



OPEN ACCESS

EDITED BY

Aydin Eresen,
University of California, Irvine, United States

REVIEWED BY

Sönke Peters,
University Medical Center Schleswig-Holstein,
Germany
Maryam Zamanian,
Isfahan University of Medical Sciences, Iran

*CORRESPONDENCE

Myriam Edjlali
✉ myriam.edjlali@aphp.fr

RECEIVED 01 August 2025

ACCEPTED 18 September 2025

PUBLISHED 08 October 2025

CITATION

Mastilović M, Heinzlef O, Federau C,
Muñoz-Ramírez V, Blanchere M, Boban J,
Cotton F and Edjlali M (2025) Evaluation of
two AI techniques for the detection of new
T2/FLAIR lesions in the follow-up of multiple
sclerosis patients.

Front. Neurol. 16:1678073.

doi: 10.3389/fneur.2025.1678073

COPYRIGHT

© 2025 Mastilović, Heinzlef, Federau,
Muñoz-Ramírez, Blanchere, Boban, Cotton
and Edjlali. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Evaluation of two AI techniques for the detection of new T2/FLAIR lesions in the follow-up of multiple sclerosis patients

Milica Mastilović^{1,2}, Olivier Heinzlef^{3,4}, Christian Federau^{5,6,7},
Verónica Muñoz-Ramírez⁸, Marie Blanchere^{3,4},
Jasmina Boban^{2,9}, Francois Cotton^{10,11} and Myriam Edjlali^{1,12*}

¹Laboratoire d'imagerie Biomédicale Multimodale (BioMaps), Université Paris-Saclay CEA, CNRS, Inserm, Service Hospitalier Frédéric Joliot, Orsay, France, ²Faculty of Medicine, University of Novi Sad, Novi Sad, Serbia, ³Department of Neurology, Poissy-Saint-Germain-en-Laye Hospital, Poissy, France, ⁴CRC SEP IDF Ouest, Poissy, Garches, France, ⁵AI Medical AG, Zollikon, Switzerland, ⁶University of Zürich, Zürich, Switzerland, ⁷Kyoto Prefectural University of Medicine, Kyoto, Japan, ⁸Pixyl Research and Development Laboratory, Grenoble, France, ⁹Oncology Institute of Vojvodina, Sremska Kamenica, Serbia, ¹⁰Radiology Department Centre Hospitalier Lyon-Sud, Hospices Civils de Lyon, Oullins-Pierre-Bénite, France, ¹¹CREATIS, INSERM U1044, CNRS UMR 5220, UCBL1, Villeurbanne, France, ¹²Department of Radiology, APHP, Hôpitaux Raymond-Poincaré & Ambroise Paré, Paris, France

Background: Multiple sclerosis is an inflammatory demyelinating disease of the CNS. Annual MRI exams are crucial for disease monitoring. Interpreting high T2/FLAIR lesion loads can be laborious. AI aids in lesion detection, and choosing between different solutions can be challenging.

Aim: This study compares two distinct software, Pixyl.Neuro.MS® and Jazz®, to assess their performance in T2/FLAIR lesion detection between two-time points.

Methods: Retrospective analysis included follow-up MRIs from 35 MS patients. Pixyl.Neuro.MS® automatically segments and classifies lesions. Jazz® automates the reading process and image display. Two readers (15 and 4 years of experience) conducted radiological analysis, followed by AI-assisted readings. A number of new lesions (NL) and reading times were recorded, with ground truth (GT) established by consensus. AI-detected lesions were classified as true (TP) and false positives (FP). Statistical analysis used SPSS ($p < 0.05$).

Results: Pixyl.Neuro.MS® readings averaged 2 min 46 s \pm 1 min 4 s while using Jazz® 3 min 33 s \pm 2 min 24 s. Over 50% of the population had a high lesion load (>20 lesions). Both software significantly improved NL detection ($p < 0.01$ for both), revealing them in more patients than standard readings. Standard reports found 8 NL in 2 patients, while AI-assisted readings detected at least 17 TP in 7 patients and rejected 61 FP lesions. GT detected 21 lesions in 19 patients.

