared to primary eye disorders [5]. Multilayer segmented OCTs are therefore used and the distance between pairs of the retinal layers, called retinal thickness map, are calculated (Figure 1). The area covered by B-scans around the macula may be oriented. As one possible hypothesis, the effect of compensating the orientation angle is studied in this work. For this purpose, the thickness maps are rotated to have a unique format as input to the next processing steps. The angle between a horizontal line through the disc center and the disc–foveal line ($angle\_Fovea\_ONH\_SLine$) (Figure 2(a)); and the relative direction of each B-scan to a horizontal line through the disc center ($slope\_Bscan$) (Figure 2(b)) are calculated. The left eyes are also flipped. The value of correcting rotation ($rotation\ angle -$ Figure 2(c)) is calculated by:

$$\begin{cases} rotation\ angle = angle\_Fovea\_ONH\_SLine & if\ \ slope\_Bscans\ = 0 \\ rotation\ angle = abs(-90 - angle\_Fovea\_ONH\_SLine) & if\ \ 88 < slope\_Bscan <= 90 \\ rotation\ angle = (angle\_Fovea\_ONH\_SLine\_deg) + (slope\_Bscan) & otherwise \end{cases} \quad (1)$$

The rotated thickness maps are cropped to a unique size of 450×450 pixels (Figure 2(d)). The thickness maps (with/without rotation) from different retinal layers including mRNFL, GCIP, sum of GCIP and INL layers (GCIP+INL), parallel use of GCIP and INL (GCIP/INL), parallel use of mRNFL, GCIP, INL, ONL, and the total macular thickness are considered as input to the classification stage.

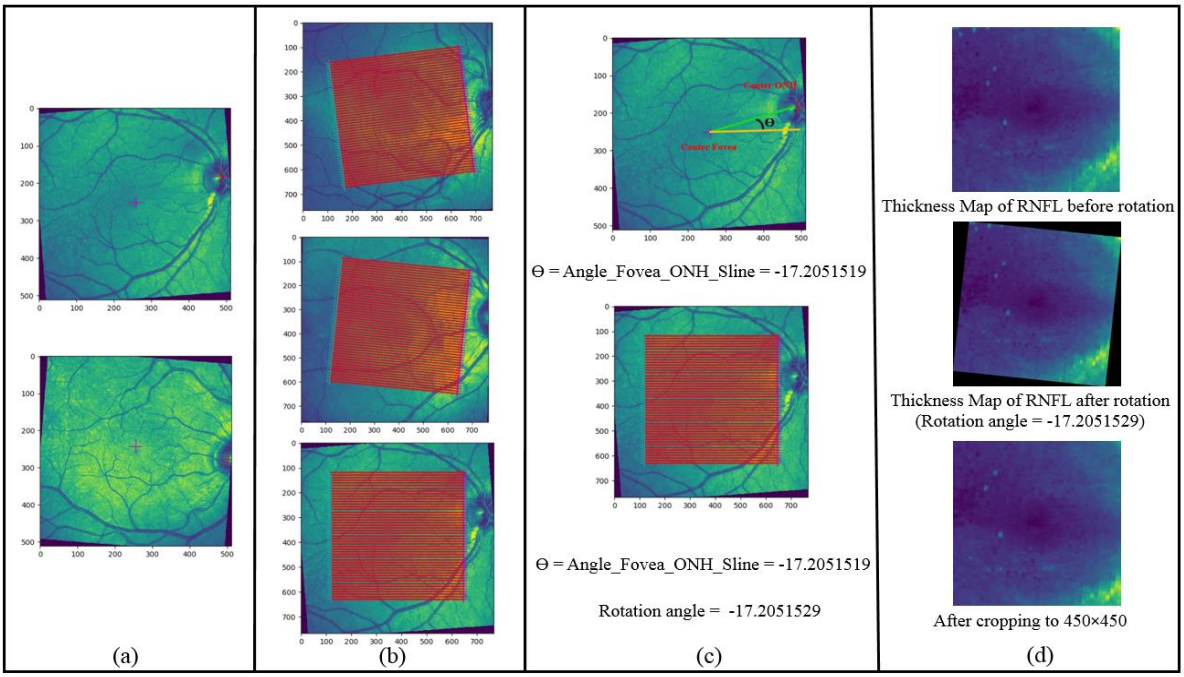

Figure 2: (a) SLO images in clockwise and counterclockwise rotations. (b) B-scans in different directions. (c) Example of finding the rotation angle), (d) Process of rotating a thickness map.

