AJNR

Overcoming the Clinical–MR Imaging Paradox of Multiple Sclerosis: MR Imaging Data Assessed with a Random Forest Approach

K. Kac?ar, M.A. Rocca, M. Copetti, S. Sala, S. Mesaros, T. Stosic Opincal, D. Caputo, M. Absinta, J. Drulovic, V.S. Kostic, G. Comi and M. Filippi

K. Kačar
M.A. Rocca
M. Copetti
S. Sala
Š. Mesaroš
T. Stosić Opinćal
D. Caputo
M. Absinta
J. Drulović
V.S. Kostić
G. Comi
M. Filippi

Overcoming the Clinical−MR Imaging Paradox of Multiple Sclerosis: MR Imaging Data Assessed with a Random Forest Approach

BACKGROUND AND PURPOSE: In MS, the relation between clinical and MR imaging measures is still suboptimal. We assessed the correlation of disability and specific impairment of the clinical functional system with overall and regional CNS damage in a large cohort of patients with MS with different clinical phenotypes by using a random forest approach.

MATERIALS AND METHODS: Brain conventional MR imaging and DTI were performed in 172 patients with MS and 46 controls. Cervical cord MR imaging was performed in a subgroup of subjects. To evaluate whether MR imaging measures were able to correctly classify impairment in specific clinical domains, we performed a random forest analysis.

RESULTS: Between-group differences were found for most of the MR imaging variables, which correlated significantly with clinical measures (r ranging from −0.57 to 0.55). The random forest analysis showed a high performance in identifying impaired versus unimpaired patients, with a global error between 7% (pyramidal functional system) and 31% (Ambulation Index) in the different outcomes considered. When considering the performance in the unimpaired and impaired groups, the random forest analysis showed a high performance in identifying patients with impaired sensory, cerebellar, and brain stem functions (error below 10%), while it performed poorly in defining impairment of visual and mental systems (error of 91% and 70%, respectively). In analyses with a good level of classification, for most functional systems, damage of the WM fiber bundles subserving their function, measured by using DTI tractography, had the highest classification power.

CONCLUSIONS: Random forest analysis, especially if applied to DTI tractography data, is a valuable approach, which might contribute to overcoming the MS clinical−MR imaging paradox.

ABBREVIATIONS: BMS = benign MS; CC = corpus callosum; CST = corticospinal tract; EDSS = Expanded Disability Status Scale; FA = fractional anisotropy; GM = gray matter; MCP = middle cerebellar peduncle; MD = mean diffusivity; MPRAGE = magnetization-prepared rapid acquisition of gradient echo; NAWM = normal-appearing white matter; NBV = normalized brain volume; PPMS = primary-progressive MS; RRMS = relapsing-remitting MS; SCP = superior cerebellar peduncle; SPMS = secondary-progressive MS

In MS, MR imaging is the most valuable paraclinical tool to monitor disease progression, either natural or modified by treatment.[1,2] Nevertheless, the correlation between clinical and MR imaging measures is still suboptimal.[1,2]

Several factors have been considered to explain the so-called clinical−MR imaging paradox of MS: First, the intrinsic limitations of the clinical scale most commonly applied to measure disability in MS, the EDSS, an ordinal scale ranging from 0 (normal neurologic examination) to 10 (death), which, for values higher than 4.0, is strongly weighted toward impairment of deambulation[3]; second, the poor specificity of conventional MR imaging measures for the most destructive aspects of MS pathology[4]; and third, the inability of conventional MR imaging to detect and grade the involvement of structures, including the GM, spinal cord[2] and strategic WM fiber bundles, such as the CST[5] and the CC,[6] which might be critical for the development of irreversible clinical deficits or for the impairment of specific functional domains.

