Performance Evaluation for Multiple Sclerosis Identification Models Based on MR Imaging and Machine Learning

W. Luo[1], L. Le[1,2], A. M. Ulug[1,3], A. Mazhari[4], N. Pinter[4], S. Magda[1], R. K. Haxton[1], R. Melton[1], C. Airriess[1]

[1]CorTechs Labs, San Diego; [2]University of California, San Diego ; [3]Bogazici University; [4]Dent Neurologic Institute

INTRODUCTION

- Previous studies have shown a high correlation between various brain atrophy, and physical and cognitive impairments in MS patients[1,2].
- Brain T1 MRI scans are part of the imaging procedure for diagnosis and follow-up process for MS patients[3].
- The objective of this study is to evaluate the performance of machine learning models based on features extracted from T1 images for identifying MS subjects and studying feature importance selected by models.

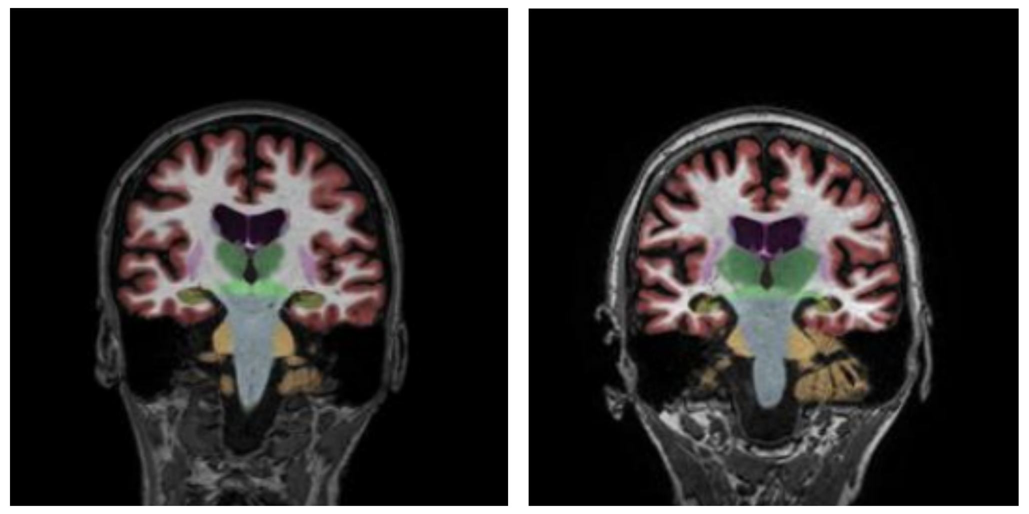


Figure 1. MS patient (left) vs. Normal subject (right). The color-coded NeuroQuant brain segmentation maps are shown

METHODS

- 3D T1- weighted scans acquired from normal and MS subjects processed by NeuroQuant 3.0 (CorTechs Labs Inc., San Diego) to generate volumetric brain information including the volume of cerebral white matter (WM) hypo-intensities.
- Random Forests algorithm [Figure 2] was used for creating the machine learning model. Brain structure volumes normalized by Intracranial Volume (ICV) was used as the input data.

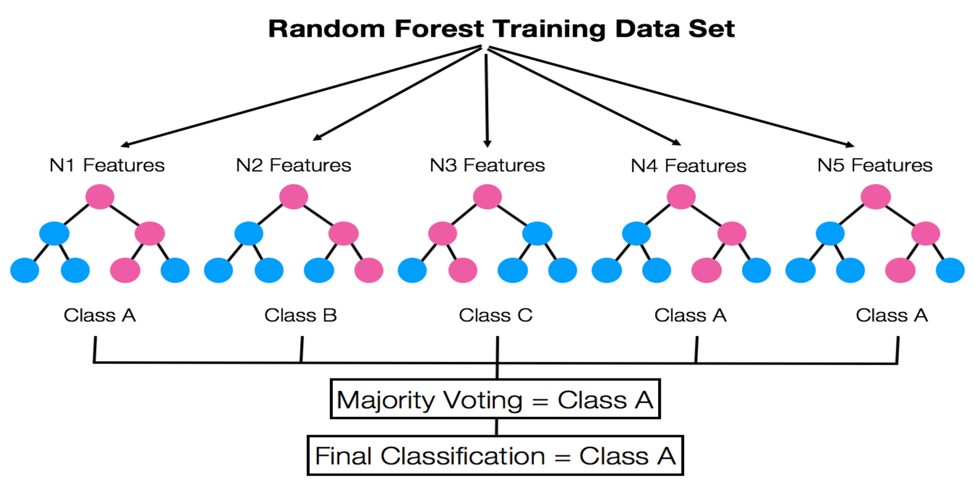


Figure 2. Random Forest Algorithm uses randomly selected feature to construct the decision tree. The final decision is based on the most votes

