

Interest of structured reporting and combined automated co-registration and lesion color-coding maps for longitudinal magnetic resonance imaging analysis in patients with multiple sclerosis : the MS-LOBI-SR study protocol.

Sébastien KERDRAON ¹, Ichem MOHTARIF ¹, Kevin ROSE ¹, Douraïed BEN SALEM ^{1,2}, Brieg DISSAUX ^{1,3} and Julien OGNARD ^{1,2}

1. Radiology department, University Hospital of Brest, Western Brittany, France

2. University of Western Brittany, UBO ; Laboratory of Medical Information Processing – LaTIM INSERM UMR 1101, France

2. University of Western Brittany, UBO ; Anatomy laboratory

Abstract

MS is characterized by chronic and immune-mediated inflammatory lesions of the brain white matter, disseminated in time and space. As the technology improved over the next three decades, MRI quickly grew to become the single most important paraclinical diagnostic and monitoring tool available. A lesion will be defined as having a high T2/FLAIR signal ≥ 3 millimeters in diameter. The purpose of this study, which includes 95 patients, was to investigate the benefit of a computer assisted detection reading using combined automated co-registration and lesion color-coding method approach, over a conventional reading approach in follow-up examinations of patients with MS, especially regarding diagnostic accuracy, inter/intra-reader agreement and required reading time. The impact of the use of a structured report on these variables will also be evaluated. The main judgment criteria will be : diagnostic accuracy (effectiveness) to assess the progression-activity of the disease based on the appearance of a T2w-FLAIR hyperintensity or a T1w enhancement on the follow-up examination, compared between consensus corrected CADR and CR readings. The reading of the examinations will be made by radiologists and radiology technician.

Introduction

MS is characterized by chronic and immune-mediated inflammatory lesions of the brain white matter (WM), disseminated in time and space (1). Besides cognitive impact, MS leads to the highest number of permanent disability in young adults (2). Europe rank among the regions with its highest incidence and prevalence (3).

The diagnosis is based on a bundle of clinical, paraclinical and evolutionary arguments. As the technology improved over the next three decades, MRI quickly grew to become the single most important paraclinical diagnostic and monitoring tool available. The ability to demonstrate lesion dissemination in space and time has led to the common use of MRI as a paraclinical measure to support a diagnosis of MS, including in patients with clinically isolated demyelinating syndromes(4).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

In addition, MRI is used to monitor the progress of disease in patients with clinically definite MS, including assessment of lesions and atrophy (5).

Current imaging markers with relevance for MS are associated with disease activity and progression, and include, among other features, number or volume of hyperintense brain lesions visible on T2-weighted MRI images, contrast-enhancing T1 lesions, increased annual brain volume loss and T1-hypointense “black holes” (6). All of these points position the MRI as a routine examination, and exploration of longitudinal changes in patients with MS is therefore very frequent.

Mainly the increase in the spatial resolution of acquisitions combined with the number/load of lesion may impair diagnostic accuracy due to a laborious interpretation. This visual MRI analysis is known as a tedious and time-consuming task that may suffer from a lack of sensitivity and intra- and inter-observer reproducibility. The heterogeneity of used machines (manufacturer, magnetic field) for the longitudinal follow up, also add to the variability of interpretation and can even have an impact regarding the experience of the reader. Hagens et al. (7) calculated an inter-rater agreement on involvement per anatomical region and dissemination in space and time that was moderate to good for 1.5 and 3T magnetic fields, while 3T slightly improved agreement between experienced raters, but slightly decreased agreement between less-experienced raters.

The processing of medical images makes it possible to assist in the interpretation and analysis of the images and more particularly in the longitudinal follow-up of a patient (8). This helps to smooth out differences in interpretation between readers, promotes the same interpretation at different levels of experience, and allows for an increase in reading speed. Wang et al. (9) have showed that a semi-automated assistive software that aids in the identification of new lesions has a beneficial effect for both neuroradiologists and non-neuroradiologists, though the effect is more profound in the non-neuroradiologist group.

