SUPPLEMENTARY TABLES

Supplementary Table 1: Additional statistics with age as a covariate.

Fig.	Comparison	p-Value		Comparison	p-Value	
		w/o age	with age		w/o age	with age
1A	CIS17 : MS17	0.001	0.005			
1C	BL_1-8 : BL_>8 BL_1-8 : FU_1-8 BL 1-8 : FU >8	<0.0005 <0.0005 0.522	<0.0005 <0.0005 0.284	BL_>8 : FU_1-8 BL_>8 : FU_>8 FU 1-8 : FU >8	<0.0005 <0.0005 0.001	<0.0005 <0.0005 0.001
1D	BL_GD- : BL_GD+ BL_GD- : FU_GD- BL_GD- : FU_GD+	<0.0005 <0.0005 0.818	<0.0005 <0.0005 0.473	BL_GD+ : FU_GD- BL_GD+ : FU_GD+ FU_GD- : FU_GD+	<0.0005 <0.0005 <0.0005	<0.0005 <0.0005 <0.0005
1E	w/o age: r = -0.170 with age: r = -0.156	<0.0005	<0.0005			
1F		0.015	0.046			
1G	w/o age: r = 0.092 with age: r = 0.090	0.008	0.011			
2B		<0.0005	<0.0005			
2C		0.036	0.029			
2D		0.035	0.034			
2E		<0.0005	<0.0005			
3B		0.002	0.002			
3C	w/o age: r = 0.223 with age: r = 0.219	<0.0005	<0.0005			
3D	high : moderate high : basic high : no DMT	<0.0005 <0.0005 <0.0005	<0.0005 <0.0005 <0.0005	moderate : basic moderate : no DMT basic : no DMT	0.179 0.102 0.914	0.259 0.316 1.000
3E	high : moderate high : basic high : no DMT	0.4 0.001 0.001	0.930 0.008 0.011	moderate : basic moderate : no DMT basic : no DMT	0.450 0.163 1.000	0.461 0.393 1.000
4B	@BL @FU2 @FU4	<0.0005 0.007 0.128	<0.0005 <0.0005 0.092			
4C	[top] w/o age: r = 0·179 with age: r = 0·176	<0.0005	<0.0005	[bottom] w/o age: r = 0.133 with age: r = 0.152	0.001	<0.0005
4D	[FU2] escalation : no switch escalation : de-escalation no switch : de-escalation	0.001 0.026 0.966	<0.0005 0.022 1.000			
4E	@BL @FU2 @FU4	0.035 0.025 0.204	0.040 0.030 0.183			

w/o age: age was not considered as a covariate; with age: statistical analysis was performed by considering age as a covariate; CIS17: clinically isolated syndrome [2017]; MS17: relapsing remitting multiple sclerosis [2017]; BL: baseline; FU: two-year follow-up; sNfL: serum neurofilament light chain; 1-8: 1-8 T2 lesions observed in the MRI scan; >8: more than 8 T2 lesions in the MRI; GD+: gadolinium enhancing lesions in the MRI scan; no DMT, basic, moderate, high: different treatment groups. P-values were obtained by analysing the data using the Mann-Whitney-U-Test, Kruskal-Wallis-Test, mixed linear models, two-way mixed ANOVA and one-way ANOVAs. Bonferroni-adjustment or Tukey's post-hoc test was applied to correct for multiple comparison. Correlation analyses were performed by Spearman's Rank Correlation and partial non-parametric correlations for covariates.

Supplementary Table 2: Baseline data of reclassified CIS[2017] patients and dropouts.

Variable		Dropouts	Reclassified	p-Value
n		111	258	
		Median (IQR)		
Age (years)		33 (25-44)	33 (26-41)	0.481[a]
Disease duration (months)		2 (1-3)	2 (1-2)	0.411[a]
		n (%)		p-Value
Sex	male	37 (33.3%)	76 (29.5%)	0.459[b]
	female	74 (66.7%)	182 (70.5%)	
T2-lesion count	1-8	34 (30.6%)	76 (29.8%)	0.874[b]
	> 8	77 (69.4%)	179 (70.2%)	

IQR: interquartile range; CIS: clinically isolated syndrome; sNfL: serum neurofilament light chain; Mann-Whitney-U tests (a) were conducted to compare group differences. Distributions were compared using a chi-square tests of homogeneity (b).

