




RESEARCH ARTICLE

Prior optic neuritis detection on peripapillary ring scans using deep learning

Seyedamirhosein Motamedi¹ , Sunil Kumar Yadav^{1,2}, Rachel C. Kenney^{3,4}, Ting-Yi Lin¹, Josef Kauer-Bonin^{1,2}, Hanna G. Zimmermann¹ , Steven L. Galetta⁴, Laura J. Balcer⁴, Friedemann Paul^{1,5}  & Alexander U. Brandt^{1,6}

¹Experimental and Clinical Research Center, a cooperation between the Max Delbrück Center for Molecular Medicine in the Helmholtz Association and Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

²Nocturne GmbH, Berlin, Germany

³Departments of Radiology and Radiological Sciences and Electrical and Computer Engineering, Vanderbilt University Medical Center, Nashville, Tennessee, USA

⁴Departments of Neurology, Population Health and Ophthalmology, New York University, New York, New York, USA

⁵Department of Neurology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

⁶Department of Neurology, University of California, Irvine, California, USA

Correspondence

Alexander U. Brandt, Department of Neurology, University of California, Irvine, 252 Health Sciences Rd, Irvine, California 92697, USA. E-mail: aubrandt@hs.uci.edu

Funding Information

This study was supported in part by Berlin Institute of Health (project “DEEP-Neuroretina” to Alexander U. Brandt), and by the Kathleen C. Moore Foundation (to Alexander U. Brandt).

Received: 29 April 2022; Revised: 22 June 2022; Accepted: 7 July 2022

Annals of Clinical and Translational Neurology 2022; 9(11): 1682–1691

doi: 10.1002/acn3.51632

Abstract

Background: The diagnosis of multiple sclerosis (MS) requires demyelinating events that are disseminated in time and space. Peripapillary retinal nerve fiber layer (pRNFL) thickness as measured by optical coherence tomography (OCT) distinguishes eyes with a prior history of acute optic neuritis (ON) and may provide evidence to support a demyelinating attack. **Objective:** To investigate whether a deep learning (DL)-based network can distinguish between eyes with prior ON and healthy control (HC) eyes using peripapillary ring scans. **Methods:** We included 1033 OCT scans from 415 healthy eyes (213 HC subjects) and 510 peripapillary ring scans from 164 eyes with prior acute ON (140 patients with MS). Data were split into 70% training, 15% validation, and 15% test data. We included 102 OCT scans from 80 healthy eyes (40 HC) and 61 scans from 40 ON eyes (31 MS patients) from an independent second center. Receiver operating characteristic curve analyses with area under the curve (AUC) were used to investigate performance. **Results:** We used a dilated residual convolutional neural network for the classification. The final network had an accuracy of 0.85 and an AUC of 0.86, whereas pRNFL only had an AUC of 0.77 in recognizing ON eyes. Using data from a second center, the network achieved an accuracy of 0.77 and an AUC of 0.90 compared to pRNFL, which had an AUC of 0.84. **Interpretation:** DL-based disease classification of prior ON is feasible and has the potential to outperform thickness-based classification of eyes with and without history of prior ON.

Introduction

Multiple sclerosis (MS) is the most common chronic autoimmune demyelinating disorder of the central nervous system.¹ In the absence of a pathological biomarker, diagnosing clinically definite MS (CDMS) requires dissemination of demyelination and neurological symptoms in time and space.² The McDonald criteria allow the

diagnosis of MS at the time of a clinically isolated syndrome (CIS) by utilizing biomarkers that primarily include T2-weighted lesions on brain magnetic resonance imaging (MRI) to establish paraclinical evidence of dissemination in time and space.³ While the McDonald criteria are ubiquitously used in clinical care, their sensitivity and specificity for clinically definite MS are only moderate⁴ causing delays in diagnosis⁵ and

diagnostic certainty.⁶ Furthermore, longitudinal studies have shown earlier treatment of ON with MS-disease modifying therapies improves visual and physical disability over time, rendering earlier diagnosis of MS by adding ON as a lesion site important.⁷

The optic nerve is a key site of disease in MS.⁸ Post-mortem studies reveal significant retinal ganglion cell loss in more than 70% of MS patient eyes.^{9,10} While acute optic neuritis (ON) represents the initial manifestation of MS in up to 25% of patients, and approximately 50% of MS patients experience acute ON at some point in their disease, demyelinating plaques in the optic nerve are found in up to 99% of MS patients at *post mortem*.^{11,12} Recognizing optic nerve pathology, that is signs of prior ON, could thus serve as an additional criterion supplementing the current McDonald criteria.¹³ During the last revision of the criteria this potential was recognized, but optic nerve assessment was ultimately not included in the revised criteria because of a lack of supporting data, which were mainly derived from MR imaging studies that may or may not have included dedicated orbital images or of the optic nerve with fat-saturation and contrast.¹⁴

Optical coherence tomography (OCT) obtains high-resolution structural measurements of retinal thickness that reflect the integrity of the optic nerve. These measurements have been suggested as diagnostic biomarkers in MS that not only reflect structural but also functional visual outcomes in MS.^{15–17} As such, the peripapillary retinal nerve fiber layer thickness (pRNFL) is reduced within weeks to months after ON, and increased magnitudes of thickness reduction are associated with worsening degrees of visual function loss.¹⁸ However, the large individual ranges of normal OCT measurements render absolute thickness measurements of pRNFL alone less useful for detecting prior ON.^{19,20}

Artificial intelligence, using deep learning and convolutional neural networks, allow image classification based on raw imaging data.²¹ Deep learning-based methods are trained directly on imaging data, and resulting networks have shown remarkable success in recognizing retinal pathologies in primary eye disorders.²² Recently, we have shown that deep learning can be used for automatic quality control of OCT images.²³ We and others have further shown that deep learning-based image segmentation outperforms classical intra-retinal layer segmentation on OCT images.²⁴

We hypothesized in this study that deep learning-based recognition of prior ON by peripapillary ring scans is feasible and outperforms thickness-based classification using absolute pRNFL values. To investigate this, we trained a deep neural network on a carefully-curated data set of pRNFL scans from healthy subjects and eyes with history of acute ON of patients with MS. We then evaluated

network performance for recognizing prior ON under several conditions, including independent validation data.