Zopfs et al. (10) used automated co-registration and lesion color-coding of MS-associated FLAIR-lesions in followup MRI of MS patients and found that it increased diagnostic accuracy and reduced the reading time significantly. They also summed up different approaches and tools available to facilitate detection of disease progression, such as subtraction or automated co-registration techniques (Galletto AJNR 2018), voxel-based registration of pairs of images (12), subtraction of previously coregistered and intensity-normalized pairs of images (13).

With the recent advent of artificial intelligence, several machine learning techniques have been proposed to improve automated delineation of pathologies (14), however, many of these techniques may not be suitable in a clinical environment, as they either require special informatic skills to be implemented or difficult to include in a clinical workflow (15).

In recent years, structured reporting (SR) has been used in different fields of radiology to overcome report variability and increase standardization (16). Moreover, evidence is increasing that the implementation of SR may increase the standardization of reports, improve communication between the radiologist and the referring physician, potentially alter patient management, and simplify clinical

decision making (17). SR may be useful in analyzing patient outcomes, which is relevant in the setting of clinical trials (18). In the context of a pathology leading to potentially high inter-observer variability, it seemed necessary to standardize the reading, using a structured report.

Therefore, the purpose of this study was to investigate the benefit of a computer assisted detection reading (CADR) using combined automated co-registration and lesion color-coding method approach, over a conventional reading approach (CR) in follow-up examinations of patients with MS, especially regarding diagnostic accuracy, inter/intra-reader agreement and required reading time. The impact of the use of a structured report on these variables will also be evaluated.

Material and methods

Ethics

The study will be conducted in accordance with the ethical regulations of the Declaration of Helsinki including its later amendments. Considering the retrospective design, all cerebral MRI exams were already performed for clinical indications, no MRI will be conducted solely for the purpose of this study. An request of reviewing of this study plan was deposited to the CERIM (Comité d'éthique de recherche en imagerie médicale). The need of informed patient consent will be adapted to this reviewing board decision.

Subjects

Identification of potential study subjects will be conducted based on a combined exploration of the picture archiving and communication system (PACS, Centricity, GE), the radiological information system (RIS, Xplore), and the hospital information system (MEVA) screened, for a period of time adapted to the number needed of subjects, for patients 18 years of age or older who were diagnosed with MS, that underwent two consecutive MR imaging examinations of the brain for follow-up with complete standardized protocol, already included in the OFSEP (Office Français de la Sclérose en Plaque) registry,. Exclusion criterion will be retained: examination of lesser quality not allowing a correct analysis of the images. The examinations will therefore only be the ones performed in our institution (mutli-machine, single center).

Imaging data

Detailed scan parameters are given for the two MRI systems of the unit : Philips Elition X 3T and GE Optima MR450W 1.5T. Protocols of MRI examinations are implemented in accordance to national guidelines (OFSEP) and described in the tables below :

	DWI	T1w TFE	T2w FLAIR	T1w TSE FS
1.5T				
Acquisition plane	Axial	Sagittal	Sagittal	Axial
Encoding direction	AP	AP	AP	RL
FOV	240x265	240x240	250x250	256x256
Pixel size	1.6x2.3x3	1x1x1	1x1x1	1x1x1
Acceleration	Sense 2	Cs Sense 3	Cs Sense 7	Cs sense 6
TE	66	3.8	288	28
TR	3771	8.3	8000	600
TI	NA	NA	2300	NA
NSA	1	1	1	1
Fat Suppresion	SPIR	No	SPIR	SPIR
Time	1'31	3'32	5'20	3'58
3T				
Acquisition plane	Axial	Axial	Sagittal	Sagittal
Encoding direction	AP	AP	AP	RL
FOV	240x240	256x205	256x216	256x230
Pixel size	1,9x1,7x4	1x1x1	1,2x1,2x1,6	1x1x1
Acceleration	ASSET 2	ASSET 1,5	ARC 2/ Hypersense 1,1	ARC 2 Hypersense 1,15
TE	69,7	2,9	136	15
TR	8500	7,2	8000	500
TI			2112	
NSA	1	1	1	1
Fat Suppresion		No	FatSat	FatSat
Time	1'51	2'26	4'12	3'48

Table 1. MRI acquisition parameters

Reading process

A lesion will be defined as having a high T2/FLAIR signal ≥ 3 millimeters in diameter. The presence of a new T2w-FLAIR hyperintensity lesion will be defined as nonartifactual, new, bright areas clearly visible against the background.