During the past 2 decades, much effort has been devoted to overcoming the clinical−MR imaging paradox of MS. This includes the development of new clinical scales (eg, the Multiple Sclerosis Functional Composite Scale),[7] which provide a better estimation than the EDSS of the overall clinical impairment by assessing other domains affected by the disease, such as cognition and upper limb function, and the application of modern MR imaging techniques, which are sensitive to the heterogeneous pathologic substrates of MS, to quantify disease-related abnormalities in different CNS compartments.[1,2] More recently, advanced MR imaging techniques, including DTI tractography, have provided novel tools to investigate in

vivo the structural damage of "strategic" WM fiber bundles of the brain.[8] Such an effort has undoubtedly ameliorated the strength of the correlation between clinical and MR imaging quantities in MS,[1,2] but which MR imaging measures correlate best with a patient's clinical status and whether regional rather than global CNS damage plays a critical role in determining disability still remain to be determined.

In this study, we collected conventional and quantitative MR imaging data of the brain and spinal cord from a large sample of patients with MS, including all the major clinical phenotypes of the disease, to assess the correlation of disability and impairment of clinical functional systems with involvement of the entire CNS or damage to specific CNS structures at different stages of the disease. To this end, we applied a random forest approach,[9] which, by permuting and bootstrapping the observations, uses a set of classification trees for obtaining robust estimates of the associations among different variables.

Materials and Methods

Subjects

We recruited consecutively 172 patients with MS (women/men = 110:62; mean age = 43.4 years, range = 19–70 years; mean disease duration = 12.2 years, range = 0.01–40 years; median EDSS score = 4.0, range = 0.0–8.5) and 46 age- and sex-matched healthy controls (women/men = 29:17; mean age = 40.8 years, range = 21–69 years) with no previous history of neurologic dysfunction and normal findings on neurologic examinations.

The subjects were selected from 2 different centers (70 from center A and 102 from center B). Twenty-two patients had clinically isolated syndromes with clinical and MR imaging features suggestive of MS[10]; 51, RRMS[11]; 44, SPMS[11]; 20, BMS (EDSS score of ≤3.0 and disease duration of ≥15 years)[12,13]; and 35 patients had PPMS.[14]

All patients underwent a complete neurologic examination within 2 days of the MR imaging study, with rating of the EDSS score[3] and evaluation of the degree of impairment in the different functional systems of the EDSS. Corticosteroid treatment for clinical relapses had to be terminated at least 4 weeks before MR imaging.

The demographic and clinical characteristics of the different cohorts of patients studied are presented in On-line Tables 1 and 2.

Local ethics committee approval and written informed consent from each subject were obtained.

MR Imaging

Conventional MR imaging and DTI scans of the brain were acquired from all subjects by using identical 1.5T magnetic field strengths (Avanto; Siemens, Erlangen, Germany). The following MR imaging sequences of the brain were performed in all subjects: 1) axial dual-echo TSE; 2) axial pulsed gradient spin-echo echo-planar diffusion MR imaging; and 3) sagittal 3D T1-weighted MPRAGE. In center B, we also performed the following pulse sequences of the cervical cord: 1) sagittal T2-weighted TSE; 2) sagittal 3D T1-weighted MPRAGE; and 3) axial 2D gradient-echo with and without an off-resonance radio-frequency saturation. The acquisition parameters of the MR imaging sequences used are described in the On-line Appendix.

MR Imaging Postprocessing

Brain T2 lesion loads were quantified. NBV was measured by using Structural Imaging Evaluation of Normalized Atrophy software (http://www.fmrib.ox.cc.uk/fsl/siena/index.html).[15] From diffusion-weighted images, the diffusion tensor was estimated for each voxel,[16] and MD and FA maps were derived.[17] The average MD of the NAWM and GM and FA of the NAWM were measured.[18]

An atlas-based automated approach was used to obtain DTI-derived measures of brain WM tracts.[5,19,20] This procedure involved the following steps: 1) the production of a reference FA image in the standard space (the FA atlas) by using a reference group of 24 healthy subjects (15 women and 9 men, mean age = 31.8 years, range = 21–40 years); 2) the definition of probability maps of the CC, CST, thalamocortical connection, inferior fronto-occipital fasciculus, uncinate fasciculus, cingulum, arcuate fasciculus, inferior longitudinal fasciculus, optic radiation, SCP, and MCP on the FA atlas; 3) the nonlinear alignment of individual subject's MD and FA maps to the FA atlas; and 4) the application of the WM tract probability maps to the individual subject's images to measure mean tract MD and FA values. The use of a nonlinear transformation algorithm and of FA images to drive the transformation (with pieces of information specific to each tract morphology) allowed an optimal overlap between single subject data and the atlas to be obtained. Furthermore, major errors were excluded during coregistration, because though the process is completely automatic, all data were visually checked during all steps of the analysis.