Methods

Study population

In this study, we retrospectively included data from relapsing–remitting MS (RRMS) and CIS patients with a history of ON from two longitudinal observational studies at the Experimental and Clinical Research Center (ECRC) of Charité—Universitätsmedizin Berlin. Data were acquired between April 2011 and September 2021. Inclusion criteria were minimum age of 18 years and a diagnosis of RRMS or CIS according to the 2017 revised McDonald criteria.³ Exclusion criteria were acute ON within 90 days before OCT examinations and any other neurological or eye disorder known to affect the retina (e.g., glaucoma, diabetes, and refractive error of >6 diopters). We only included eyes with a documented history of ON; this was based on providers' notes in medical records and patient attack history. Additionally, age- and sex-matched data from healthy controls (HC) with no history of neurological or eye disorder from the ECRC's database were included in this study. Repeated measurements during follow-up were allowed in this study. Table 1 shows an overview of the study cohort.

A confirmatory cohort, consisting of 61 OCT scans from 40 eyes of 31 patients with RRMS with a history of ON (65% women, age at baseline: 41 ± 12 years) and 102 OCT scans from 80 eyes of 40 healthy controls (53% women, age at baseline: 29 ± 10 years) from a longitudinal observational cohort study at the Department of

Table 1. Demographic description of HC and ON cohorts of the exploratory (Berlin) data.

	HC	ON
No. of subjects [N]	213	140
No. of eyes [N]	415	164
No. of OCT scans [N]	1033	510
Age at baseline [years] ¹	36 ± 12	37 ± 10
Sex (female) [N (%)] ²	142 (67)	93 (66)
Time since onset at baseline (median [range]) [years]	–	1.12 [0.26–43.08]
Time since last ON (median [range]) [years]	–	3.02 [0.25–44.83]
Diagnosis at baseline	–	108 RRMS 32 CIS

CIS, clinically isolated syndrome; HC, healthy controls; OCT, optical coherence tomography; ON, optic neuritis; RRMS, relapsing–remitting multiple sclerosis.

¹Age match: $p = 0.12$.

²Sex match: $p = 1$.

Neurology, New York University Grossman School of Medicine, New York, NY, was included in this study by incorporating the same inclusion and exclusion criteria. RRMS and HC groups were well matched in the confirmatory cohort by sex (p value: 0.43) but not by age (p value: <0.001). The mean time from disease onset for the RRMS group was 10.02 years with minimum of 0.25 and maximum of 23 years. This cohort had a more diverse racial/ethnic origin, in HC (ethnicity: Not Hispanic or Latino = 34, Hispanic or Latino = 4, 2 missing; race: Caucasian = 27, African-American = 1, Asian or Pacific Islander = 7, Other = 4, and 1 missing) and patient data (ethnicity: Not Hispanic or Latino = 19, Hispanic or Latino = 3, 9 missing; race: Caucasian = 16, African-American = 3, Asian or Pacific Islander = 1, Other = 1, 10 missing).

Optical coherence tomography

All OCT scans of both exploratory and confirmatory cohorts were performed using Spectralis spectral-domain OCT (Heidelberg Engineering, Heidelberg, Germany), with automatic real-time (ART) averaging and activated eye tracking. Measurements were performed in a dimly lit room without pupil dilation unless required. Peripapillary ring scans were obtained either as single 2D OCT B scans (12° , 1536 A scans per B scan, 496 pixels per A scan, $16 \leq \text{ART} \leq 100$) or extracted from the optic nerve head radial and circular (ONH_RC) protocol taken with anatomic positioning system (APS) (inner ring scan with constant circle diameter of 3.5 mm, 768 A scans per B scan, 496 pixels per A scan, $16 \leq \text{ART} \leq 100$). All scans were quality controlled according to the OSCAR-IB criteria²⁵ by an experienced grader. Seven OCT scans from the exploratory cohort and one OCT scan from the confirmatory cohort were excluded from the study because of inadequate OCT scan quality or inaccurate intraretinal segmentation. OCT data are reported in accordance with the Advised Protocol for OCT Study Terminology and Elements (APOSTEL) recommendations.²⁶ This study was conducted in accordance to the artificial intelligence extension of the OSCAR-IB criteria (OSCAR-AI),²⁷ with the exception that the criterion of Openness could not be fulfilled because of data protection constraints on data-sharing.

Artificial Neural Network

Training test and validation data sets

The OCT scans included in this study were randomly allocated to the training, validation and test data sets. We allocated 70% of the data for model training, 15% for model validation during the training and 15% for model

testing after the training. Data allocation of repeated measures of an individual subject required that all measurements of this individual only appeared in one data set. Random allocation was repeated until having around 70% of patients and eyes in the training data set and 15% of them in each of the validation and test data sets. This process was necessary in order to ensure the homogeneity of all data sets with regard to the number of patients and eyes in addition to the number of OCT scans. In the end, we allocated 728 OCT scans from 290 eyes of 149 HC and 352 OCT scans from 115 eyes of 98 patients to the training data set, 152 OCT scans from 32 HC (63 eyes) and 79 OCT scans from 21 patients (25 eyes) to the validation data set, and 153 OCT scans from 32 HC (62 eyes) and 79 OCT scans from 21 patients (24 eyes) to the test data set.