All scans will be centrally collected and checked for completeness. The scans will be rated independently by four raters during two central reading sessions: one experienced rater (J.O., neuroradiologist for 5 years), two MS researchers considered as less-experienced raters (K.R., I.M., radiology residents for 3 years), and a non-radiologist (S.K., radiology technician for 10 years).

The disagreements will be resolved in consensus by two experienced raters non-implicated in the central reading, in the same reading conditions of readings (D.B.S, neuroradiologist for 20 years, B.D., neuroradiologist for 3 years).

The order and type of readings will be randomized in and between sessions, and won't be the same for all the raters, that will be fully blinded of clinical data.

There will be four types of readings, according to the design of the study : CADR with SR, CR with SR, CADR with NSR and CR with NSR.

Readings will be performed at the same workstation under standardized reading conditions with a time interval of 12 weeks between the conventional reading and the CADR reading, to avoid a recall bias.

For radiologists, the process will consist in performing the four types of readings. The radiology technician not carrying out in his current practice the radiological reports, he will perform only the reading using SR (CADR with SR, CR with SR).

The average time required for loading the images within the PACS will be recorded in the conventional reading session. The average time required for loading the applications and processing the images will be recorded. Furthermore, the time from being presented with the images to making the diagnosis will be recorded for each follow-up examination pair. All time measurements will be performed by an assistant not involved in the readouts.

Computer assisted detection reading CADR

To perform CADR, the readers will use a CE-certified and FDA-approved software that facilitates follow-up examinations of the same patient (MR Longitudinal Brain Imaging [LoBI]; Philips Healthcare). The software is integrated into the vendor's image viewer (Intellispace Portal, Version 11; Philips Healthcare), however is not constrained to MR images generated by Philips Healthcare scanners and generally applicable. The software automatically performs a rigid co-registration. Then, the application performs an intensity normalization and subsequent subtraction of the selected sequences. After co-registration, normalization, and subtraction, both sequences are linked at the same

anatomic level, allowing manual correction if necessary. Additionally, the software creates an overlay map, which highlights a focal increase in signal intensity in red and a focal decrease in signal intensity in blue, respectively. The color intensity of the color-coding can be adjusted seamlessly. Both linked sequences and the overlay map are displayed side by side and can be viewed simultaneously on the same anatomic level. (19)