Volumes of T2 visible lesions in the different WM fiber bundles were derived by applying fiber bundle probability maps, obtained by using DTI tractography, and calculating the volume of lesions located inside each.[5] Figure 1 shows an example of WM fiber bundle reconstruction in an individual subject before constructing the probability maps.

Cervical cord hyperintense lesions were quantified. Cord cross-sectional area and average magnetization transfer ratio values were measured.[21,22] The details of MR imaging data analysis are described in the On-line Appendix.

Statistical Analysis

Between-group differences were assessed by using the nonparametric Kruskal-Wallis test. Correlations between clinical and MR imaging−derived quantities were estimated by using the Spearman rank correlation coefficient.

Random forest analysis was performed to classify clinically impaired-versus-unimpaired patients by using a set of MR imaging covariates (including measures of global brain and cord damage as well as selective damage to brain WM fiber bundles) as detailed in the On-line Appendix. Because the number of observations for each value of the functional systems and EDSS scales was small, potentially leading to unreliable results, we dichotomized the functional systems into impaired (≥1) and nonimpaired (0) and the EDSS scale according to a cutoff of 4.0 (which identifies fully ambulatory patients). According to the random forest technique, 100 000 trees were built to classify compromised patients.[9] The training set used to grow each tree was a 0.632+ bootstrap resample of the observations.[23] Trees were allowed to grow to their full size without pruning; nodes with at least 1 event and a minimum total size of 5 were used as stopping rules. The best split at each node was selected from a random subset of covariates. The left-out observations (ie, "out of bag" observations) were then predicted to obtain the classification error rate of the considered tree. We assessed the predictive ability of the classifier, aggregating the single-tree error rates. Furthermore, the random forest framework allowed us to estimate the importance of a variable by looking at how much the classification error increases when "out of bag" data for that variable are permuted while all others are left unchanged. We followed Strobl et al[24] to avoid possible bias in variable selection: Indi-
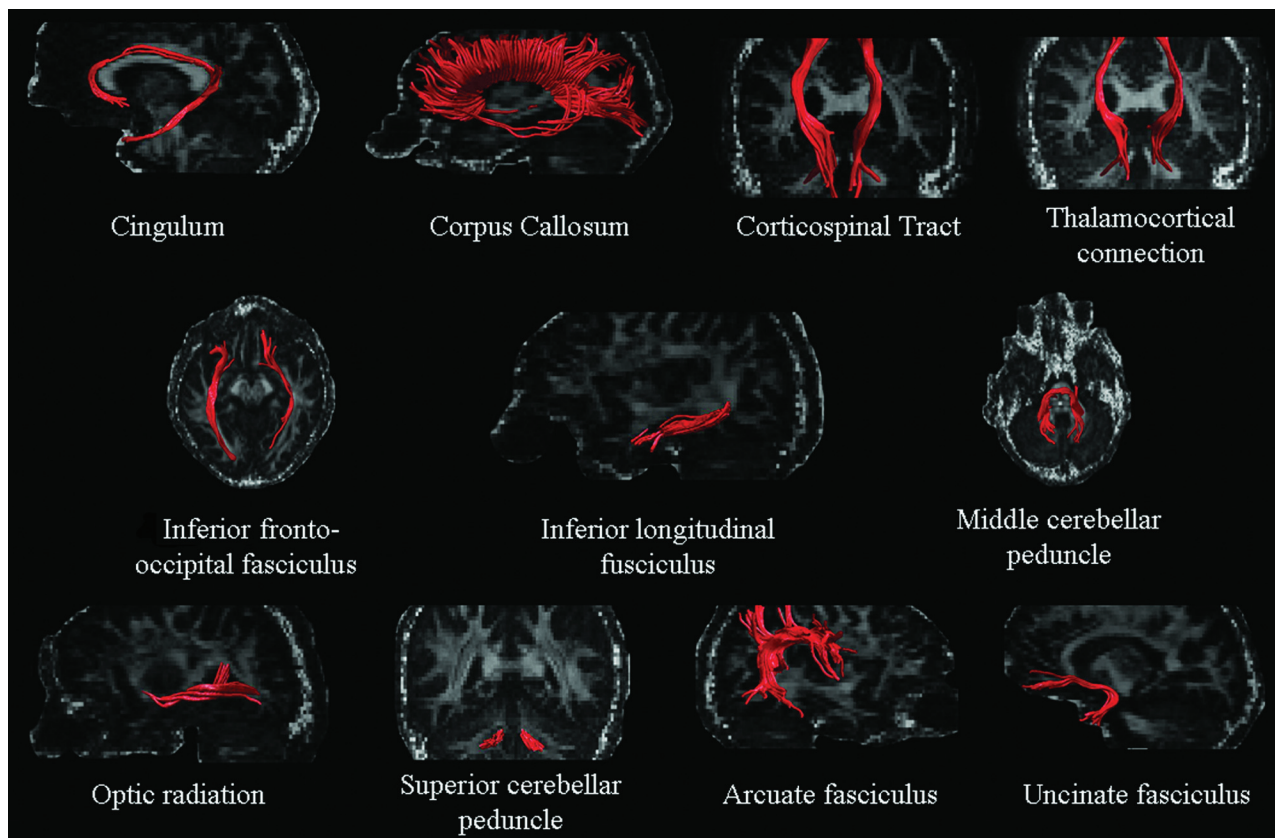
Fig 1. Illustrative behavior of the main WM fiber bundles studied in a single subject, before spatial normalization. See text for further details.