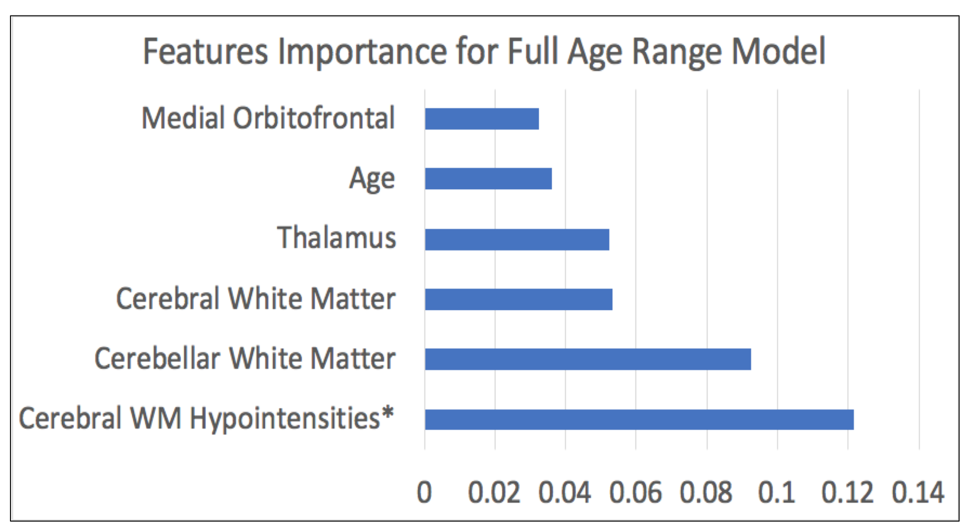
Groups	Male n (age range)	Female n (age range)
Normal	477(18-71, mean 40)	512 (18-71, mean 44)
MS	102(18-70, mean 43)	361(18-71, mean 42)

Table 1. Summary of Study Datasets

- One-third of randomly selected data were used for testing while the rest for training. The model was adjusted for imbalanced normal and MS data.
- The performances and the importance of the features selected by the model are evaluated for the full age range model and specific age range models.
- Visualization of the classification algorithm is based on Multiple Dimension Scaling (MDS) of proximity measurement.
- Four different individual test cases with the known diagnosis were used for visualization purpose.

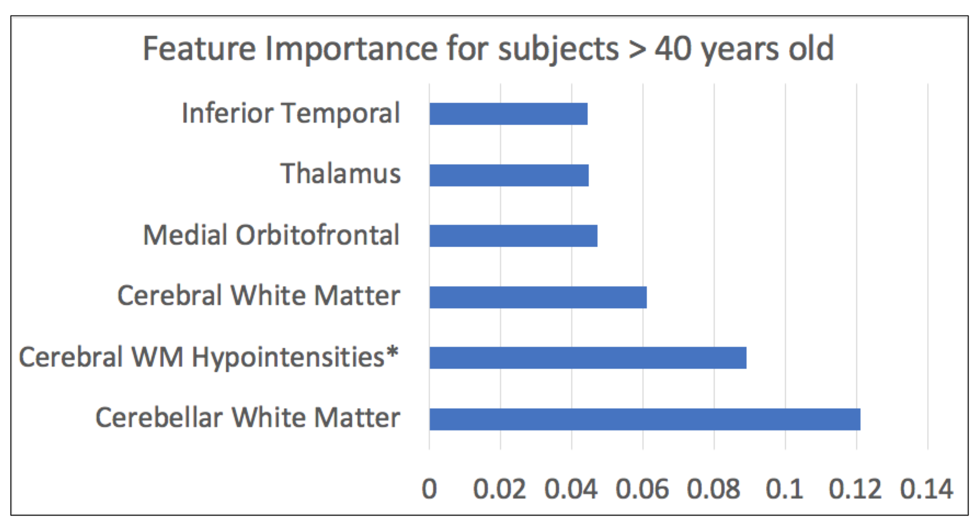
RESULTS

	Precision	Recall	F-1 Score
Normal	0.89	0.98	0.93
MS	0.93	0.74	0.82
Accuracy			0.90



Full age range model (mean AUC = 0.97)

	Precision	Recall	F-1 Score
Normal	0.89	0.96	0.93
MS	0.90	0.74	0.81
Accuracy			0.89



Specific age range model (mean AUC = 0.96)

Note: **Precision** is the ratio *tp / (tp + fp)* where *tp* is the number of true positives and *fp* the number of false positives. **Recall (or called sensitivity)** is the ratio *tp / (tp + fn)*. **F-1 score** can be interpreted as a weighted average of the precision and recall. **AUC** is the Area Under the Curve. The model performance is better when AUC is close to 1.