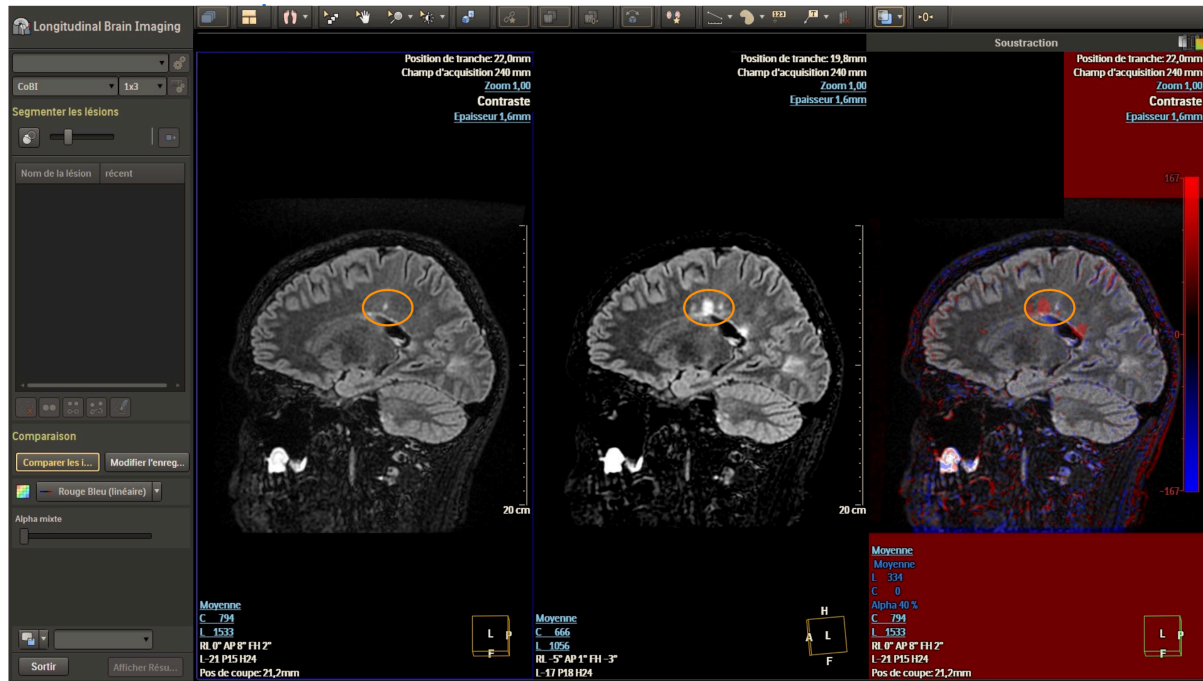


Fig 1. LOBI Interface. The yellow circle shows the apparition of a new hypersignal.

Conventional reading CR

Conventional readings will be performed as the reader habits, using CUV software (Centricity Universal Viewer, GE Healthcare)

Structured report SR

The SR will be generated using RADR software, a web application (<https://radr.fr/>, accessed on the 2022-06-05) that allows via a simple interface to creates reports through point and click. For the MS module, the anonymization, clinical information (such as type of MS, treatment, recent attack or not, EDSS score) as well as the sequences carried out are informed. The interface permits to select via simple clicks pre parcellated areas of the brain (lobes, corpus callosum, brainstem), and category of WM or cortex involved. The user can also inform the apparition of the lesion. A structured report is generated. A standardized table can be exported as a .Json file. For all scans, the number of inflammatory lesions larger than 3 mm in size will be scored and categorized according to the anatomical region. The number of enhancing lesions per region will be also reported if contrast media was used during the examination. Activity-progression of the disease will then be determined by each reader.

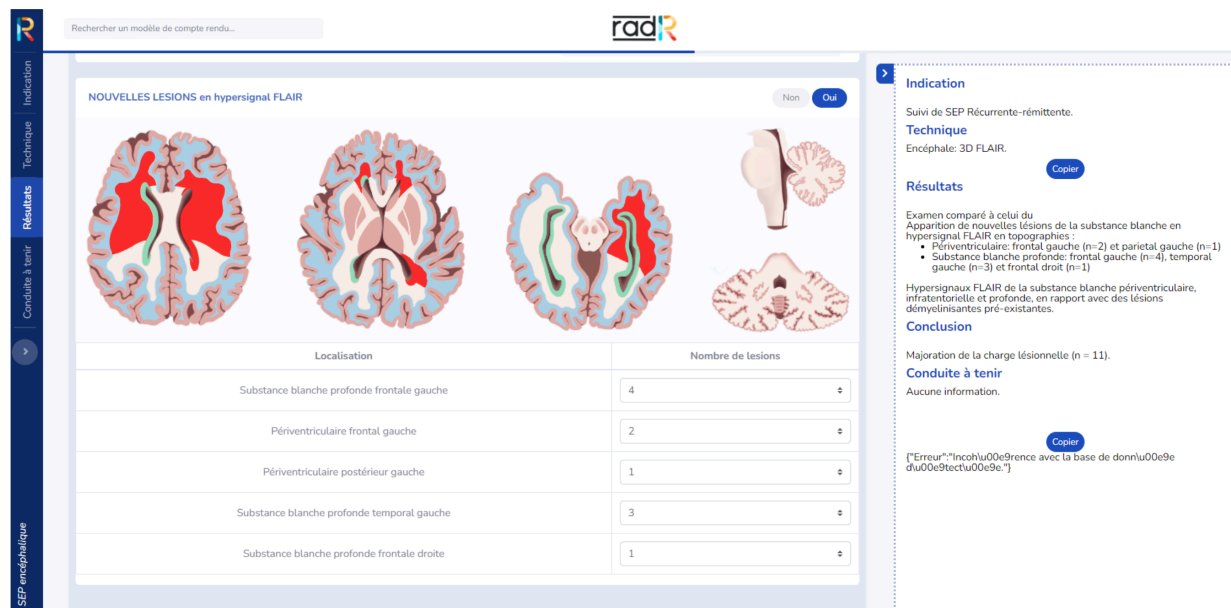


Fig2 RadR Interface

Non-structured report NSR

NSR will be performed as the radiologist own habits, using medical voice dication (Dragon)