vidual classification trees were built by using subsampling without replacement and adopting a conditional permutation scheme.[25]

A *P* value <.05 was considered significant. All analyses were performed by using SAS Release 9.1 (SAS Institute, Cary, North Carolina). For the random forest analysis, we used the package randomForest, Version 4.5 implemented in R software (http://cran.r-project.org/web/packages/randomForest/index.html).

Results

On-line Table 3 summarizes the main structural MR imaging findings from the different study groups. Most of the MR imaging measures were significantly different among groups. Several correlations were found between EDSS and impairment of individual functional systems and MR imaging measures of the following: 1) global brain damage (*r* ranging from −0.57 to 0.51, *P* values ranging from .04 to <.0001), 2) damage to specific WM fiber bundles (*r* ranging from −0.51 to 0.55, *P* values ranging from .05 to <.0001), and 3) cervical cord damage (*r* ranging from −0.48 to 0.26, *P* values ranging from .03 to <.0001). None of the MR imaging quantities analyzed was correlated with impairment of the sensory functional system.

The results of the random forest analysis in the whole sample of patients and in the different patient subgroups are reported in On-line Table 4. For each clinical end point, the global error of the random forest estimation and the error in unimpaired and impaired patients, separately, are provided. A list of the first 4 most important variables that classified patients correctly is also given. Empty cells are due to the absence of a group of patients (eg, 100% impaired and 0% unimpaired).

In the entire sample of patients, the random forest analysis showed relatively low global error rates in classifying correctly impaired and unimpaired patients according to the EDSS scale and to the different functional system subscales (global error range from 7.5% to 30.7%). When we considered the performance in impaired and unimpaired patients separately, the random forest analysis achieved a good fit in the classification of impaired patients (with the exception of the visual and mental functional system), whereas it had heterogeneous performances in identifying unimpaired patients (with best performance for the visual and worst performance for the pyramidal functional system). In the different clinical disease phenotypes, the random forest analysis showed heterogeneous performance. Low performances were observed when the design was strongly unbalanced (On-line Table 4).

The ranking of variable importance showed that, for most of the functional systems, damage to the WM fiber bundles subserving their functions had the highest classification accuracy (eg, Ambulation Index versus cord area and cerebellar damage; cerebellar function versus CST and SCP damage).