Preprocessing

Figure 1 summarizes all preprocessing steps. All OCT scans went through intraretinal segmentation of inner limiting membrane (ILM), lower boundary of retinal nerve fiber layer, and Bruch's membrane (BM) using Nocturne ONE software (Nocturne GmbH, Berlin, Germany). All segmentation results were manually quality controlled as described above. Based on the segmentation, a label image was created for each OCT scan with four different classes: (1) The area above the ILM, (2) pRNFL, (3) the area between the lower boundary of RNFL and BM, and (4) the area below BM. After segmentation, both OCT and label images were flattened using the BM as reference. Scans were then resized to a fixed size of 512×512 pixel using bi-linear interpolation. Finally, the intensity of pixels located between ILM and BM (foreground) was linearly normalized for each individual scan, and the intensity of the pixels above ILM and below BM (background) was set to zero (black background).

Data augmentation

To increase data size, diversity, translation invariance, classification accuracy and possibly reduce over-fitting, we performed model training additionally on synthetically modified data created using data augmentation. Augmentation was performed by applying random horizontal and vertical circular shift of up to 20 pixels to all OCT images of the training data set. To test if data augmentation results in improvements, the model was trained on five different training data sets with these scenarios: (1) No data augmentation (510 ON OCT images, 1033 HC OCT images), (2) ON data augmentation by a factor of 2 using random vertical shift (1020 ON OCT images, 1033 HC OCT images), (3) ON data augmentation by a factor of 2

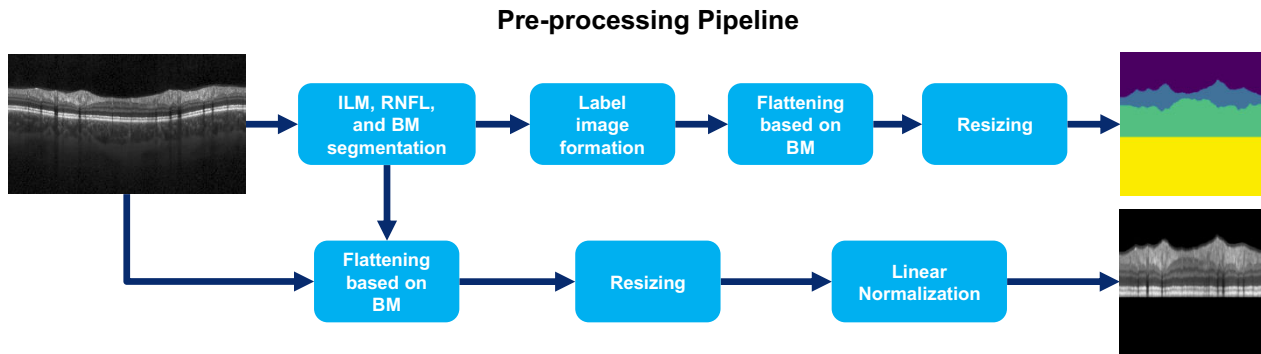


Figure 1. The pipeline for the pre-processing step. BM, Bruch's membrane; ILM, inner limiting membrane; RNFL, retinal nerve fiber layer.

using random vertical and horizontal shift (1020 ON OCT images, 1033 HC OCT images), (4) ON data augmentation by a factor of 4 and HC data augmentation by a factor of 2 using random vertical shift (2040 ON OCT images, 2066 HC OCT images), and (5) ON data augmentation by a factor of 4 and HC data augmentation by a factor of 2 using random vertical and horizontal shift (2040 ON OCT images, 2066 HC OCT images). The main reason to augment more on ON data was that repeated OCT measurements at follow ups are more similar to each other in healthy controls than ON patients who have a possibility of having disease worsening and repeated acute ON episodes between the measurements. Another reason behind performing more data augmentation on ON was to compensate for the imbalance between the size of the HC and ON data. The reason behind applying mainly vertical shift was that OCT ring scans have more positioning similarity in the horizontal direction than the vertical direction.

Implementation

The model was fully developed in Python version 3.7 using Keras with TensorFlow backend. Early stopping regularization was used to avoid over-fitting during training by stopping the training process after the convergence of training and validation accuracy. Model training was performed on an on-premise Linux workstation with two Intel Xeon Gold 6144 (Intel Corporation, Santa Clara, CA) CPUs and four NVIDIA GeForce GTX 1080 Ti (NVIDIA Corporation, Santa Clara, CA) GPUs. Model testing was performed on a Windows desktop with an Intel Core i9 11980HK CPU and a NVIDIA GeForce RTX 3080 GPU.

Statistical analysis

The convenience sample size in this pilot study was not based on an a priori sample size estimation. To confirm

group matching, sex differences between groups were tested using χ^2 test and age differences were tested using 2-sample Wilcoxon test. Classification performance was evaluated using area under the curve (AUC) from receiver operating characteristic (ROC) curves. The optimal cutoff for the classifier was estimated based on the intersection of a 45° tangent line and the ROC curve, which is equivalent to the maximum of the summation of the sensitivity and specificity. F1 score, a measure of accuracy, was calculated as: $\frac{\text{True positive}}{\text{True positive} + \frac{1}{2}(\text{False positive} + \text{False negative})}$, with ON as the positive class and HC as the negative. $p < 0.05$ were considered significant. All statistical analysis were performed in R version 3.5.²⁸

Ethics statement

The study was approved by the local ethics committee at Charité—Universitätsmedizin Berlin (EA1/182/10 and EA1/163/12). Independent validation data from NYU were collected under IRB approval at NYU. The study was conducted according to the Declaration of Helsinki in its currently applicable version and the applicable German and European laws. All participants gave written informed consent.