To extract different topological information from each thickness map, regions of interest typically follow those defined by the Early Treatment Diabetic Retinopathy Study (ETDRS) [19]. ETDRS concentric circles are calculated with diameters of 1 mm, 3mm, and 6 mm around the fovea, divided into quadrants and forming nine macular areas demonstrated in Figure 1. As alternative topologies, we also used different resolutions of the thickness maps in squares ranging between 20×20, 30×30, and 40×40 pixels. A combination of retinal layers, classifiers, evaluation, and dimension reduction approaches are used and summarized in Figure 3.

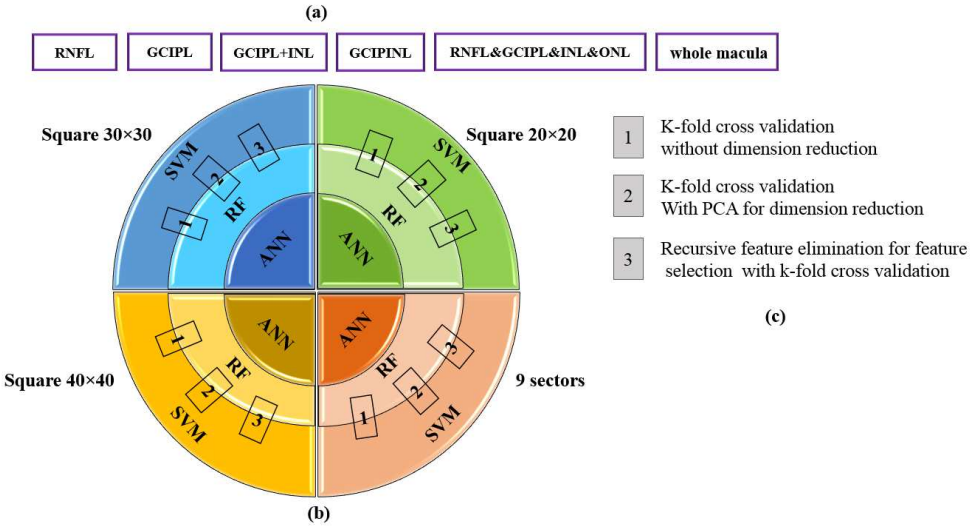

Figure 3: (a) Retinal layers investigated in the proposed method. (b) Classification models: SVM, RF, and ANN on four groups of extracted features: square 20×20, square 30×30, square 40×40, and 9 ETDRS sectors. (c) Different evaluation and dimension reduction methods in the study.

Dimension reduction

To decrease the model complexity and avoid overfitting, we used principal component analysis (PCA) [23] that deduces information from the feature set to make a new feature subspace. Recursive feature elimination (RFE) [24] is also used to select subsets of the main features.

Machine learning algorithms and the evaluation method

Machine learning algorithms are used to explain the patterns in the data and to extract information from it. The algorithms in this study are limited to SVM, RF and ANN.

Support Vector Machine (SVM)

Support Vector Machine is driven by a linear function $w^T x + b$ that predicts the classes according to the sign of this function [25]. In two-class problems, SVM looks for a hyper-plane to divide two different classes with a maximum margin. When the original data is not separable linearly, a nonlinear transformation with a kernel function can be used to transfer the feature space to a higher dimension space with good separability [26]. Kernel functions used in this study are linear, polynomial, radial basis, and sigmoid.

Random Forest (RF)

RF includes many decision trees, and each decision tree prepares a classification for input data. RF gathers the trees and chooses the most voted prediction as the result. The input of each tree is sampled data from the whole dataset. Moreover, a subset of features is randomly chosen from the optimal features to grow the tree at each node[27,28]. We used the grid search method [29] to optimize parameters of a random forest like the number of trees, criterion (the function to measure the quality of a split) including Gini and entropy, and maximum features (the number of features to be considered when looking for the best split) such as sqrt, log2, and auto modes.

Artificial Neural Network (ANN)

An artificial neural network (ANN) includes an input layer of neurons, one or two hidden layers, and an output layer that is the universal function approximator of the interconnection of human neurons [30]. In this study, we found a good performance with a sequential model with four

dense layers. The neurons in each layer are 100, 80, 20 and 1, respectively by grid search method. Rectified linear activation (ReLu) is used as activation function in the first three layers, and the last layer uses sigmoid activation function.

Evaluation methods

Ten-fold patient-wise CV is used with no combination of subjects' eyes in the training and test folds. This approach reduces overestimation of prediction accuracy [17] in instance-wise CV with leakage of information between training and testing phases. Classification performance is evaluated according to the confusion matrix and the values of accuracy, precision, recall, and f1-score are reported.

Interpretability

One of the main limitations of AI in medical applications its black box nature contradicting with interpretability in trustworthy AI. Conventional machine learning methods are mostly designed to work with vectors as input. Therefore, the images are changed into vectors, and the original image structure is ignored. On the other hand, recent methods like Convolutional Neural Networks are introduced as powerful competitors, preserving the image structure, and providing image-based interpretability, expected to be humanly interpretable [31].