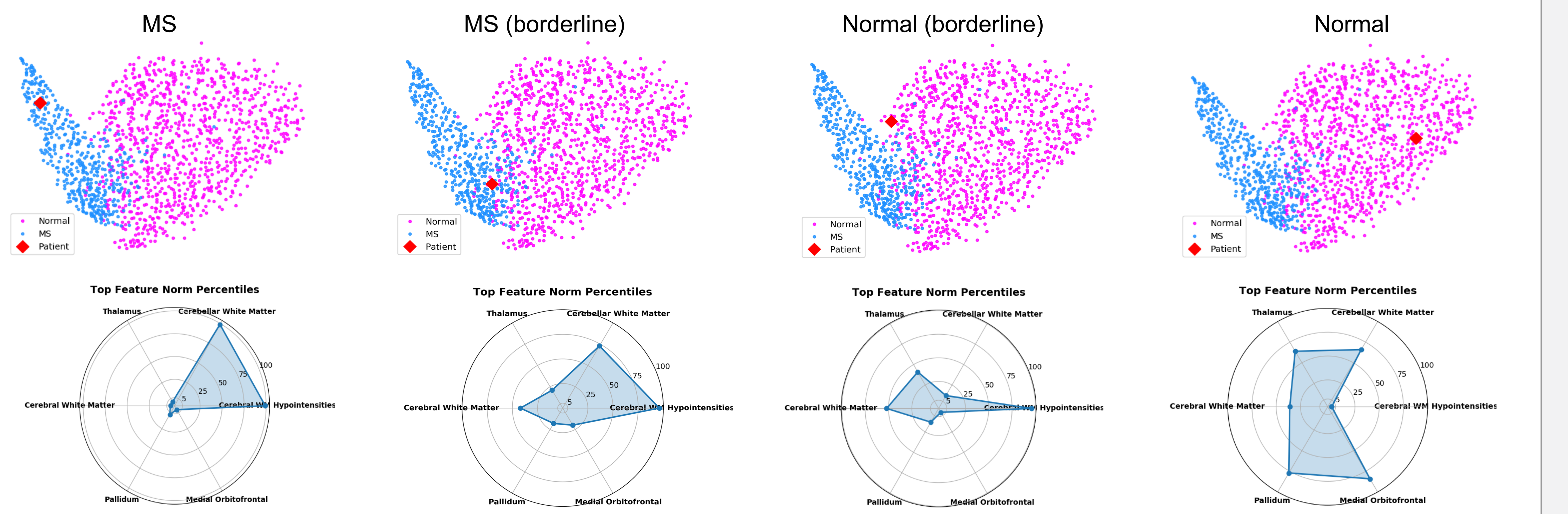


Figure 3. The test cases were classified based on the full age range model. The top row is the visualization of the classification algorithm based on MDS and how individual data point decided by the algorithm. On the bottom is the spie chart that visualize the six top features' percentile values.