Results

Network architecture

The proposed convolutional neural network (CNN) for the classification of eyes with a history of ON vs HC eyes is shown in Figure 2. The designed CNN accepts two-channel inputs consisting of the OCT scans (Channel 1) as well as their segmentation map (Channel 2), both with dimensions of 512×512 pixels. Our proposed model uses dilated convolution, which results in exponential expansion of the receptive field by linearly increasing the dilation factor, without increasing the kernel size.²⁹ The

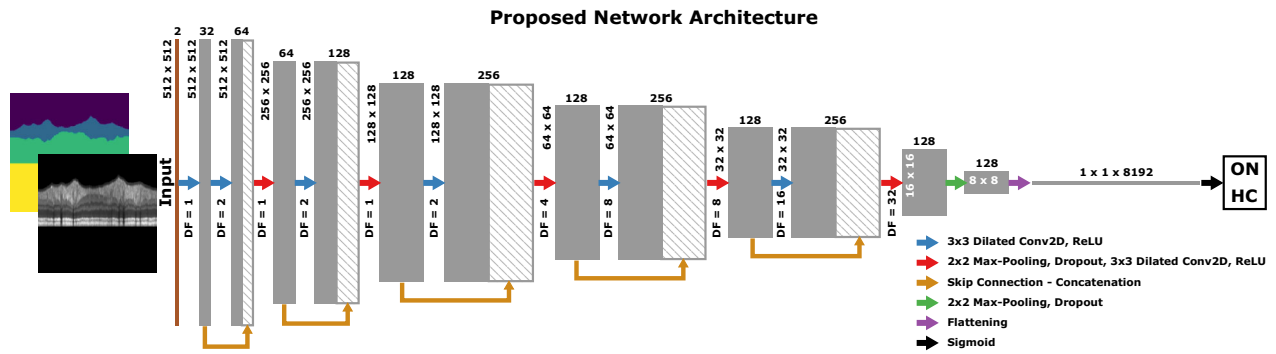


Figure 2. The architecture of the proposed model. Conv2D, 2D convolution layer; DF, dilation factor; HC, healthy controls; ON, optic neuritis; ReLU, rectified linear activation function.

model also employs residual learning using skip connections, which helps it to avoid the vanishing gradients problem.³⁰ The network has five dilated residual blocks, each consists of two convolutional layers with kernel size of 3×3 and different dilation factors, followed by concatenation of the feature maps created by each convolutional layer, a max-pooling layer of size 2×2 and stride of two, and a dropout layer with dropout rate of 0.4. After the dilated residual blocks, the model has another 3×3 dilated convolutional layer with a dilation factor of 32, followed by max-pooling and dropout layers with the same features as the previous layers. Each convolutional layer is followed by a rectified linear unit (ReLU) activation function. In order to make the final prediction, the output of the last layer is flattened and then fed to a dense layer, followed by a sigmoid activation function. For the training, binary cross entropy (BCE) loss was used as $-\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$, where y is the classification label (1 for ON and 0 for HC) and $p(y)$ is the output of the model which is the predicted probability of an input image being from an eye with a history of ON. Adaptive moment estimation (Adam) algorithm was used to optimize the model during the training with a fixed learning rate of 0.0001.

Model validation with data augmentation

The proposed model was trained on training data sets with different data augmentation scenarios. Table 2 summarizes the performance of the model on the validation data set. As shown in the table, the models trained on the training data sets with only ON data augmentation have better overall performance in identification of eyes with prior ON from healthy controls, and therefore these models and the model trained on the training data set without any augmentation were selected to be tested on the test data set.

Recognizing signs of optic neuritis in MS eyes

Table 3 summarizes the performance of the selected models on the test data. The optimal cutoff for each classifier based on its performance on the test data was calculated and used to report the metrics in Table 3. As shown, the model trained on the training data set with only ON data augmentation with vertical shift had the best overall performance and thus was selected. The model was used with the classification cutoff calculated based on the test data in this study. ROC curve analysis for the selected classifier

Table 2. Model performance with data augmentation on validation data set.

Aug.	Type	TP	FN	FP	TN	Acc.	Sen.	Spec.	F1
None	–	44	35	6	146	0.82	0.56	0.96	0.68
2 × ON	Vert.	44	35	5	147	0.83	0.56	0.97	0.69
2 × ON	Vert. – Horiz.	42	37	2	150	0.83	0.53	0.99	0.68
4 × ON – 2 × HC	Vert.	45	34	11	141	0.81	0.57	0.93	0.67
4 × ON – 2 × HC	Vert. – Horiz.	43	36	5	147	0.82	0.54	0.97	0.68

The best result for each evaluation metric is highlighted in bold. Acc., accuracy; F1, F1 score; FN, false negative; FP, false positive; HC, healthy controls—negative class; Horiz., horizontal shift; ON, optic neuritis—positive class; Sen., sensitivity; Spec., specificity; TP, true positive; TN, true negative; Vert., vertical shift.

Table 3. Model performance with data augmentation on test data set.

Aug.	Type	Cut.	TP	FN	FP	TN	Acc.	Sen.	Spec.	F1
None	–	0.493	50	29	8	145	0.84	0.63	0.95	0.73
2 × ON	Vert.	0.469	58	21	14	139	0.85	0.73	0.91	0.77
2 × ON	Vert. – Horiz.	0.467	53	26	9	144	0.85	0.67	0.94	0.75

The best result for each evaluation metric is highlighted in bold. Acc., accuracy; Cut., optimal cutoff; F1, F1 score; FN, false negative; FP, false positive; HC, healthy controls—negative class; Horiz., horizontal shift; ON, optic neuritis—positive class; Sen., sensitivity; Spec., specificity; TP, true positive; TN, true negative; Vert., vertical shift.

is shown in Figure 3A, next to ROC curve for pRNFL for comparison. Figure 4A shows pRNFL for both eyes with a history of ON and HC eyes, separated by the classification result.