In this study, we propose a novel approach to add interpretability to current machine learning approaches. We used occlusion sensitivity [32] and modified it to fit the vector-like inputs. After training the model, we created a black mask with the size of 10×10 pixels and moved it on the test set with a single step to sweep the whole image. The location of pixels covered by the mask are transferred to vector-shaped positions (Figure 4). The masked vector is sent as input to the model and the accuracy is calculated. It is expected that the occlusion of regions with important discriminative information, lead to lower accuracy. The interpretability is shown by regenerating the occlusion with the original image size, with value of accuracy in location of each pixel (called the heat map). An interpretability heatmap indicates how important each location is concerning the class and visualizes the regional contribution to classification.

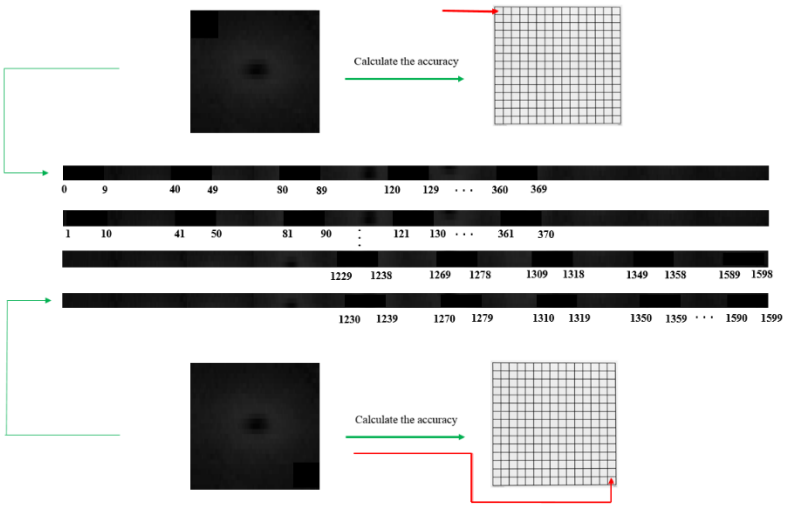

Figure 4: The proposed process for creating a black mask, moving it on the test set, and transferring the locations to vector-shaped positions

Results

For classification purposes, different topologies of the thickness maps around the macula in squares with resolutions of 20×20, 30×30, 40×40 pixels, and mean thicknesses in 9 sectors of ETDRS are considered. The effect of compensating rotation on thickness maps is examined. The classification models are first trained and tested on Charité (first) dataset. To show the generalizability of the method, the trained classifier with the best performance (on the first dataset) is tested on the Isfahan (second) dataset. The proposed occlusion sensitivity is also shown for interpretability.

Different combinations of the features, two different dimension-reduction methods and different machine learning methods are used with 10-fold patient-wise CV on Charité (first) dataset. The comparison of metrics on each parameter is presented by keeping the other parameters fixed on the best-performing set. Table 1 compares effectiveness of different retinal layers and the effect of rotation in correct classification. The selection of best topology is performed based on Table 2. Accuracy of different classification methods with different topologies is presented in Table 3. For SVM method in classification, Table 4 compares the performance of the different kernels. Moreover, dimension reduction methods are compared in Table 5.

Table 1: Comparison of input features in classification of MS and HC. The other parameters are fixed on the best-performing set of information (square size of 40×40, Linear SVM as classifier with 10-fold cross-validation, PCA for dimension reduction). The effect of rotations is shown in the upper and lower part of the table, respectively.

Square 40×40 – (10-fold with PCA) - without rotation				
	Accuracy	Precision	Recall	F1-score
mRNFL	79%	47%	41%	43%
GCIP	87%	72%	60%	64%
GCIP&INL(GCIP/INL)	**88%**	**78%**	**63%**	**68%**
GCIP+INL	82%	56%	51%	52%
mRNFL&GCIP&INL&ONL	84%	64%	58%	59%
Whole macular volume	80%	52%	45%	47%
GCIP & whole macular volume	80%	51%	49%	49%
GCIP & INL & macular volume	81%	54%	52%	52%
Square 40×40 - (10-fold with PCA) - with rotation				
	Accuracy	Precision	Recall	F1-score
mRNFL	74%	33%	33%	33%
GCIP	82%	56%	56%	55%
GCIP&INL(GCIP/INL)	82%	56%	51%	52%
GCIP+INL	83%	63%	50%	54%
mRNFL&GCIP&INL&ONL	82%	58%	47%	50%
Whole macular volume	80%	53%	44%	45%
GCIP & whole macular volume	80%	51%	47%	48%
GCIP & INL & macular volume	79%	47%	51%	48%