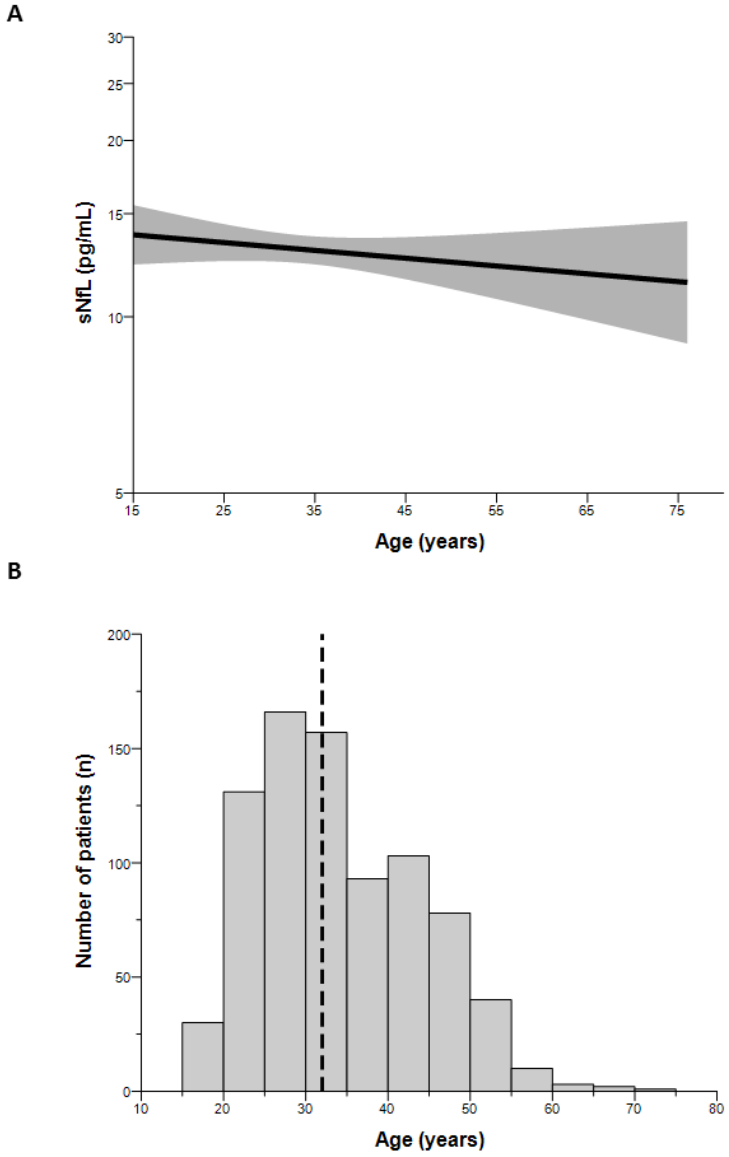
Supplementary Table 3: Additional support vector machine (SVM) analyses: OCBs or Gd+ lesions were analyzed without sNfL and with different sNfL percentiles.

sNfL percentile	number of cases (total n = 258)	sNfL cutoff (pg/ml)	% sensitivity [95% CI]	% specificity [95% CI]	% accuracy [95% CI]	% PPV	% NPV
-		-	71.9 [62.3, 80.5]	75·9 [64.9, 84.4]	79·4 [68.4, 88.6]	48.9	88.4
50th	136	10.1	66.2 [57.4, 74.8]	68·3 [59.2, 77.2]	72·7 [61.5, 82.7]	51.8	81.3
60th	114	12.0	69.2 [60.2, 78.3]	71·3 [62.5, 82.1]	75·6 [64.7, 86.1]	48.7	91.4
70th	85	14.8	72·2 [63.5, 81.1]	74·6 [65.7, 83.9]	77·2 [66.5, 88.5]	50.2	91.3
80th	57	20.0	75·1 [64.8, 84.7]	72·4 [61.8, 81.2]	78·2 [67.8, 87.5]	46.8	90.2

Specificity (%) = TN/(TN+FP) * 100
Sensitivity (%) = TP/(TP+FN) * 100
Accuracy (%) = (TP+TN)/T * 100
PPV = TP/(TP+FP) *100
NPV = TN/(TN+FN) *100