To account for correlation between repeated measurements per eye, the performance of the model was tested on a data set with only one OCT image per eye, randomly selected from the test data set. The model had an accuracy of 0.83 with the sensitivity of 0.67, specificity of 0.89, and F1 score of 0.68 for the classification of eyes with a history of ON and HC on the sub data set of the test data.

Model confirmation

We tested our proposed model on an independent cohort of MS patients with a history of ON and HC, to investigate whether the results from the exploratory data could be confirmed on independent data. Our model had an accuracy of

0.77 with sensitivity of 0.89, specificity of 0.71, and F1 score of 0.74 (true positive: 54, false negative: 7, false positive: 30, true negative: 72). ROC curve analysis of the model on the confirmatory cohort as well as using pRNFL thickness to discriminate between ON and HC is shown in Figure 3B. Figure 4B shows the pRNFL values for this cohort against the classification result.

Similar to the exploratory data, we tested the performance of our model on a subset of the confirmatory cohort with only one OCT scan per eye, to account for the effect of repeated measurements per eye. Here, the model had an accuracy of 0.72 with sensitivity of 0.85, specificity of 0.65, and F1 score of 0.67.

Discussion

In this study we show that deep-learning based classification of prior ON on peripapillary ring scans is feasible

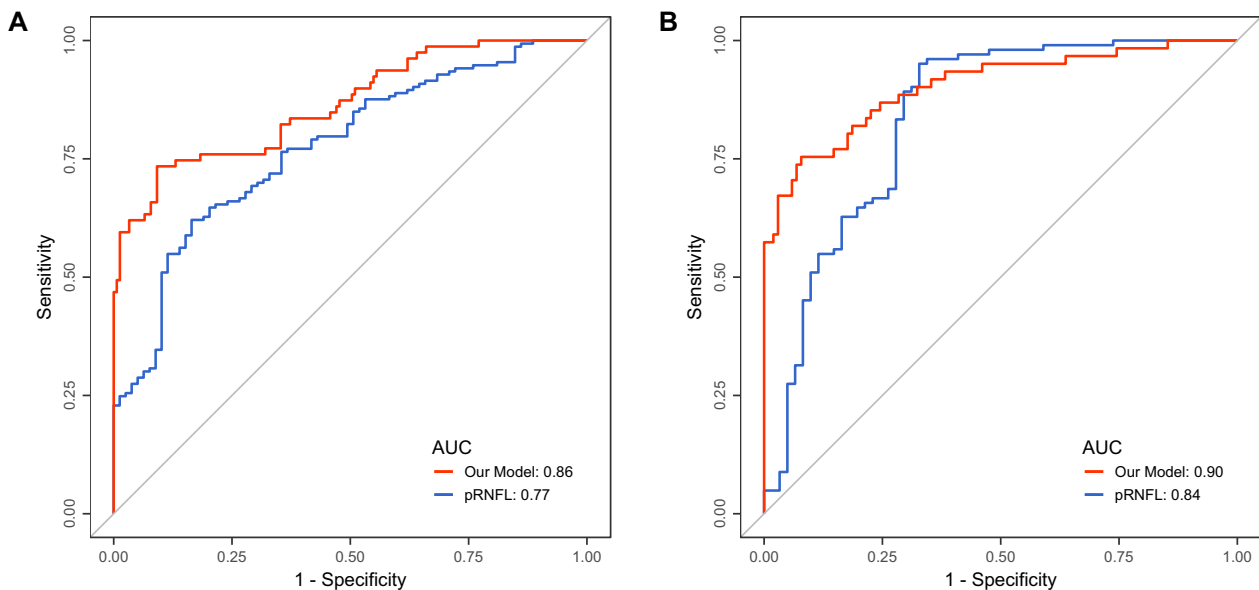


Figure 3. ROC curves for the proposed model and pRNFL identifying eyes with prior ON and HC in (A) exploratory and (B) confirmatory data. AUC, area under ROC curve; HC, healthy control; ON, optic neuritis; pRNFL, peripapillary retinal nerve fiber layer thickness; ROC, receiver operating characteristic.