DISCUSSION

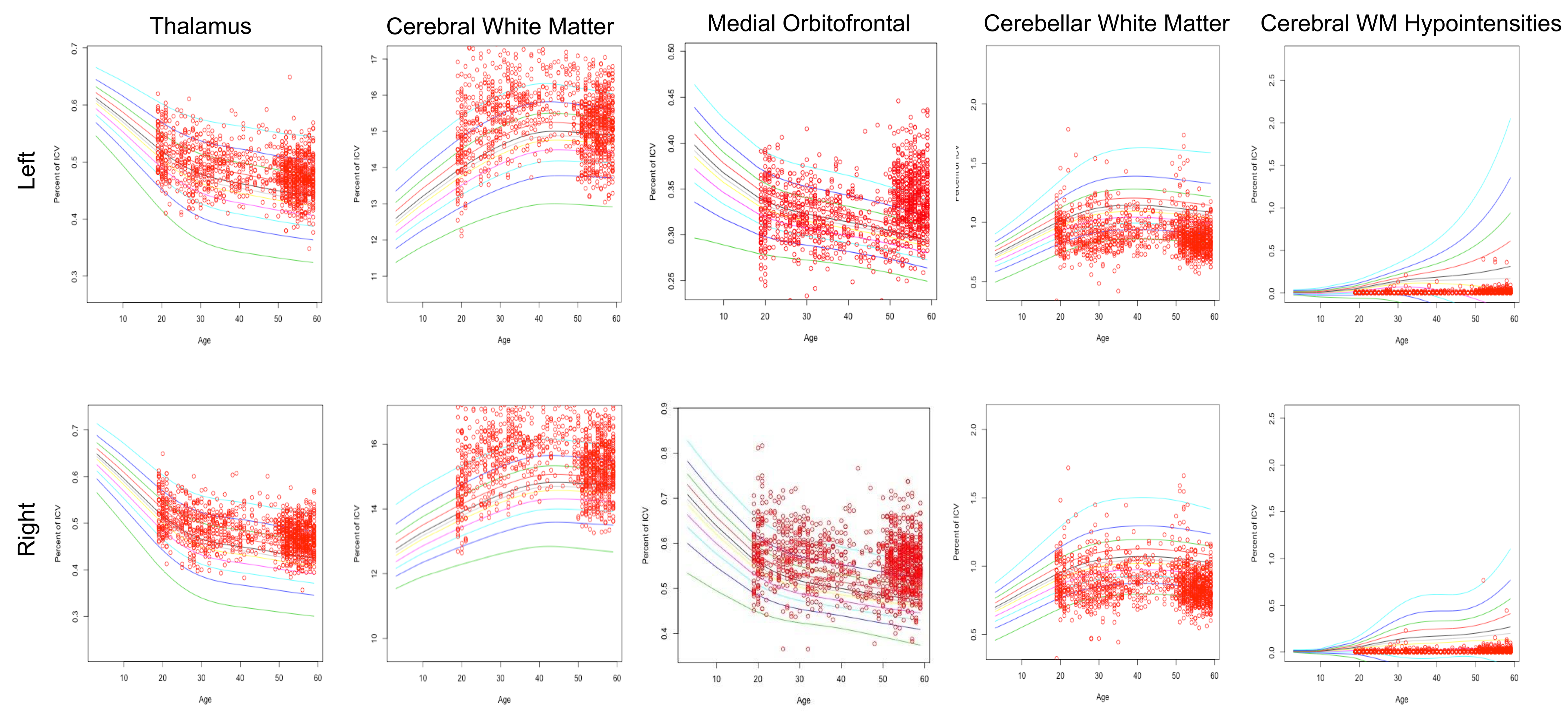


Figure 4. Normal data are plotted on the top of the MS curves created from MS subjects for various brain structures

- Brain structure volumes selected as features by the machine learning model show disparities in comparison to the normal data, as shown in Figure 4 above, where the scatter plots of the normal data deviate from the MS percentiles created from the MS data.
- Models constructed with data from varying age ranges selected different key features and weighted them differently, thus enabled different precision and sensitivity performance for MS classification. This also suggests the evolving effects of disease upon the brain during progression.

CONCLUSION

- Machine learning models showed high classification performance between normal and MS.
- These models might be useful in helping clinical decision making and with the initial differentiation and diagnostic MS workup.

DISCLOSURE

Dr. A.Mazhari has speaking engagements for Sanofi, Genentech, Serono; and he sits in Advisory Boards for Novartis, Biogen, and Rom3 rehab.

REFERENCES

[1] Lcierny, D, et. al. (2017). The age at onset in Multiple Sclerosis is associated with patient's prognosis. *Bratisl Lek Listy*, 118(6): 374:377
[2] Sastre-Garriga J. and Pareto D. et. al. (2017). Brain Atrophy In Multiple Sclerosis, Clinical Relevance and Technical Aspects, *Neuroimag Clin N Am,* 27: 289-300
[3] Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., et al. (2011). Diagnosis Criteria for Multiple Sclerosis: 2010 Revision to the McDonald Criteria. *Ann. Neurol.* 69, 292-302. doi: 10.1002/ana.22366
[4] Brass SD, Chen NK, Mulkern RV, et al. (2006). Magnetic resonance imaing of iron deposition in neurological disorders. Top Magn Reson Imaging, 17:31-40
[5] Powell T, Sussman JG, Davies-Jones GA. (1992). MR imaging in acute multiple sclerosis: ringlike appearance in plaques suggesting the presence of paramagnetic free radicals. AJNR Am J Neuroradiol, 13:1544-1546
[6] Janardhan V, et al. (2007). Multiple Sclerosis: hyperintaense lesions in the brain on nonenhanced T1-weighted MR images evidenced as areas of T1 shortening. Radiology, 244:823-831