Conclusion: Both AI software have been found to enhance NL detection in MS patients, outperforming standard methods. These tools offer crucial advantages for accurate disease monitoring.

KEYWORDS

artificial intelligence, deep learning, magnetic resonance imaging, multiple sclerosis, lesion evaluation

1 Introduction

Multiple sclerosis (MS) is a widespread inflammatory demyelinating disease characterized by the presence of demyelinating lesions predominantly located in the central nervous system, typically distributed across periventricular, cortical/juxtacortical, infratentorial, and spinal cord regions (1, 2). The global prevalence of MS has witnessed a notable increase from 2.3 million individuals in 2013 to 2.8 million in 2020, highlighting the significant burden this condition imposes on healthcare systems worldwide. Despite extensive research efforts, a definitive cure for MS remains elusive, with current therapeutic interventions primarily focused on mitigating the risk of relapses and disability progression (3–7).

MAGNIMS group recommends at least yearly follow-up MRI exams of MS patients, or more frequently if there is a modification in the treatment strategy with a 3–6 months MRI after the initiation of treatment (8). OFSEP working group (9) highlights the importance of gadolinium contrast agent deposition in the brain and suggests that contrast is unnecessary in every follow-up exam. Therefore, accurately detecting new or enlarging T2/FLAIR lesions on follow-up magnetic resonance imaging (MRI) scans serves as a critical marker for tracking the evolution of MS disease progression (8, 10, 11). However, evaluating MS lesions poses considerable challenges due to the inherent variability in lesion distribution and load among patients. Visual lesion detection, the traditional approach, is not only time-consuming but also lacks consistent reproducibility (12).

In response to this challenge, the field has witnessed a surge in the development of artificial intelligence (AI) solutions tailored to MRI imaging in MS patients. These AI applications encompass a spectrum of software aimed at lesion detection and segmentation in MRI images, leveraging both conventional machine learning and deep learning techniques (13–17). Some AI software tools perform a longitudinal analysis of images to detect any new lesions directly between two visits. In contrast, others prioritize automating the reading process using AI to provide assistance to radiologists. However, navigating the multitude of available radiological software options presents a significant challenge for clinicians and researchers alike (18, 19).

Radiological workflow in everyday clinical practice consists of several steps that cannot be skipped, including protocol adaptation, display of images, co-registration of images, image interpretation in addition to other relevant clinical data, and finally, the summarization of all relevant findings in a final report. Software like Jazz® and Pixyl Neuro.MS® are examples of AI tools that radiologists can use to assist with various steps in their workflow. These two software have distinct functionalities: Jazz® software (20) does a co-registration of images, displaying them to the radiologist in a more simplified manner, enabling single-click switching between previous and current images; conversely, Pixyl Neuro.MS® software (21) performs lesion segmentation, providing radiologists with an AI-generated lesion mask with color-coded lesions to indicate their evolution.

Hence, this study aims to address this question by comparing the performance of two distinct software platforms, each with its unique functionalities, in aiding T2/FLAIR new lesion detection during the follow-up of MS patients.

Our aim was to determine the efficacy of these software solutions in facilitating the detection and characterization of MS lesions on MRI

scans, ultimately contributing to improved patient care and management strategies.

2 Materials and methods

This retrospective study included 35 patients with MS who were referred to our institution for second opinion and follow-up for severe disease progression or high burden of MS lesions. All patients provided their written informed consent. Ethical approval was obtained from the Institutional Review Board of CPP Ile-de-France VI (ID RCB: 2019-A03066-51).

2.1 MRI acquisition

All patients underwent both a prior and a follow-up MRI scan using the same Siemens 3 T Skyra scanner (Erlangen, Germany). The standard OFSEP protocol for imaging of MS patients was used, including 3D FLAIR, thin slice 3D T1, axial DWI and ADC, and if needed 3D T1 post contrast, SWI, T2, DTI. However, for this retrospective study only FLAIR sequences were used for lesion analysis with the following parameters: TR = 7,000 ms, TE = 2.9 ms, TI = 900 ms, slice thickness = 1 mm, dFOV = 36.1 × 25.6 cm, resolution = 256 × 240. All images were pseudonymized before analysis.