The random forest analysis in patients recruited in center B confirmed the results obtained in the whole sample (data not shown). Such an analysis, resulted in better error estimation (with reliable ranking of variable importance) for the following subgroup analysis:

- Pyramidal functional system in patients with clinically isolated syndromes: global error = 15%, unimpaired error = 25%, impaired error = 12.5%; variable importance: CST average lesion FA, CST average lesion MD, CC NAWM MD, CC NAWM FA.

- Visual functional system in patients with SPMS: global error = 40.0%, unimpaired error = 26.7%, impaired error = 80.0%; variable importance: NAWM MD, CC NAWM MD, OR NAWM MD, NBV.
- Sensory functional system in patients with SPMS: global error = 31.6%, unimpaired error = 80.0%, impaired error = 14.3%; variable importance: NAWM FA, SCP NAWM MD, CC NAWM MD, MCP NAWM MD.

Discussion

We applied a random forest analysis to assess whether MR imaging measures of CNS structural damage could contribute to correctly classifying patients with MS with or without clinical disability, as determined by the EDSS scale, and with or without impairment to the different functional systems of this scale.

Most previous studies lacked a large-enough sample sizes; typically, only a restricted selection of MR imaging measures was considered as data input.[1,2] The random forest approach is an established highly accurate classifier, which can handle a large number of input variables even when the number of observations is small. Using a different bootstrap sample of the data and a different subset of predictors, chosen randomly to construct each tree of the forest, it overcomes false-positive discoveries. Unlike a classic correlation analysis, no adjustment for multiple comparisons is needed. Furthermore, it estimates the importance of variables in the classification by using the permuted "out of bag" observations: By introducing an appropriate level of randomness, it produces accurate estimates of associations handling the problem of correlated predictors and showing which, among them, is the best one.[9] The same is not possible by using other standard statistical procedures. Possible bias in variable selection was present in the first version of random forest analysis when the predictors varied in their scale of measurement or their number of categories and when the predictors were highly correlated.[24] According to recent developments,[25] such problems were avoided in our study by building individual classification trees by using subsampling without replacement and adopting a conditional permutation scheme, where each predictor is permuted only within selected subgroups of observations to preserve the correlation structure between the permuted predictor and the other variables.

With this new conditional permutation scheme, the importance measure can reveal the spurious correlations and account for them.

The combination of all the available MR imaging techniques for the quantification of macro- and microscopic MS-related damage to the brain and spinal cord in a single study would have required an extremely long acquisition time and would have been hampered by problems with patient compliance. As a consequence, we, a priori, chose a set of sequences to be applied, on the basis of their ability to provide information on different aspects of MS pathology. Our MR imaging acquisition scheme included the following: T2-weighted scans, which are sensitive to macroscopic MS-related abnormalities[26]; T1-weighted scans, which allow measuring atrophy and, as a consequence, provide an estimate of irreversible tissue damage[26]; DTI, which allows quantifying injury to critical WM pathways[8]; and magnetization transfer MR imaging, which reliably quantifies the extent of MS damage to the cervical cord.[22]

In line with the results of many previous MR imaging reports in MS,[1,2] we found significant between-group differences for most of the MR imaging variables studied. As expected, considering the number of variables included, the classic analysis showed poor-to-moderate correlations between clinical EDSS scores and some of the MR imaging measures of structural damage to the brain and spinal cord. Similar results were obtained when considering the impairment of the different functional systems, with the exception of the sensory one, where impairment was not related to any MR imaging measure.

The random forest analysis provided several additional important pieces of information from this dataset. First, by using a set of MR imaging covariates, we were able to classify, with relatively low error rates, impaired-versus-unimpaired patients according to their EDSS and different functional system scores, with a global error between 7% (for the pyramidal functional system) and 31% (for the Ambulation Index) in all the different outcomes considered, indicating that such an approach could classify correctly a proportion between 70% and 93% of our patients.