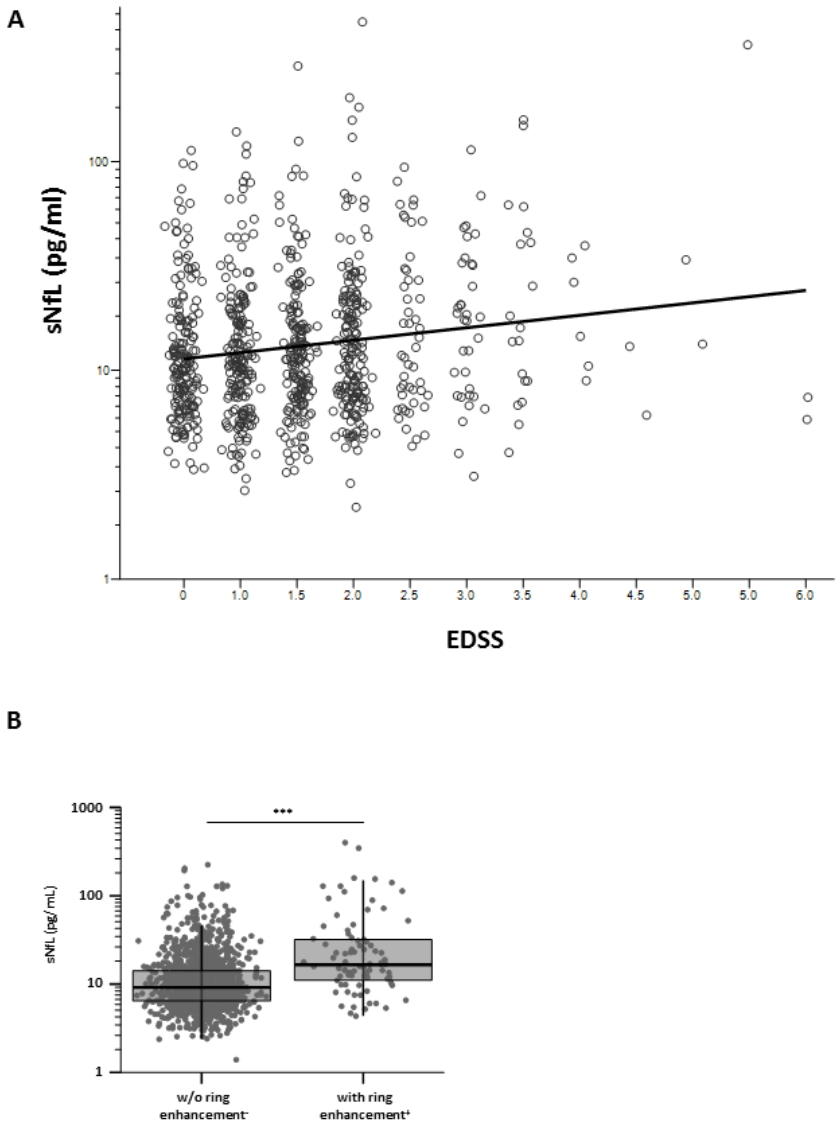
The entities in the above equations are: (TN (true negatives), TP (true positives), FN (false negatives), FP (false positives), PPV (positive predictive value), NPV (negative predictive value)), and T is the total number of data under test. CI: confidence interval, OCB: oligoclonal bands; Gd^+ lesions: gadolinium enhancing MRI lesions; sNfL: serum neurofilament light chain.

SUPPLEMENTARY FIGURES



Supplementary Fig. 1

Supplementary Figure 1. Relation between age and sNfL values. A) In our young cohort, no correlation was observed between sNfL and age ($r = -0.044$, $p = 0.211$). Correlation was determined by Spearman's rank correlation coefficient and is plotted with the 95% confidence interval. B) Age distribution in our cohort. Dotted line marks median age (32 years).



Supplementary Fig. 2

Supplementary Figure 2. Relation between EDSS and ring-enhancing lesions and sNfL. A) Significant weak correlation between sNfL levels and EDSS values ($r = 0.130$, $p < 0.0005$). Correlation was determined by Spearman's rank correlation coefficient. B) sNfL levels in patients with ($n = 83$) and without ring-enhancing gadolinium lesions ($n = 1335$, $p < 0.0005$, $\eta^2 = 0.18$). Group differences were analysed by Mann-Whitney-U-Test.