2.2 Image analysis workflow

To evaluate MS lesion detection, we implemented a structured three-step analysis:

2.2.1 Standard clinical reading

The standard clinical reports used for comparison were generated in routine practice by different board-certified radiologists with varying years of experience. This heterogeneity reflects real-world clinical conditions and ensures that the study results are representative of daily reporting practice.

2.2.2 Reassessment using Jazz® software

MRI scans were reanalyzed using Jazz® (version 1.0.0, AI Medical, Switzerland), an AI-assisted software designed to facilitate lesion detection through automated co-registration of longitudinal FLAIR sequences. The software integrates a dedicated visualization environment tailored for multiple sclerosis follow-up.

In practical terms, Jazz® enables radiologists to directly compare prior and follow-up scans at the same anatomical location, allowing instantaneous toggling between time points with a single click (mouse or keyboard shortcut). This synchronized display helps the reader identify subtle signal changes. In addition, the software provides a “lesion locking” option: once a lesion is selected, the tool automatically tracks its anatomical location across different time points, ensuring consistent assessment of potential lesion evolution (Figure 1).

For this study, two independent readers reassessed the cases with both software solutions: one senior neuroradiologist with 15 years of

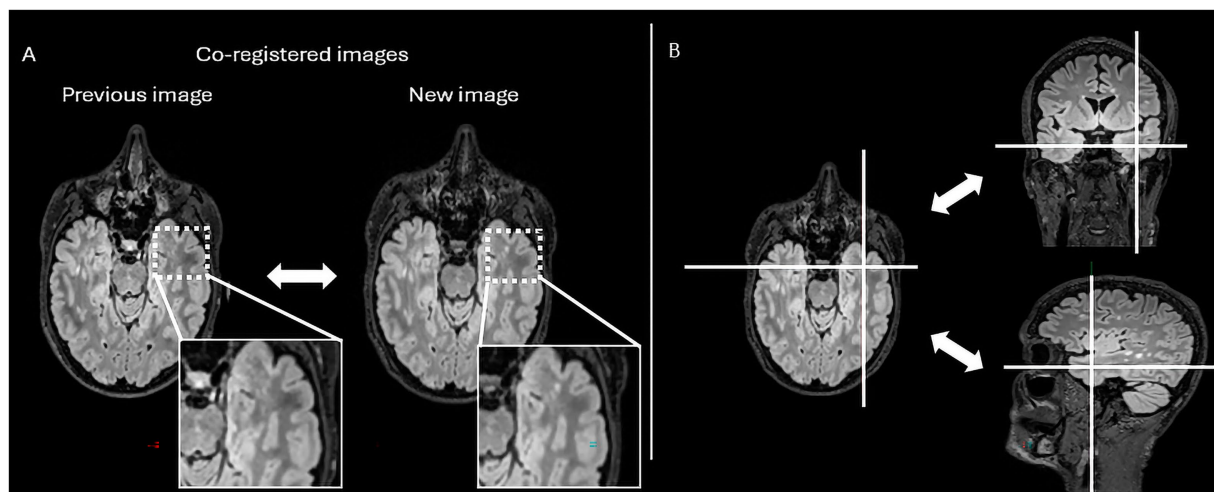


FIGURE 1

The figure represents the reading interface in Jazz® software. It allows a comparison of previous and new MRI FLAIR images of a patient with MS. The reader can easily switch from previous to new image, and vice versa, just by mouse clicking or using a keyboard shortcut (A), while there is as well lesion locking option (B) when the software automatically detects lesion's anatomical location.

experience and one radiology trainee with 4 years of experience. The workflow using Jazz® software included (i) review of the paired longitudinal FLAIR images within the Jazz® interface, (ii) use of the integrated annotation tool to label new lesions, and (iii) recording of the time required for each case. At the end of the evaluation, Jazz® automatically generated a structured quantitative report summarizing the number of new lesions identified, linked to the annotated regions.

The software is designed to fit into clinical routine by allowing radiologists to quickly switch between native images and AI-assisted visualizations without altering the standard reporting process. While the primary focus in this study was on new lesion detection, Jazz® also provides quantitative lesion counts that can be exported as part of the final report.