When we considered the performance in the unimpaired and impaired groups separately, in the entire sample of patients, the random forest analysis showed an extremely high performance in identifying patients with impairment of sensory, cerebellar, and brain stem functions (with an error below 10%), while it performed poorly in defining impairment of the visual and mental systems (with an error of 91% and 70%, respectively) as well as in the pyramidal functional system. This unequal behavior of patient classification is likely due to the unbalanced degree of impairment in the different functional systems (in this case, only an increase of the sample size would allow improving the performance). Indeed, only a few patients experienced a clinical impairment of the visual and mental functional system, while most had pyramidal functional system impairment according to our criteria. Admittedly, the rating of these 2 functions was performed during a routine neurologic assessment in the proximity of the MR imaging study. As a consequence, we cannot exclude the possibility that a more precise estimation of impairment of these 2 functional systems through a neuro-ophthalmologic and a neuropsychological evaluation would have resulted in a better classification. The same considerations also apply to the results obtained from the random forest analysis in the different disease phenotypes. Clearly, these results are strongly influenced by the dichotomization we applied. Such an approach resulted in inestimable error rates when we had only clinically impaired or unimpaired patients (eg, all patients with BMS had an EDSS below 3.0, and all patients with SPMS and PPMS had locomotor impairment). Similarly, we had high error rates in case of strongly unbalanced groups (eg, most patients with clinically isolated syndromes had low EDSS scores and most patients with PPMS had high EDSS scores).

The random forest analysis also enabled us to define the importance of variables, which can contribute to better defining the relationship between clinical and MR imaging−derived quantities. When considering the results obtained in the entire group of patients with MS, this analysis showed that for most of the functional systems, clinical impairment was associated with a structural injury to the CNS structures strictly related to the function of the system of interest. For instance, the Ambulation Index was associated with

cord and cerebellar damage; brain stem functional system impairment, with MCP damage; and cerebellar functional system impairment, with SCP and CST damage. For most of the functional systems (except for bowel and bladder, and visual functional systems), MR imaging measures of global brain damage, including T2 lesion load, NBV, and NAWM average FA, were also found to have a role in determining a patient's clinical status. These findings suggest that disability in this disease might result from both an injury to specific CNS structures/pathways and the extent of overall damage to the entire brain. However, our classification was based on data from a mixed group of patients with MS with different disease clinical phenotypes. As a consequence, it is tempting to speculate that regional-versus-global structural MR imaging damage might play different roles in patients with MS according to the disease stage. This hypothesis is at least partially supported by the results obtained from variable classifications in the different disease phenotypes. Although such an analysis was limited by the presence of a few empty cells, it showed that in patients with clinically isolated syndromes, for most of the functional systems, damage to selected WM fiber bundles rather than overall CNS damage had the greatest impact on clinical status, whereas in those with BMS and PPMS, global and regional damage seemed to have a similar role.

The random forest approach also allowed us to define a correlation between structural MR imaging variables and clinical scales, which did not emerge during the classic statistical analysis. This was the case, for instance, for impairment of the sensory functional system, which was not related to any MR imaging measure in the classic analysis, while it was associated with thalamocortical connection damage, NAWM FA, and NBV when the random forest approach was used.

Despite these results suggesting that the random forest approach might be a valuable tool to overcome the clinical−MR imaging paradox of MS, our study is not without limitations. These include the absence of corrections for geometric distortions of DTI data, which might result in a mismatch with non-echo-planar imaging–based images, dichotomization of clinical impairment (which is likely to have influenced the random forest performance for the pyramidal functional system), the lack of proper neuropsychological and neuro-ophthalmologic assessment, and the relatively small number of patients classified according to their disease phenotypes. In this perspective, none of the variables considered showed a significant association with the variable "center," suggesting that the application of this approach in large multicenter datasets might be a valuable strategy to be applied in future studies.

Conclusions

Random forest analysis, especially if applied to DTI tractography data, is a valuable approach, which might contribute to overcoming the MS clinical−MR imaging paradox.