Variables

For the SR readings, the number of T2w-FLAIR hyperintensities, number of T1w enhancing lesions, localization of the lesions will be recorded through the exported .json file from RADR. Will be adjudicated by the two experienced raters that were not part of central readings : the presence of a new T2w hyperintensity, the presence of a T1w enhancement, the progression-activity of the disease status.

For the NSR, given the unconstrained design of reporting, chosen to approach a clinical time of reading/reporting, the counts of the lesions won't be assessed. Only qualitative analysis will be made on the reports by the two experienced raters that were not part of central readings, extracting : the presence of a new T2w hyperintensity, the presence of a T1w enhancement, the progression-activity of the disease status. Each NSR will be stored according to the number of participation of the patient, to a separate informatics file.

For the CADR readings, failure or success of coregistration-fusion will be reported.

The recorded times will be stored in a separate informatics file.

Clinical and demographical data will be extracted from patient informatics medical file and will be stored in a separate informatics file.

At the end of the readings all files will be made available to the two experienced neuroradiologist to adjudicate the data, and then provided to the statistician for the analysis.

Primary endpoint

The main objective of this study is to evaluate the diagnostic performance of CADR compared to CR in the follow-up of MS.

Secondary endpoints

The secondary objectives will be studied using a quantitative and qualitative analysis of the reading :

- to evaluate the diagnosis performance of readings process
- to evaluate the variability of readings
- to evaluate the variability of reportings
- to evaluate the impact of CADR on the variability of readings
- to compare the variability of readings according to the reader experience / radiologist/non-radiologist status
- to detail the variability of readings according to lesion localization using SR
- to evaluate the impact of CADR on the reading time
- to evaluate the impact of SR on the reporting time
- to evaluate the acceptability of CADR

Primary judgment criteria

The main judgment criteria will be : Diagnostic accuracy (effectiveness) to assess the progression-activity of the disease based on the appearance of a T2w-FLAIR hyperintensity or a T1w enhancement on the follow-up examination, compared between consensus corrected CADR and CR readings.

Secondary judgment criterion

The secondary criterion will be based on the count of T2w hyperintensities, count of T1w enhancement (quantitative), progression-activity status (qualitative):

-AUC-ROC, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value

-Intra and Inter-observer agreements between CADR-SR and CR-SR / CADR-NSR and CR-NSR

-Intra and Inter-observer agreements between CADR-SR and CADR-NSR / CR-SR and CR-NSR

-Intra and Inter-observer agreements between CADR-SR and CR-SR

-Intra and Inter-observer agreements between CADR-SR and CR-SR in the subgroups of experiences

-Intra and Inter-observer agreements for lesion localisations using SR

-Reading times

-Reporting times

- Number of unexpected interruptions of the software, presence of a sub-optimal analysis described by the software.

Statistics

Number of subject needed

According to Zopfs et al. (10.) who retrospectively assessed 70 follow-up MRI of 53 patients with MS using the same CADR software, diagnostic accuracy increased from 67 to 90% for a and from 70 to 87% for the other using CADR versus CR. Diagnostic accuracy (effectiveness), is the a proportion of correctly classified subjects (True Positive TP +True Negative TN) among all subjects (TP+TN+False Positive FP+False Negative). According to the literature prevalence and contingency table (Zopfs et al. (10.)) of changes between follow-up exams, the expected effect of a 20% change of diagnostic accuracy, the number of subject needed to depict this difference with an absolute precision of 0.10 is beyond 63 and 95.