2.2.3 Reassessment using Pixyl.Neuro.MS® software

Six months later, the same two readers reanalyzed the images using Pixyl.Neuro.MS® (version 1.8.7, Pixyl SAS, France) a CE-marked and FDA-cleared cloud-based medical device for quantitative brain MRI analysis. Pixyl.Neuro® is organized into several dedicated modules tailored to different clinical contexts; in this study, the MS module (Pixyl.Neuro.MS®) was used for longitudinal monitoring of multiple sclerosis patients.

The workflow is as follows: DICOM images exported from the MRI scanner are pseudonymized and uploaded to the Pixyl platform. In routine clinical practice, images can be retrieved automatically from the PACS and results are returned directly to the hospital PACS as DICOM overlays and structured reports. For this research setting, however, all analyses were launched manually, and a dedicated web interface and a secure data sharing system was used to facilitate the review of outputs.

The MS module integrates a complete processing pipeline: (i) automated image preprocessing and quality control, (ii) segmentation

using a supervised deep learning algorithm based on convolutional neural networks (CNNs), and (iii) automated generation of results and reports in DICOM format.

Pixyl.Neuro.MS® processes 3D T2-FLAIR sequences independent of the MRI manufacturer. The software automatically segments white matter hyperintensity lesions and quantifies lesion load by anatomical regions. When two time points are available, its longitudinal module generates a color-coded activity map (red for new lesions, blue for stable lesions, yellow for enlarging lesions) together with a quantitative report summarizing lesion number, volume, and classification (stable, new, or growing; Figure 2).

In this study, we specifically used the longitudinal setting, and the radiologists concentrated solely on the detection of new lesions.

To prevent recognition bias, the cases were presented in a different order from the previous evaluation.

While two authors of the manuscript are affiliated with the software companies, several measures were taken to ensure an independent and unbiased validation. First, the retrospective analysis and validation were performed by authors with no ties to either company. Second, the MRI cases included in the study were not previously used for software development or training, ensuring that both AI solutions were tested on independent real-world data. Third, the AI-assisted outputs were compared against a ground truth established by experienced neuroradiologists who had no involvement in software development. Together, these safeguards minimize the risk of industry bias and strengthen the objectivity of our findings.

2.3 Ground truth definition

The final ground truth for lesion detection was established through a side-by-side review of prior and follow-up MRIs by the same two readers, who reached a consensus on the exact number of new T2/FLAIR lesions. The ground truth evaluation defined a new