References

1. Bakshi R, Thompson AJ, Rocca MA, et al. **MRI in multiple sclerosis: current status and future prospects.** *Lancet Neurol* 2008;7:615–25
2. Barkhof F, Calabresi PA, Miller DH, et al. **Imaging outcomes for neuroprotection and repair in multiple sclerosis trials.** *Nat Rev Neurol* 2009;5:256–66
3. Kurtzke JF. **Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS).** *Neurology* 1983;33:1444–52
4. Lassmann H. **The pathologic substrate of magnetic resonance alterations in multiple sclerosis.** *Neuroimaging Clin N Am* 2008;18:563–76, ix
5. Pagani E, Filippi M, Rocca MA, et al. **A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis.** *Neuroimage* 2005;26:258–65
6. Mesaros S, Rocca MA, Riccitelli G, et al. **Corpus callosum damage and cognitive dysfunction in benign MS.** *Hum Brain Mapp* 2009;30:2656–66
7. Cutter GR, Baier ML, Rudick RA, et al. **Development of a multiple sclerosis functional composite as a clinical trial outcome measure.** *Brain* 1999;122(pt 5):871–82
8. Rovaris M, Agosta F, Pagani E, et al. **Diffusion tensor MR imaging.** *Neuroimaging Clin N Am* 2009;19:37–43
9. Breiman L. **Random forests.** *Mach Learn* 2001;45:5–32
10. Polman CH, Reingold SC, Edan G, et al. **Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria."** *Ann Neurol* 2005;58:840–46
11. Lublin FD, Reingold SC. **Defining the clinical course of multiple sclerosis: results of an international survey—National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis.** *Neurology* 1996;46:907–11
12. Hawkins SA, McDonnell GV. **Benign multiple sclerosis? Clinical course, long term follow up, and assessment of prognostic factors**. *J Neurol Neurosurg Psychiatry* 1999;67:148–52
13. Pittock SJ, McClelland RL, Mayr WT, et al. **Clinical implications of benign multiple sclerosis: a 20-year population-based follow-up study.** *Ann Neurol* 2004;56:303–06
14. Montalban X, Sastre-Garriga J, Filippi M, et al. **Primary progressive multiple sclerosis diagnostic criteria: a reappraisal.** *Mult Scler* 2009;15:1459–65
15. Smith SM, De Stefano N, Jenkinson M, et al. **Normalized accurate measurement of longitudinal brain change.** *J Comput Assist Tomogr* 2001;25:466–75
16. Basser PJ, Mattiello J, LeBihan D. **MR diffusion tensor spectroscopy and imaging.** *Biophys J* 1994;66:259–67
17. Basser PJ, Pierpaoli C. **Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI.** *J Magn Reson B* 1996;111:209–19
18. Ceccarelli A, Rocca MA, Falini A, et al. **Normal-appearing white and grey matter damage in MS: a volumetric and diffusion tensor MRI study at 3.0 Tesla.** *J Neurol* 2007;254:513–18
19. Rocca MA, Pagani E, Absinta M, et al. **Altered functional and structural connectivities in patients with MS: a 3-T study.** *Neurology* 2007;69:2136–45
20. Rocca MA, Valsasina P, Ceccarelli A, et al. **Structural and functional MRI correlates of Stroop control in benign MS.** *Hum Brain Mapp* 2009;30:276–90
21. Losseff NA, Kingsley DP, McDonald WI, et al. **Clinical and magnetic resonance imaging predictors of disability in primary and secondary progressive multiple sclerosis.** *Mult Scler* 1996;1:218–22
22. Filippi M, Rocca MA. **Magnetization transfer magnetic resonance imaging of the brain, spinal cord, and optic nerve.** *Neurotherapeutics* 2007;4:401–13
23. Efron B, Tibshirani R. **Improvements on cross-validation: the .632+ bootstrap method.** *J Am Stat Ass* 1997;92:548–60
24. Strobl C, Boulesteix AL, Zeileis A, et al. **Bias in random forest variable importance measures: illustrations, sources and a solution.** *BMC Bioinformatics* 2007;8:25
25. Strobl C, Boulesteix AL, Kneib T, et al. **Conditional variable importance for random forests.** *BMC Bioinformatics* 2008;9:307
26. Filippi M, Rocca MA. **Conventional MRI in multiple sclerosis.** *J Neuroimaging* 2007;17(suppl 1):3S–9S