Supplementary Material and Methods

Receiver operation characteristic (ROC) and support vector machine (SVM) analyses

A classifier is a parameter with a suitable optimal threshold that is used in a classification algorithm. In this study we applied binary classification algorithms (i.e. testing the classification between the two groups termed "RRMS" and "CIS"). The performance of a classifier is evaluated by three main metrics, namely specificity, sensitivity and accuracy. Specificity indicates the ability of a classifier to detect negative cases. Sensitivity represents the ability of a classifier to detect the positive cases. Accuracy represents the overall performance of a classifier. It indicates the percentage of correctly classified positive and negative cases among the total number of cases.

Specificity (%) = TN/(TN+FP) * 100
Sensitivity (%) = TP/(TP+FN) * 100
Accuracy (%) = (TP+TN)/T * 100
PPV = TP/(TP+FP) *100
NPV = TN/(TN+FN) *100

The entities in the above equations are: TN (true negatives), TP (true positives), FN (false negatives), FP (false positives), PPV (positive predictive value), NPV (negative predictive value); T is the total number of data under test. Receiver operating characteristic (ROC) analysis was used as a linear classifier to determine the three parameters mentioned above.

We have used the support vector machine (SVM) analysis[1-3] which is a non-linear classifier in addition to validate the linear classification results. For the SVM analysis, the algorithm looked for an optimally separating threshold between the two data sets by maximising the margin between classes' closest points. The general notion on SVM classifiers would favor the larger class. In general, the majority class will have a high accuracy in prediction (sensitivity if the positive class is the majority and specificity if the negative class is the majority) and the minority class will have a low accuracy. These procedures are not useful for our applications here in this study. A main challenge in the class-imbalanced classification is to develop a classifier that can provide good accuracy for the minority class prediction[4,5]. A class-imbalanced classifier typically modifies a standard classifier by a correction strategy or by incorporating a new strategy in the training phase to account for differential class sizes. Therefore, we used the SVM-based correction classifier which is a SVM threshold adjustment (SVM-THR)[6]. To have robust threshold adjustment, we used the 10-fold cross validation approach for defining the threshold for each of the estimation. The open source machine learning library LIBSVM[7] was used as a toolbox. Specially, a composite risk score consisting of the factors presence or absence of Gd+ lesions and presence of 1-8 T2 lesions or more was calculated. Next, we added sNfL concentrations in the risk score and tested the prediction of EDSS and MSFC at two-year follow-up. In another approach, we tested the CIS[2010] patients and classified them according to 2017 McDonald criteria based on the information of the further variables (i.e. OCBs and Gd+ lesions, defined as response variable). We then added sNfL (>90th percentile) as an additional variable and repeated the SVM analysis for testing the classification.

References:

1. Michels L, Muthuraman M, Anwar AR, et al. Changes of Functional and Directed Resting-State Connectivity Are Associated with Neuronal Oscillations, ApoE Genotype and Amyloid Deposition in Mild Cognitive Impairment. *Front Aging Neurosci* 2017; **9**: 304.
2. Siller N, Kuhle J, Muthuraman M, et al. Serum neurofilament light chain is a biomarker of acute and chronic neuronal damage in early multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England)* 2018; **25**(5): 678-86.
3. Muthuraman M, Fleischer V, Kolber P, Luessi F, Zipp F, Groppa S. Structural Brain Network Characteristics Can Differentiate CIS from Early RRMS. *Front Neurosci* 2016; **10**: 14.
4. Provost F, Fawcett T. Robust classification for imprecise environments. *Machine Learning* 2001; **42**(3): 203-31.
5. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. *Intelligent Data Analysis* 2002; **6**(5): 429-49.
6. Chen JJ, Tsai CA, Moon H, Ahn H, Young JJ, Chen CH. Decision threshold adjustment in class prediction. *SAR and QSAR in environmental research* 2006; **17**(3): 337-52.
7. Chang CC, Lin CJ. LIBSVM: A Library for Support Vector Machines. *Acm T Intel Syst Tec* 2011; **2**(3).