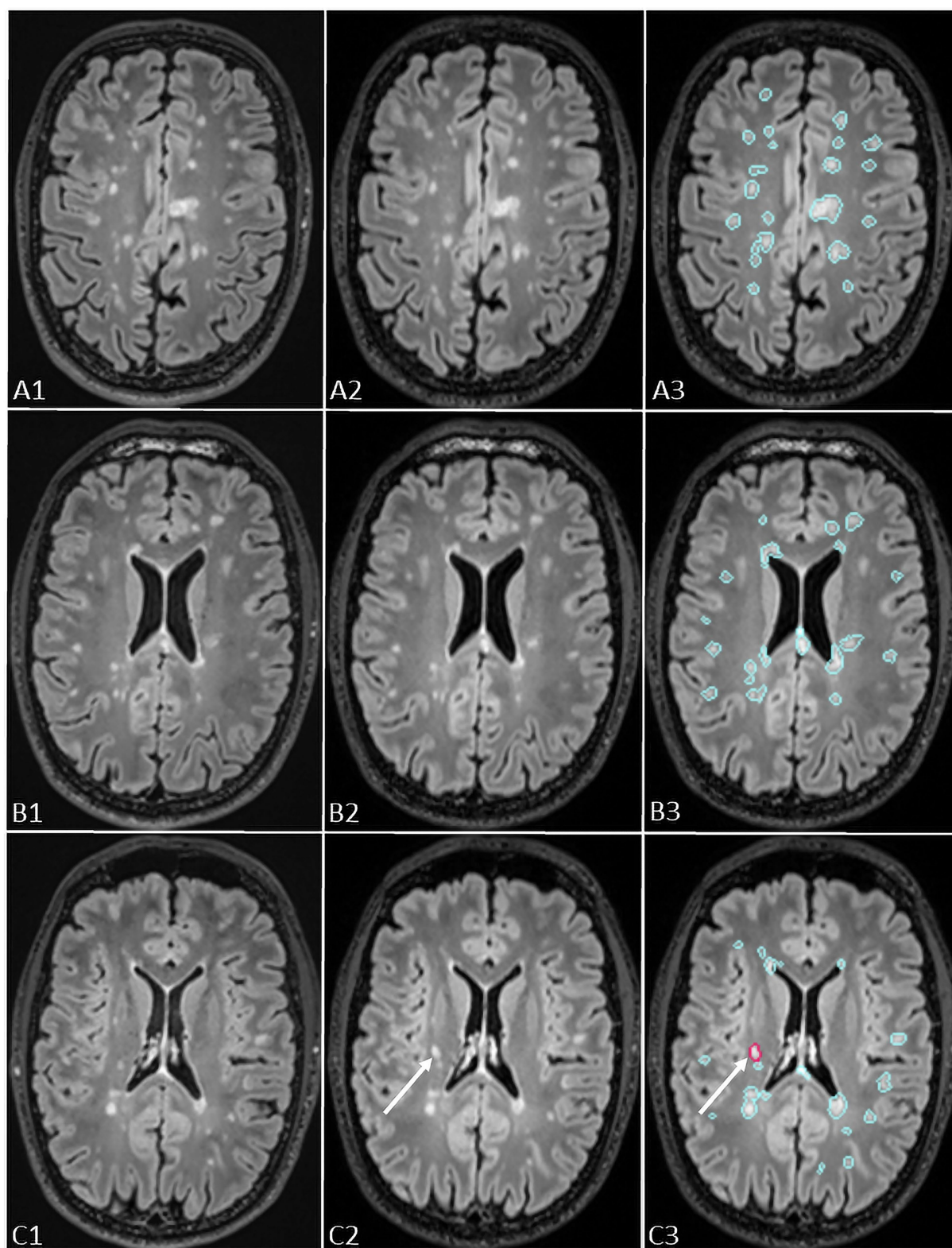


FIGURE 2

The figure shows representation of longitudinal evaluation of previous and new FLAIR images of a MS patient with high lesion load using Pixyl.Neuro. MS[®] segmentation mask. The lesions are color-coded in the segmentation mask: blue – stable lesion, red – new lesion. The white arrows in C3 indicate a new lesion detected using Pixyl.Neuro. MS[®], confirmed by comparison of new (C2) and previous (C1) exams.

lesion as an area of abnormality measuring at least 3 mm in the axial plane (1), located in typical MS regions (periventricular, juxtacortical, infratentorial) or within non-specific white matter. Only 3D FLAIR

sequences were used for lesion detection. AI software outputs (from Jazz[®] and Pixyl.Neuro. MS[®]) were incorporated into the consensus process to assess the validity of detected lesions.

2.4 Statistical analysis

Statistical analysis was performed using the SPSS Statistics 27 package (IBM, SPSS, NY). Descriptive statistics included mean, median, and standard deviation for quantitative variables, while qualitative values were expressed as ratios. The Wilcoxon signed-rank test was used to determine statistically significant differences between: (1) the number of patients with new lesions detected by ground truth versus standard radiological reports; (2) the number of new lesions detected using Jazz® and Pixyl.Neuro.MS® compared to standard reports. Significance was set at $\alpha < 0.05$. Correlation between readers' performances was assessed using Pearson's correlation coefficient, with the following interpretation: $r \leq 0.19$ —very low correlation; 0.2–0.39—low correlation; 0.4–0.59—moderate correlation; 0.6–0.79—high correlation; 0.8–1—very high correlation. This structured approach ensures a clear workflow, minimizing bias while assessing the added value of AI-assisted lesion detection in MS follow-up imaging.