Analysis

Statistics will be performed by an independent statistician, using STATA (MP 16, StataCorp). Quantitative variables will be presented as means and medians (interquartile range [IQR]), and categoric variables, as percentages. Mixed models will fit to compare the 4 reading methods (CADR-SR ; CADR-NSR ; CR-SR ; CR-NSR) : A Poisson regression will be used for the number of new lesions (CADR-SR and CR-SR); a logistical regression, for binary variables such as the presence of at least 1 new lesion/state of progression-activity of the disease (all the readings); and a γ regression, for the reading time. The κ coefficient (qualitative) and intraclass correlation coefficient (quantitative) will

be used to assess interobserver and intraobserver agreement and will be interpreted as suggested in previous publications (Landis and Koch): excellent agreement ($\kappa \geq 0.8$), good agreement ($\kappa \geq 0.6$), moderate agreement ($\kappa \geq 0.4$), poor agreement ($\kappa \leq 0.4$); and poor, fair, good, and excellent agreement categories will be qualified according to Cicchetti. The difference between dichotomous variables will be evaluated using McNemar's test and the difference between continuous variables using the Wilcoxon test. Statistical significance will be set to $p \leq 0.05$.

References

1. Karussis D. The diagnosis of multiple sclerosis and the various related demyelinating syndromes: a critical review. *J Autoimmun.* 2014 Feb-Mar;48-49:134-42. doi: 10.1016/j.jaut.2014.01.022. Epub 2014 Feb 10. PMID: 24524923.
2. Ramagopalan SV, Sadovnick AD. Epidemiology of multiple sclerosis. *Neurol Clin.* 2011 May;29(2):207-17. doi: 10.1016/j.ncl.2010.12.010. PMID: 21439437.
3. Simpson S Jr, Blizzard L, Otahal P, Van der Mei I, Taylor B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *J Neurol Neurosurg Psychiatry.* 2011 Oct;82(10):1132-41. doi: 10.1136/jnnp.2011.240432. Epub 2011 Apr 8. PMID: 21478203.
4. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, Fujihara K, Galetta SL, Hartung HP, Kappos L, Lublin FD, Marrie RA, Miller AE, Miller DH, Montalban X, Mowry EM, Sorensen PS, Tintoré M, Traboulsee AL, Trojano M, Uitdehaag BMJ, Vukusic S, Waubant E, Weinshenker BG, Reingold SC, Cohen JA. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2018 Feb;17(2):162-173. doi: 10.1016/S1474-4422(17)30470-2. Epub 2017 Dec 21. PMID: 29275977.
5. Wattjes MP, Rovira À, Miller D, Yousry TA, Sormani MP, de Stefano MP, Tintoré M, Auger C, Tur C, Filippi M, Rocca MA, Fazekas F, Kappos L, Polman C, Frederik Barkhof, Xavier Montalban; MAGNIMS study group. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis--establishing disease prognosis and monitoring patients. *Nat Rev Neurol.* 2015 Oct;11(10):597-606. doi: 10.1038/nrneurol.2015.157. Epub 2015 Sep 15. PMID: 26369511.
6. Gramsch C, Nensa F, Kastrup O, Maderwald S, Deuschl C, Ringelstein A, Schelhorn J, Forsting M, Schlamann M. Diagnostic value of 3D fluid attenuated inversion recovery sequence in multiple sclerosis. *Acta Radiol.* 2015 May;56(5):622-7. doi: 10.1177/0284185114534413. Epub 2014 May 27. PMID: 24867222.
7. Hagens MH, Burggraaff J, Kilsdonk ID, Ruggieri S, Collorone S, Cortese R, Cawley N, Sbardella E, Andelova M, Amann M, Lieb JM, Pantano P, Lissenberg-Witte BI, Killestein J, Oreja-Guevara C, Wuerfel J, Ciccarelli O, Gasperini C, Lukas C, Rovira A, Barkhof F, Wattjes MP; MAGNIMS Study Group. Impact of 3 Tesla MRI on interobserver agreement in clinically isolated

syndrome: A MAGNIMS multicentre study. *Mult Scler*. 2019 Mar;25(3):352-360. doi: 10.1177/1352458517751647. Epub 2018 Jan 12. PMID: 29327668; PMCID: PMC6393953.

8. García-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med Image Anal*. 2013 Jan;17(1):1-18. doi: 10.1016/j.media.2012.09.004. Epub 2012 Sep 29. PMID: 23084503.