Table 2: Comparison of square size in the classification of MS and HC. The other parameters are fixed on the best-performing set of information (GCIP/INL without rotation as input feature, Linear SVM as a classifier with 10-fold cross-validation, and PCA for dimension reduction)

	Accuracy (SVM-linear)	Precision (SVM-linear)	Recall (SVM-linear)	F1-score (SVM-linear)
Square 20×20	84%	64%	57%	57%
Square 30×30	86%	74%	57%	61%

Square 40×40	**88%**	**78%**	**63%**	**68%**
9 sectors	84%	75%	33%	44%

Table 3: Comparison of machine learning methods in classification of MS and HC. The other parameters are fixed on the best-performing set of information (GCIP/INL in square size of 40×40 without rotation as input feature, and PCA for dimension reduction)

	Accuracy (SVM-liner)	Accuracy (RF)	Accuracy (ANN)
Square 20×20	**84%**	84%	82%
Square 30×30	**86%**	85%	84%
Square 40×40	**88%**	85%	85%
9 sectors	84%	85%	82%

Table 4: Comparison of kernels for SVM method in classification of MS and HC. The other parameters are fixed on the best-performing set of information (GCIP/INL in square size of 40×40 without rotation as input feature, SVM as a classifier with 10-fold cross-validation, and PCA for dimension reduction)

	Accuracy	Precision	Recall	F1-score
Linear	**88%**	78%	**63%**	**68%**
Polynomial	85%	87%	28%	41%
Radial basis (RBF)	86%	91%	33%	47%
Sigmoid	83%	67%	38%	47%

Table 5: Comparison of different dimension reduction methods in classification of MS and HC. The other parameters are fixed on the best-performing set of information (GCIP/INL in square size of 40×40 without rotation as input feature, Linear SVM as a classifier with 10-fold cross-validation)

	Accuracy	Precision	Recall	F1-score
10-fold cross validation without dimension reduction	88%	79%	59%	65%
10-fold cross validation with PCA for dimension reduction	**88%**	78%	**63%**	**68%**
10-fold cross validation with RFE for dimension reduction	86%	76%	50%	57%

To explore the application of RFE in cross-validation, the importance of each feature is obtained through a coefficient attribute and features with a correlation coefficient above a threshold of 0.8 are removed. The diagram of accuracy against the number of features is shown in Figure 5.

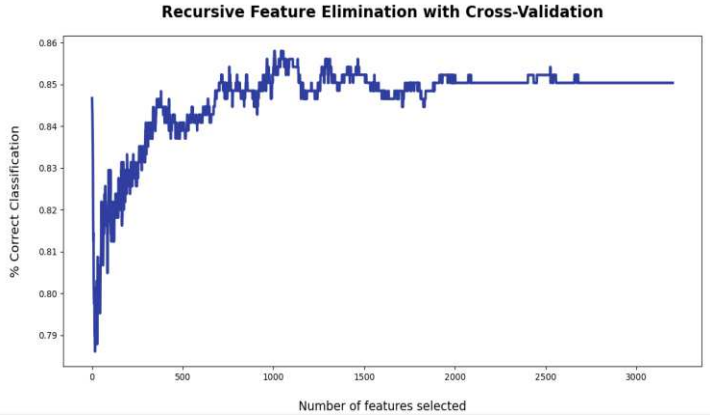

Figure 5: RFE with cross-validation diagram that shows accuracy against the number of features

To show the generalizability of the method, the trained classifier with the best performance on the Charité dataset (GCIP/INL in square size of 40×40 without rotation as input feature, Linear

SVM as a classifier with 10-fold cross-validation, and PCA for dimension reduction) is tested on Isfahan dataset and the performance is shown inTable 6.

Table 6: Classification of MS and HC on Isfahan (second) dataset using a best-performing classifier trained on Charité (first) dataset.

	Accuracy	Precision	Recall	F1-score
SVM(linear) with PCA for dimension reduction and 10-fold cross validation	88%	89%	88%	84%

Visual interpretability

The proposed method for visual interpretability is demonstrated by plotting the heatmap of the occlusion sensitivity. The results in the previous section showed that GCIP/INL (parallel use of GCIP and INL) are the most effective layers in distinguishing MS patients from HCs; therefore, these two layers on the best-performing set of hyperparameters are used for analyzing the interpretability in Figure 6.