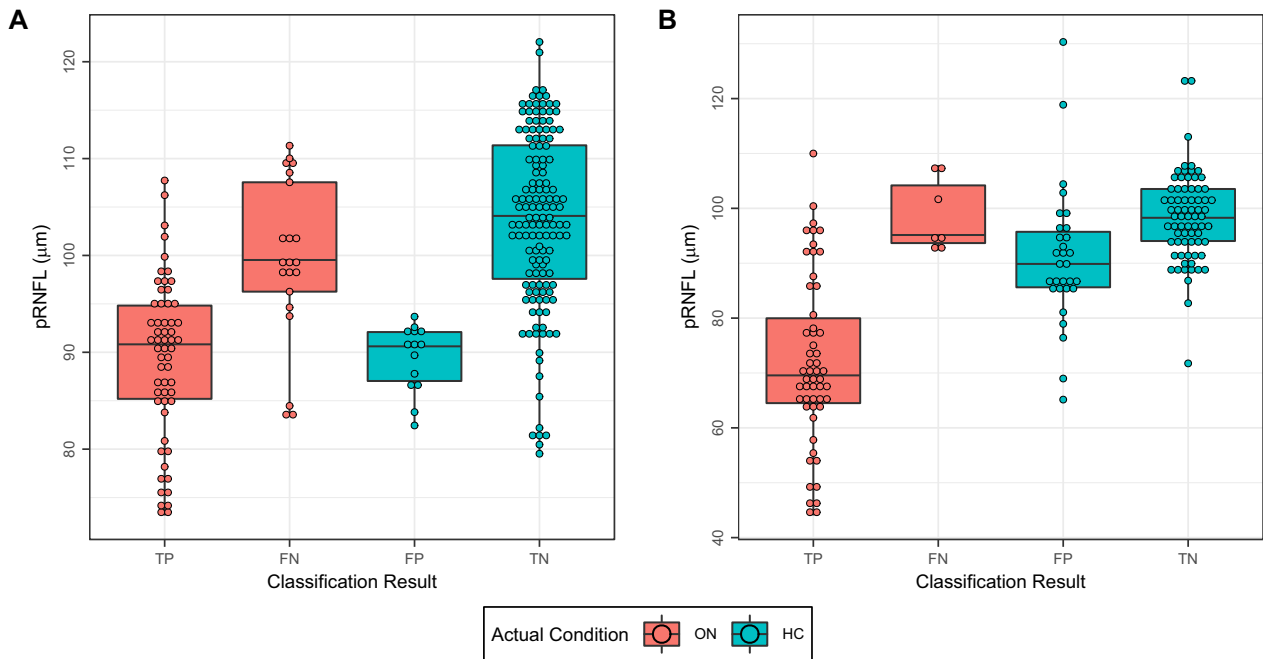


Figure 4. pRNFL stratified by classification result for (A) exploratory and (B) confirmatory data. HC, healthy control; FN, false negative; FP, false positive; ON, optic neuritis; pRNFL, peripapillary retinal nerve fiber layer thickness; TN, true negative; TP, true positive.

and outperforms pRNFL thickness-based classification in our initial cohort and a second independent data set. The good model performance in an independent cohort confirms the generalizability of our approach.

OCT is a simple, fast, and readily available method to query retina and optic nerve for neuroaxonal damage. OCT-derived measures, including pRNFL, have been suggested as imaging biomarkers for MS diagnosis.¹³ Signs of prior ON have been estimated using interocular differences (IOD) of retinal layers on the assumption that a typically unilateral ON would lead to neuroaxonal damage in one eye but not the other.^{31,32} Percentage rather than absolute IOD may be superior over absolute IOD in detecting subtle changes.³³ Clinically meaningful cut-off values may be defined in comparison to healthy controls²⁰ or based on large reference studies with MS-derived OCT measures.^{32,34} Machine learning based classification of eyes with ON history utilizing standard OCT parameters has been promising.³⁵ However, not all ON lead to measurable neuroaxonal damage, and in a previous study about 15% of patients do not suffer relevant damage during an ON episode.³⁶ This limits the value of methods relying on absolute OCT parameter thickness or inter-eye differences. Indeed, the proposed method recognized scans with normal pRNFL thickness correctly as ON, suggesting the DL-based classification may rely on other differences than only average thickness.

Other methods may provide alternative or complementary information regarding optic nerve pathology in MS. A conversion equation recently developed on an international cohort can provide a useful way to convert between two different OCT devices clinically and in research studies (Kenney et al. in press). Alternatively, that is, optic nerve MRI³⁷ or visual evoked potentials (VEP)³⁸ may provide valuable biomarkers to determine dissemination in time and space criteria and for diagnosing MS.

Our choice for the convolutional neural network was a dilated residual network. Residual networks enhances the overall training with reduced training loss by applying residual skip connections to mainly avoid vanishing gradient problems.³⁰ Convolutional networks reduce the receptive field in deeper layers and make decisions based on small limited feature maps, leading to the loss of spatial information.²⁹ This problem can be avoided by either increasing the kernel size of convolutional layers or using the concept of dilated convolution. The latter option is preferable because it does not add more model parameters while increasing the receptive field, which is essential in our classification task, as the network needs to preserve spatial resolution to capture the structural relationship between different anatomical regions on peripapillary ring scans.

Our pilot study has notable weaknesses. Our network was trained on data from one device in one clinical

center, whereas we were able to confirm model performance using additional data from a second, independent center. Thus, our study serves as a proof-of-principle, but the approach would need to be developed further to including multi-center, multi-device, multinational and multi-racial/ethnic data to be clinically applicable. The model in this study was trained, tested, and validated on Spectralis data. The segmentation method we used may have effects on the pRNFL thickness reference data. However, all scans were thoroughly quality controlled, including segmentation, which should lead to similar pRNFL values regardless of the underlying segmentation method. Regarding deep learning-based classification, biggest hurdle for applying our method to data from devices other than Spectralis is the data format. To design a method that would work on different devices' raw data, an interim step for data harmonization and transforming data into the same feature space would be required, which was beyond the scope of this study. Race and ethnicity need to be representative in the training data to avoid cultural bias. The training data of this study was reflective of the German population (predominantly Caucasian), which makes its applicability in racial/ethnic more diverse populations questionable. For example, African American MS patients display more severe neuroretinal damage, which may impair classification in both DL- and thickness-based diagnosis.³⁹ The good performance in the confirmatory cohort, which was more diverse in race/ethnicity somewhat mitigates this effect, but the cohorts were too small to conclusively investigate a potential of cultural bias. Comorbid diseases may affect retinal OCT scans and thickness measurements, and should also be considered,³⁴ which was not possible in this study. Furthermore, the sample size for comparing performance against pRNFL is small, and confirmation in a larger and more diverse sample is necessary. Lastly, ON diagnosis was based on clinical records, which may have led to misclassification of some training data, if records were erroneous.