9. Wang W, van Heerden J, Tacey MA, Gaillard F. Neuroradiologists Compared with Non-Neuroradiologists in the Detection of New Multiple Sclerosis Plaques. *AJNR Am J Neuroradiol*. 2017 Jul;38(7):1323-1327. doi: 10.3174/ajnr.A5185. Epub 2017 May 4. PMID: 28473341; PMCID: PMC7959901.

10. Zopfs D, Laukamp KR, Paquet S, Lennartz S, Pinto Dos Santos D, Kabbasch C, Bunck A, Schlamann M, Borggreffe J. Follow-up MRI in multiple sclerosis patients: automated co-registration and lesion color-coding improves diagnostic accuracy and reduces reading time. *Eur Radiol*. 2019 Dec;29(12):7047-7054. doi: 10.1007/s00330-019-06273-x. Epub 2019 Jun 14. PMID: 31201526.

11. Galetto?

12. Tan IL, van Schijndel RA, Fazekas F, Filippi M, Freitag P, Miller DH, Yousry TA, Pouwels PJ, Adèr HJ, Barkhof F. Image registration and subtraction to detect active T(2) lesions in MS: an interobserver study. *J Neurol*. 2002 Jun;249(6):767-73. doi: 10.1007/s00415-002-0712-6. PMID: 12111312.

13. Horsfield MA, Rocca MA, Pagani E, Storelli L, Preziosa P, Messina R, Camesasca F, Copetti M, Filippi M. Estimating Brain Lesion Volume Change in Multiple Sclerosis by Subtraction of Magnetic Resonance Images. *J Neuroimaging*. 2016 Jul;26(4):395-402. doi: 10.1111/jon.12344. Epub 2016 Mar 28. PMID: 27019077.

14. Van Heerden J, Rawlinson D, Zhang AM, Chakravorty R, Tacey MA, Desmond PM, Gaillard F. Improving Multiple Sclerosis Plaque Detection Using a Semiautomated Assistive Approach. *AJNR Am J Neuroradiol*. 2015 Aug;36(8):1465-71. doi: 10.3174/ajnr.A4375. Epub 2015 Jun 18. PMID: 26089318; PMCID: PMC7964704.

15. Valverde S, Salem M, Cabezas M, Pareto D, Vilanova JC, Ramió-Torrentà L, Rovira À, Salvi J, Oliver A, Lladó X. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. *Neuroimage Clin*. 2019;21:101638. doi: 10.1016/j.nicl.2018.101638. Epub 2018 Dec 10. PMID: 30555005; PMCID: PMC6413299.

16. Larson DB, Towbin AJ, Pryor RM, Donnelly LF. Improving consistency in radiology reporting through the use of department-wide standardized structured reporting. *Radiology*. 2013 Apr;267(1):240-50. doi: 10.1148/radiol.12121502. Epub 2013 Jan 17. PMID: 23329657.

17. Dickerson E, Davenport MS, Syed F, Stuve O, Cohen JA, Rinker JR, Goldman MD, Segal BM, Foerster BR; Michigan Radiology Quality Collaborative. Effect of Template Reporting of Brain MRIs for Multiple Sclerosis on Report Thoroughness and Neurologist-Rated Quality: Results of a Prospective Quality Improvement Project. *J Am Coll Radiol*. 2017 Mar;14(3):371-379.e1. doi: 10.1016/j.jacr.2016.09.037. Epub 2016 Dec 5. PMID: 27932248.
18. Sahni VA, Silveira PC, Sainani NI, Khorasani R. Impact of a Structured Report Template on the Quality of MRI Reports for Rectal Cancer Staging. *AJR Am J Roentgenol*. 2015 Sep;205(3):584-8. doi: 10.2214/AJR.14.14053. PMID: 26295645.
19. Zopfs D, Laukamp K, Reimer R, Grosse Hokamp N, Kabbasch C, Borggrefe J, Pennig L, Bunck AC, Schlamann M, Lennartz S. Automated Color-Coding of Lesion Changes in Contrast-Enhanced 3D T1-Weighted Sequences for MRI Follow-up of Brain Metastases. *AJNR Am J Neuroradiol*. 2022 Feb;43(2):188-194. doi: 10.3174/ajnr.A7380. Epub 2022 Jan 6. PMID: 34992128; PMCID: PMC8985679.