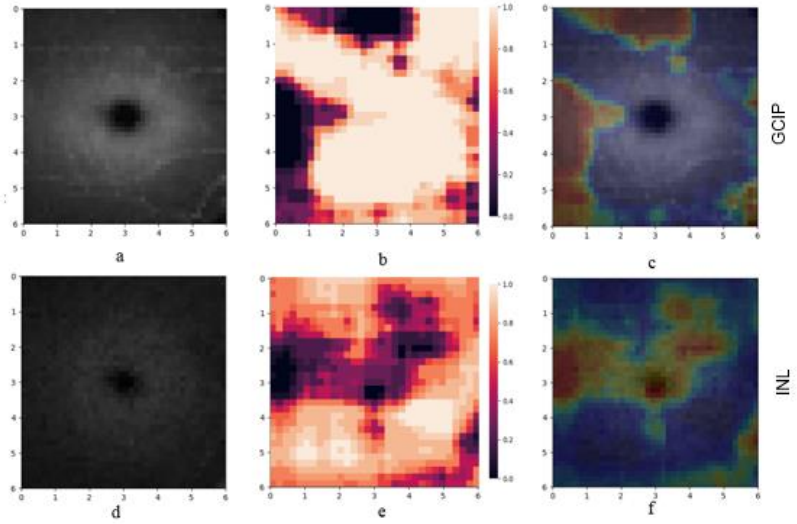

Figure 6: Visual interpretability on thickness maps of GCIP and INL. (a) Thickness map of GCIP in one sample from MS dataset (x and y axis in mm), (b) heatmap of occlusion sensitivity in the classification of MS and HC, (c) overlap of the heatmap and the GCIP layer, (d) Thickness map of INL in one sample from MS dataset, (e) heatmap of occlusion sensitivity in the classification of MS and HC, (f) overlap of the heatmap and the INL layer.

Discussion and conclusion

The model with highest accuracy based on our optimization approaches is able to discriminate MS and HCs with accuracy of 88% and F1-score of 68%, using GCIPL and INL information. Indistinct changes in retinal sub-layers are captured with multilayer segmented OCT. An interpretable result is acquired to indicate regional layer contribution to classification performance using occlusion sensitivity. The generalizability is evaluated by training on a first dataset and then testing on a second independent (but similar) dataset. The performance is similar (accuracy of 88% and F1-score of 84%) when testing on another data, which proves the generalization ability of the proposed method (more detail is presented in Table 7). To avoid overestimation, patient-wise CV is used with separate set of patients in training and test datasets. Different combinations of the retinal layers as input features, two different dimension-reduction methods and different machine learning methods are compared.

Simultaneous data from GCIP and INL (GCIP/INL) is found to be the most informative combination of the retinal layers (Table 1). This finding is in accordance with clinical studies [5] [6]. The rotation of the thickness maps did not improve the performance. One possible reason for this finding is using the traditional machine learning methods which change the image format to vectorized data. This vectorization process may be responsible for reducing the effect of the rotation.

The best topology is shown to be square size of 40×40 (Table 2). It seems that this resolution is relevant to the number of B-scans in each OCT data (40 to 51 B-scans). Namely, 40×40 square extracts the most possible information without suppressing data between the B-scans.

The interpretability heatmap of classifications with this novel proposed algorithm is a new strategy in conventional machine learning methods and makes them comparable to their main competitors like CNN. As demonstrated in Figure 6, the temporal region in the thickness map of GCIP is found to have more effect on the classification of MS disease.

Among the machine learning methods, SVM achieved the best results (Table 3) with linear kernel (Table 4). This finding seems reasonable since linear kernels are proven to be more effective when the number of features is large in comparison to the training samples [33]. Dimension reduction improved in the results and PCA method was found more appropriate (Table 5). The selected model with highest accuracy based on our optimization approaches discriminates MS and HCs with accuracy of 88% and F1-score of 68%, using GCIPL and INL information. Table 7 shows a summary of previous similar methods in comparison with the proposed algorithm. Direct comparison of the results with these works is not possible since the codes and datasets are not released in any of those works. Furthermore, none of the previous works considered the patient-wise CV and accordingly higher performance is reported with leakage of information between train and test data in instance-wise approaches. It should also be noted that in this work, the state of being affected by optic neuritis (ON) was not considered and accordingly, MS patients with/ without ON are combined for classification. Therefore, compared to works on MS with ON, a lower performance is convincing since the eye without ON show less thinning and are be less discriminable from the HCs [5,6,34]. Finally, some previous works include the pRNFL data as input of the classification and a correspondingly higher performance is achieved compared to limited focus of macular region.