In summary we show that deep learning-based recognition of ON on peripapillary ring scans is feasible. While our study's results are promising, further methodological improvements both regarding network architecture, used data modalities and training data variability are warranted.

Acknowledgment

This study was supported in part by Berlin Institute of Health (project "DEEP-Neuroretina" to Alexander U. Brandt), and by the Kathleen C. Moore Foundation (to Alexander U. Brandt).

Conflict of Interest

Seyedamirhosein Motamedi received travel funding from Americas Committee for Treatment and Research in Multiple Sclerosis (ACTRIMS). Sunil Kumar Yadav is a cofounder and holds shares in technology start-up Nocturne GmbH, which has commercial interest in OCT applications in neurology; was an employee of Nocturne GmbH; and is named as a coinventor on a pending patent application "Method and system for retinal tomography image quality". Josef Kauer-Bonin is currently an employee of Nocturne GmbH; and is named as a coinventor on a pending patent application "Method and system for retinal tomography image quality". Hanna G. Zimmermann received research grants from Novartis and speaking honoraria from Bayer. Steven L. Galetta received consulting fees from Biogen and Genentech. Laura J. Balcer is the Editor-in-Chief of the Journal of Neuro-Ophthalmology; and is named as a coinventor on patent applications related to the Mobile Universal Lexicon Evaluation System (MULES) and the Staggered Uneven Number (SUN) tests owned by NYU, who also owns the copyright on these tests. Friedemann Paul is a cofounder and holds shares in technology start-up Nocturne GmbH, which has commercial interest in OCT applications in neurology; served on the scientific advisory boards of Novartis and MedImmune; received travel funding and/or speaker honoraria from Bayer, Novartis, Biogen, Teva, Sanofi-Aventis/Genzyme, Merck Serono, Alexion, Chugai, MedImmune, and Shire; is an associate editor of Neurology: Neuroimmunology & Neuroinflammation; is an academic editor of PLoS ONE; consulted for Sanofi Genzyme, Biogen, MedImmune, Shire, and Alexion; received research support from Bayer, Novartis, Biogen, Teva, Sanofi-Aventis/Genzyme, Alexion, and Merck Serono; and received research support from the German Research Council, Werth Stiftung of the City of Cologne, German Ministry of Education and Research, Arthur Arnstein Stiftung Berlin, EU FP7 Framework Program, Arthur Arnstein Foundation Berlin, Guthy-Jackson Charitable Foundation, and NMSS. Alexander U. Brandt is a cofounder and holds shares in technology start-up Nocturne GmbH, which has commercial interest in OCT applications in neurology; is a cofounder and holds shares in Motognosis GmbH; received a grant from Berlin Institute of Health; received travel funding from Americas Committee for Treatment and Research in Multiple Sclerosis (ACTRIMS); named as a coinventor on several patent applications regarding retinal image analysis, MS serum biomarkers, MS-related myelination pathway, and motor function analysis; served on the observational study monitoring board of the Central Vein Sign in MS (CAVS-MS) study (funded by the National Institutes of Health

[NIH]); and serves as Secretary/Treasurer and Member of the Board of the International MS Visual System Consortium. All other authors have no conflicts of interest to declare.

Author Contributions

Seyedamirhosein Motamedi collected and curated the data; designed and implemented the method; performed the analyses; and drafted the manuscript for intellectual content. Sunil Kumar Yadav designed and implemented the method; and revised the manuscript for intellectual content. Rachel C. Kenney collected the confirmatory data; contributed to data interpretation; and revised the manuscript for intellectual content. Ting-Yi Lin contributed to data collection; and revised the manuscript for intellectual content. Josef Kauer-Bonin contributed to the method implementation; and revised the manuscript for intellectual content. Hanna G. Zimmermann contributed to data collection and interpretation; and revised the manuscript for intellectual content. Steven L. Galetta contributed the confirmatory data; contributed to data interpretation; and revised the manuscript for intellectual content. Laura J. Balcer contributed the confirmatory data; contributed to data interpretation; and revised the manuscript for intellectual content. Friedemann Paul contributed to study management; contributed to data interpretation; and revised the manuscript for intellectual content. Alexander U. Brandt designed and conceptualized the study; and drafted the manuscript for intellectual content.