Table 7. Summary of previous similar methods

Previous works	Number of datasets	Input retinal layers	Being affected by ON	patient-wise/ instance-wise cross validation	Performance metrics	The most discriminant retinal layer	Classification method
Garcia-Martin et al. [35] 2013	106 MS, 115 HC	Peripapillary area	29% (31) with ON, 71% (75) without ON	instance-wise	AUC=0.945	pRNFL	ANN
Garcia-Martin et al. [36] 2015	112 MS, 105 HC	Peripapillary area	36.6%(41) with ON, 63.4% (71) without ON	instance-wise	Recall=89.3% Specificity=87.6% Precision=88.5%	pRNFL	ANN

Palomar et al. [37] 2019	80 MS, 180 HC	Peripapillary, macular and extended (between macula and papilla) areas	with ON	instance-wise	Decision tree in macular area: Accuracy=97.24% AUC=0.959 In extended area: Accuracy=95.3% AUC=0.998	pRNFL	Decision tree, ANN, SVM
Cavaliere et al. [26] 2019	48 MS, 48 HC	Peripapillary and macular areas	Without ON	instance-wise	Accuracy=91% Recall=89% Specificity=92% AUC=0.97	GCL++ (between inner limiting membrane to INL) and nasal quadrant of outer and inner ring in pRNFL	SVM
Garcia-Martin et al. [38] 2020	48 MS, 48 HC	Macular area	Without ON	instance-wise	Recall=98% Specificity=98% AUC=0.83	GCL++	SVM, ANN
Zhang et al. [39] 2020	58 MS, 63 HC	Macular area	33 with ON, 25 without ON	instance-wise	Recall=64% Specificity=94%	GCIPL	logistic regression (LR), logistic regression regularized with the elastic net penalty (LR-EN), SVM
Montolio et al. [40] 2021	108 MS, 104 HC	Peripapillary and macular areas	34 with ON, 74 without ON	instance-wise	EC: Accuracy=87.7% Recall=87% Specificity=88.5% Precision=88.7% AUC=0.8775 K-NN: Accuracy=85.4% SVM: Accuracy=84.4% LSTM: Accuracy=81.7% Recall=81.1% Specificity=82.2% Precision=78.9% AUC=0.8165	pRNFL	multiple linear regression (MLR), SVM, decision tree, k-nearest neighbors (k-NN), Naïve Bayes (NB), ensemble classifier (EC), long short-term memory (LSTM) recurrent neural network
Proposed algorithm With training and testing on first dataset	106 MS, 422 HC	Macular area	With and without ON	Patient-wise	Accuracy = 88% Precision = 78% Recall =63% F1-score = 68%	GCIPL and INL	Elaborated in the text

Proposed algorithm With training on first dataset and testing on second dataset	Train: 106 MS, 422 HC Test: 67 MS 45 HC	Macular area	With and without ON	Patient-wise	Accuracy = 88% Precision = 89% Recall =88% F1-score = 84%	GCIPL and INL	Elaborated in the text

There are several limitations to the present study. First, the state of having a history of ON is not considered. Second, a longitudinal follow-up data from patients were not taken into consideration. Third, other demographic factors like age, gender, and race were not included. Such information may improve the performance of the proposed methods.

In conclusion, this machine learning approach is designed to fill the gap in previous automatic methods for discrimination of MS and HCs. The relatively big sample size with manually corrected multilayer segmented OCT is used. Various topologies from sub-retinal thicknesses maps are individually analyzed to find the best combination. Interpretability and generalizability are guaranteed with proposed approaches and the overestimated results are avoided with patient-wise techniques. Future works should be done on more comprehensive datasets to prove the effectiveness of such methods in clinical applications.

Data Availability:

The codes and datasets analyzed during the current study available from the corresponding author on reasonable request.

References

1. Reich, D.S.; Lucchinetti, C.F.; Calabresi, P.A. Multiple Sclerosis. *N. Engl. J. Med.* **2018**, *378*, 169, doi:10.1056/NEJMRA1401483.
2. Filippi, M.; Preziosa, P.; Banwell, B.L.; Barkhof, F.; Ciccarelli, O.; De Stefano, N.; Geurts, J.J.G.; Paul, F.; Reich, D.S.; Toosy, A.T. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain* **2019**, *142*, 1858–1875.
3. Graves, J.S.; Oertel, F.C.; Van der Walt, A.; Collorone, S.; Sotirchos, E.S.; Pihl-Jensen, G.; Albrecht, P.; Yeh, E.A.; Saidha, S.; Frederiksen, J. Leveraging visual outcome measures to advance therapy development in neuroimmunologic disorders. *Neurol. Neuroinflammation* **2022**, *9*.
4. Costello, F.; Burton, J.M. Retinal imaging with optical coherence tomography: a biomarker in multiple sclerosis? *Eye Brain* **2018**, *10*, 47.
5. Petzold, A.; Balcer, L.; Calabresi, P.A.; Costello, F.; Frohman, T.; Frohman, E.; Martinez-Lapiscina, E.H.; Green, A.; Kardon, R.; Outteryck, O.; et al. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol.* **2017**, *16*, 797–812, doi:10.1016/S1474-4422(17)30278-8.
6. Oertel, F.C.; Zimmermann, H.G.; Brandt, A.U.; Paul, F. Novel uses of retinal imaging with optical coherence tomography in multiple sclerosis. *Expert Rev. Neurother.* **2019**, *19*, 31–43, doi:10.1080/14737175.2019.1559051.
7. Hu, H.; Jiang, H.; Gameiro, G.R.; Hernandez, J.; Delgado, S.; Wang, J. Focal thickness reduction of the ganglion cell-inner plexiform layer best discriminates prior optic neuritis in patients with multiple sclerosis. *Invest. Ophthalmol. Vis. Sci.* **2019**, *60*, 4257–4269.
8. Shi, C.; Jiang, H.; Gameiro, G.R.; Hu, H.; Hernandez, J.; Delgado, S.; Wang, J. Visual function and disability are associated with focal thickness reduction of the ganglion cell-inner plexiform layer in patients with multiple sclerosis. *Invest. Ophthalmol. Vis. Sci.* **2019**, *60*, 1213–1223.

9. Pesapane, F.; Codari, M.; Sardanelli, F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur. Radiol. Exp.* **2018**, *2*, 1–10.
10. Oren, O.; Gersh, B.J.; Bhatt, D.L. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit. Heal.* **2020**, *2*, e486–e488.
11. Li, J.-P.O.; Liu, H.; Ting, D.S.J.; Jeon, S.; Chan, R.V.P.; Kim, J.E.; Sim, D.A.; Thomas, P.B.M.; Lin, H.; Chen, Y. Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective. *Prog. Retin. Eye Res.* **2021**, *82*, 100900.
12. Wang, Z.; Keane, P.A.; Chiang, M.; Cheung, C.Y.; Wong, T.Y.; Ting, D.S.W. Artificial intelligence and deep learning in ophthalmology. *Artif. Intell. Med.* **2020**, 1–34.
13. Maloca, P.M.; Müller, P.L.; Lee, A.Y.; Tufail, A.; Balaskas, K.; Niklaus, S.; Kaiser, P.; Suter, S.; Zarranz-Ventura, J.; Egan, C. Unraveling the deep learning gearbox in optical coherence tomography image segmentation towards explainable artificial intelligence. *Commun. Biol.* **2021**, *4*, 1–12.
14. Yoon, J.; Han, J.; Park, J.I.; Hwang, J.S.; Han, J.M.; Sohn, J.; Park, K.H.; Hwang, D.D.-J. Optical coherence tomography-based deep-learning model for detecting central serous chorioretinopathy. *Sci. Rep.* **2020**, *10*, 1–9.
15. De Fauw, J.; Ledsam, J.R.; Romera-Paredes, B.; Nikolov, S.; Tomasev, N.; Blackwell, S.; Askham, H.; Glorot, X.; O'Donoghue, B.; Visentin, D. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **2018**, *24*, 1342–1350.
16. Petzold, A.; Albrecht, P.; Balcer, L.; Bekkers, E.; Brandt, A.U.; Calabresi, P.A.; Deborah, O.G.; Graves, J.S.; Green, A.; Keane, P.A. Artificial intelligence extension of the OSCAR-IB criteria. *Ann. Clin. Transl. Neurol.* **2021**, *8*, 1528–1542.
17. Saeb, S.; Lonini, L.; Jayaraman, A.; Mohr, D.C.; Kording, K.P. The need to approximate the use-case in clinical machine learning. *Gigascience* **2017**, *6*, gix019.
18. Motamedi, S.; Gawlik, K.; Ayadi, N.; Zimmermann, H.G.; Asseyer, S.; Bereuter, C.; Mikolajczak, J.; Paul, F.; Kadas, E.M.; Brandt, A.U. Normative data and minimally detectable change for inner retinal layer thicknesses using a semi-automated OCT image segmentation pipeline. *Front. Neurol.* **2019**, *10*, 1117.
19. Ashtari, F.; Ataei, A.; Kafieh, R.; Khodabandeh, Z.; Barzegar, M.; Raei, M.; Dehghani, A.; Mansurian, M. Optical Coherence Tomography in Neuromyelitis Optica spectrum disorder and Multiple Sclerosis: A population-based study. *Mult. Scler. Relat. Disord.* **2020**, *47*, 102625.
20. Kafieh, R.; Rabbani, H.; Abramoff, M.D.; Sonka, M. Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map. *Med. Image Anal.* **2013**, *17*, 907–928, doi:10.1016/j.media.2013.05.006.
21. Kafieh, R.; Rabbani, H.; Hajizadeh, F.; Abramoff, M.D.; Sonka, M. Thickness mapping of eleven retinal layers segmented using the diffusion maps method in normal eyes. *J. Ophthalmol.* **2015**, *2015*.
22. Montazerin, M.; Sajjadifar, Z.; Pour, E.K.; Riazi-Esfahani, H.; Mahmoudi, T.; Rabbani, H.; Movahedian, H.; Dehghani, A.; Akhlaghi, M.; Kafieh, R. Livelayer: a semi-automatic software program for segmentation of layers and diabetic macular edema in optical coherence tomography images. *Sci. Rep.* **2021**, *11*, 1–13.
23. Shlens, J. A tutorial on principal component analysis. *arXiv Prepr. arXiv1404.1100* **2014**.
24. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422.
25. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep learning*; MIT press Cambridge, 2016; Vol. 1;.
26. Cavaliere, C.; Vilades, E.; Alonso-Rodríguez, M.; Rodrigo, M.J.; Pablo, L.E.; Miguel, J.M.; López-Guillén, E.; Morla, E.M.; Boquete, L.; Garcia-Martin, E. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. *Sensors* **2019**, *19*, 5323.
27. Mao, W.; Wang, F.Y. Cultural Modeling for Behavior Analysis and Prediction. *New Adv. Intell. Secur. Informatics, 1st ed.; Acad. Press Waltham, MA, USA* **2012**, 91–102.
28. Zheng, G.; Li, S.; Szekely, G. *Statistical shape and deformation analysis: methods, implementation and applications*; Academic Press, 2017; ISBN 0128104945.
29. Syarif, I.; Prugel-Bennett, A.; Wills, G. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkomnika* **2016**, *14*, 1502.
30. Wang, S.-C. Artificial neural network. In *Interdisciplinary computing in java programming*; Springer, 2003; pp. 81–100.
31. Shukla, P.; Verma, A.; Verma, S.; Kumar, M. Interpreting SVM for medical images using Quadtree. *Multimed.*