References

1. Reich DS, Lucchinetti CF, Calabresi PA. Multiple sclerosis. *N Engl J Med*. 2018;378:169-180.
2. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol*. 1983;13(3):227-231.
3. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17(2):162-173.
4. van der Vurst de Vries RM, Mescheriakova JY, Wong YYM, et al. Application of the 2017 revised McDonald criteria for multiple sclerosis to patients with a typical clinically isolated syndrome. *JAMA Neurol*. 2018;75(11):1392.
5. Kaufmann M, Kuhle J, Puhon MA, et al. Factors associated with time from first-symptoms to diagnosis and treatment initiation of multiple sclerosis in Switzerland. *Mult Scler J Exp Transl Clin*. 2018;4(4):2055217318814562.
6. Kaisey M, Solomon AJ, Luu M, Giesser BS, Sicotte NL. Incidence of multiple sclerosis misdiagnosis in referrals to two academic centers. *Mult Scler Relat Disord*. 2019;30:51-56.
7. Kenney R, Liu M, Patil S, et al. Long-term outcomes in patients presenting with optic neuritis: analyses of the MSBase registry. *J Neurol Sci*. 2021;430:118067.
8. Toosy AT, Mason DF, Miller DH. Optic neuritis. *Lancet Neurol*. 2014;13(1):83-99.
9. Green AJ, McQuaid S, Hauser SL, Allen IV, Lyness R. Ocular pathology in multiple sclerosis: retinal atrophy and inflammation irrespective of disease duration. *Brain*. 2010;133(6):1591-1601.
10. Kerrison JB, Flynn T, Green WR. Retinal pathologic changes in multiple sclerosis. *Retina*. 1994;14(5):445-451.
11. Ikuta F, Zimmerman HM. Distribution of plaques in seventy autopsy cases of multiple sclerosis in the United States. *Neurology*. 1976;26(6 pt 2):26-28.
12. Toussaint D, Périer O, Verstappen A, Bervoets S. Clinicopathological study of the visual pathways, eyes, and cerebral hemispheres in 32 cases of disseminated sclerosis. *J Clin Neuroophthalmol*. 1983;3(3):211-220.
13. Brownlee WJ, Galetta S. Optic nerve in multiple sclerosis diagnostic criteria: an eye to the eyes? *Neurology*. 2021;96(4):139-140.
14. Filippi M, Preziosa P, Meani A, et al. Prediction of a multiple sclerosis diagnosis in patients with clinically isolated syndrome using the 2016 MAGNIMS and 2010 McDonald criteria: a retrospective study. *Lancet Neurol*. 2018;17(2):133-142.
15. Brandt AU, Martinez-Lapiscina EH, Nolan R, Saidha S. Monitoring the course of MS with optical coherence tomography. *Curr Treat Options Neurol*. 2017;19(4):15.
16. Petzold A, de Boer JF, Schippling S, et al. Optical coherence tomography in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol*. 2010;9(9):921-932.
17. Petzold A, Balcer LJ, Calabresi PA, et al. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol*. 2017;16(10):797-812.
18. Costello F, Coupland S, Hodge W, et al. Quantifying axonal loss after optic neuritis with optical coherence tomography. *Ann Neurol*. 2006;59(6):963-969.
19. Bock M, Brandt AU, Dörr J, et al. Patterns of retinal nerve fiber layer loss in multiple sclerosis patients with or without optic neuritis and glaucoma patients. *Clin Neurol Neurosurg*. 2010;112(8):647-652.
20. Motamedi S, Gawlik K, Ayadi N, et al. Normative data and minimally detectable change for inner retinal layer thicknesses using a semi-automated OCT image segmentation pipeline. *Front Neurol*. 2019;10:1117.
21. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
22. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167-175.
23. Kauer-Bonin J, Yadav SK, Beckers I, et al. Modular deep neural networks for automatic quality control of retinal

- optical coherence tomography scans. *Comput Biol Med.* 2022;141:104822.
24. Yadav SK, Kafieh R, Zimmermann HG, et al. Deep learning based intraretinal layer segmentation using cascaded compressed U-Net. medRxiv. 2021:2021.11.19.21266592.
 25. Schippling S, Balk LJ, Costello F, et al. Quality control for retinal OCT in multiple sclerosis: validation of the OSCAR-IB criteria. *Mult Scler.* 2015;21(2):163-170.
 26. Aytulun A, Cruz-Herranz A, Aktas O, et al. APOSTEL 2.0 recommendations for reporting quantitative optical coherence tomography studies. *Neurology.* 2021;97(2):68-79.
 27. Petzold A, Albrecht P, Balcer L, et al. Artificial intelligence extension of the OSCAR-IB criteria. *Ann Clin Transl Neurol.* 2021;8(7):1528-1542.
 28. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2018. Accessed March 21, 2022. <https://www.R-project.org/>
 29. Yu F, Koltun V, Funkhouser T. Dilated residual networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2017:472-480.
 30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016:770-778.
 31. Nolan RC, Galetta SL, Frohman TC, et al. Optimal intereye difference thresholds in retinal nerve fiber layer thickness for predicting a unilateral optic nerve lesion in multiple sclerosis. *J Neuroophthalmol.* 2018;38:451-458.
 32. Nolan-Kenney RC, Liu M, Akhand O, et al. Optimal intereye difference thresholds by optical coherence tomography in multiple sclerosis: an international study. *Ann Neurol.* 2019;85(5):618-629.
 33. Coric D, Balk LJ, Uitdehaag BMJ, Petzold A. Diagnostic accuracy of optical coherence tomography inter-eye percentage difference for optic neuritis in multiple sclerosis. *Eur J Neurol.* 2017;24(12):1479-1484.
 34. Petzold A, Chua SYL, Khawaja AP, et al. Retinal asymmetry in multiple sclerosis. *Brain.* 2021;144(1):224-235.
 35. Kenney RC, Liu M, Hasanaj L, et al. The role of OCT criteria and machine learning in multiple sclerosis and optic neuritis diagnosis. *Neurology.* 2022;10.1212/WNL.0000000000200883.
 36. Brandt AU, Specovius S, Oberwahrenbrock T, Zimmermann HG, Paul F, Costello F. Frequent retinal ganglion cell damage after acute optic neuritis. *Mult Scler Relat Disord.* 2018;22:141-147.
 37. Outteryck O, Lopes R, Drumez É, et al. Optical coherence tomography for detection of asymptomatic optic nerve lesions in clinically isolated syndrome. *Neurology.* 2020;95(6):e733-e744.
 38. Vidal-Jordana A, Rovira A, Arrambide G, et al. Optic nerve topography in multiple sclerosis diagnosis. *Neurology.* 2021;96(4):e482-e490.
 39. Caldito NG, Saidha S, Sotirchos ES, et al. Brain and retinal atrophy in African-Americans versus Caucasian-Americans with multiple sclerosis: a longitudinal study. *Brain.* 2018;141(11):3115-3129.