Tools Appl. **2020**, *79*, 29353–29373.
32. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Proceedings of the European conference on computer vision; Springer, 2014; pp. 818–833.
33. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. A practical guide to support vector classification"(http://www. csie. ntu. edu. tw/~ cjlin/papers/guide/gu ide. pdf. **2003**.
34. Aly, L.; Noll, C.; Wicklein, R.; Wolf, E.; Romahn, E.F.; Wauschkuhn, J.; Hosari, S.; Mardin, C.; Berthele, A.; Hemmer, B. Dynamics of Retinal Vessel Loss After Acute Optic Neuritis in Patients With Relapsing Multiple Sclerosis. *Neurol. Neuroinflammation* **2022**, *9*.
35. Garcia-Martin, E.; Pablo, L.E.; Herrero, R.; Ara, J.R.; Martin, J.; Larrosa, J.M.; Polo, V.; Garcia-Feijoo, J.; Fernandez, J. Neural networks to identify multiple sclerosis with optical coherence tomography. *Acta Ophthalmol.* **2013**, *91*, e628–e634.
36. Garcia-Martin, E.; Herrero, R.; Bambo, M.P.; Ara, J.R.; Martin, J.; Polo, V.; Larrosa, J.M.; Garcia-Feijoo, J.; Pablo, L.E. Artificial neural network techniques to improve the ability of optical coherence tomography to detect optic neuritis. In Proceedings of the Seminars in ophthalmology; Taylor & Francis, 2015; Vol. 30, pp. 11–19.
37. Pérez del Palomar, A.; Cegoñino, J.; Montolío, A.; Orduna, E.; Vilades, E.; Sebastián, B.; Pablo, L.E.; Garcia-Martin, E. Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques. *PLoS One* **2019**, *14*, e0216410.
38. Garcia-Martin, E.; Ortiz, M.; Boquete, L.; Sánchez-Morla, E.M.; Barea, R.; Cavaliere, C.; Vilades, E.; Orduna, E.; Rodrigo, M.J. Early diagnosis of multiple sclerosis by OCT analysis using Cohen'sd method and a neural network as classifier. *Comput. Biol. Med.* **2020**, 104165.
39. Zhang, J.; Wang, J.; Wang, C.; Delgado, S.; Hernandez, J.; Hu, H.; Cai, X.; Jiang, H. Wavelet features of the thickness map of retinal ganglion cell-inner plexiform layer best discriminate prior optic neuritis in patients with multiple sclerosis. *IEEE Access* **2020**.
40. Montolío, A.; Martín-Gallego, A.; Cegoñino, J.; Orduna, E.; Vilades, E.; Garcia-Martin, E.; Del Palomar, A.P. Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. *Comput. Biol. Med.* **2021**, *133*, 104416.