3 Results

3.1 Patient population

The patients' demographic and MRI characteristics are given in Table 1. The analysis included 35 patients (mean age 48.8 ± 12.75) with multiple sclerosis with follow-up exams. Our cohort consisted of 75% women and 25% men. Lesions load varied from less than 10 lesions to more than 20 lesions, with the majority of patients exhibiting a severe lesion load of more than 20 lesions at the cerebral level

TABLE 1 Study population characteristics.

| Demographic and MRI characteristics | |
|--|---------------|
| Number of patients | 35 |
| Female | 26 |
| Male | 9 |
| Age, mean (SD), years | 48.8 ± 12.75 |
| Female | 50.27 ± 13.28 |
| Male | 44.55 ± 12.58 |
| EDSS, median (interquartile range) | |
| Baseline | 4 (2.5–6) |
| Follow-up | 4 (2.5–6) |
| EDSS change | 0 (0–0.5) |
| Time between scans, days (interquartile range) | 499 (234–690) |
| Lesion load, number (%) | |
| <10 lesion | 6 (17.14%) |
| 10–20 lesions | 12 (34.28%) |
| >20 lesions | 17 (48.57%) |
| Clinical form, number (%) | |
| Relapsing–remitting | 21 (60%) |
| Primary progressive | 2 (5.71%) |
| Secondary progressive | 12 (34.29%) |

(54.29%). In comparison, 28.57% had between 10 and 20 lesions, and 8.86% had less than 10 lesions.

3.2 Inter-reader reliability

A correlation analysis revealed a high and statistically positive correlation between readers when using both Pixyl.Neuro.MS® ($r = 0.73$, $p < 0.001$) and Jazz® software ($r = 0.75$, $p < 0.001$).

3.3 Reading time

Reading added time using Pixyl.Neuro.MS® software took on average $2\text{ m }46\text{ s} \pm 1\text{ m }4\text{ s}$ for both readers, while the reading time was on average $2\text{ min }12\text{ s} \pm 51\text{ s}$ for reader 1, and $3\text{ min }26\text{ s} \pm 1\text{ min }5\text{ s}$ for reader 2.

Reading added time using Jazz® software took, on average, $3\text{ m }33\text{ s} \pm 2\text{ m }24\text{ s}$ for both readers, while the reading time was, on average, $2\text{ m }25\text{ s} \pm 1\text{ m }6\text{ s}$ for reader 1, and $4\text{ m }41\text{ s} \pm 2\text{ m }48\text{ s}$ for reader 2.

3.4 New T2/FLAIR lesions

The ground truth reported 21 new lesions from 9 patients (Figure 3). With the use of Pixyl.Neuro.MS® software, reader 1 reported 19 true positive lesions from 9 patients, while reader 2 reported 20 true positive lesions from 8 patients. On the other hand, with the use of Jazz® software, reader 1 reported 19 true positive lesions in 8 patients, while reader 2 reported 17 true positive lesions in 7 patients. The standard report reported 8 new lesions in 2 patients.

A statistically significant difference exists between the number of patients with new lesions detected by radiological reading using Pixyl.Neuro.MS® and Jazz® software compared to the number of patients with new lesions detected by the standard radiological report ($z = -2.646$, $p < 0.01$ for both software).

Finally, compared to the ground truth, 61 false positive new lesions were detected by Pixyl.Neuro.MS® software alone, and radiologists did not retain these lesions (Figure 4). False positive lesions were associated with artifacts in 40% (24/61) of cases, while the remaining 60% (38/61) resulted from co-registration errors (28%) and the presence of slowly enlarging lesions (SELs; 26%).

The addition of AI support markedly improved lesion detection compared to standard clinical reports. The standard reports, produced in routine practice, showed very high specificity and PPV (100%) but low sensitivity (22.2%), reflecting the tendency to miss subtle new lesions under clinical time constraints, in a population with a large number of MS lesions per patient. When using Jazz®, both readers achieved substantially higher sensitivities (77.8–88.9%) while maintaining perfect specificity and PPV, highlighting its ability to increase lesion detection without introducing false positives. Pixyl® further increased sensitivity, reaching 100% for both readers, although Reader 2 experienced a slight decrease in specificity (96.2%) and PPV (90%), indicating a small number of false positives. When used in a fully automated mode, Pixyl® detected all true lesions (100% sensitivity and NPV) but generated a high number of false positives, leading to low specificity (26.9%)

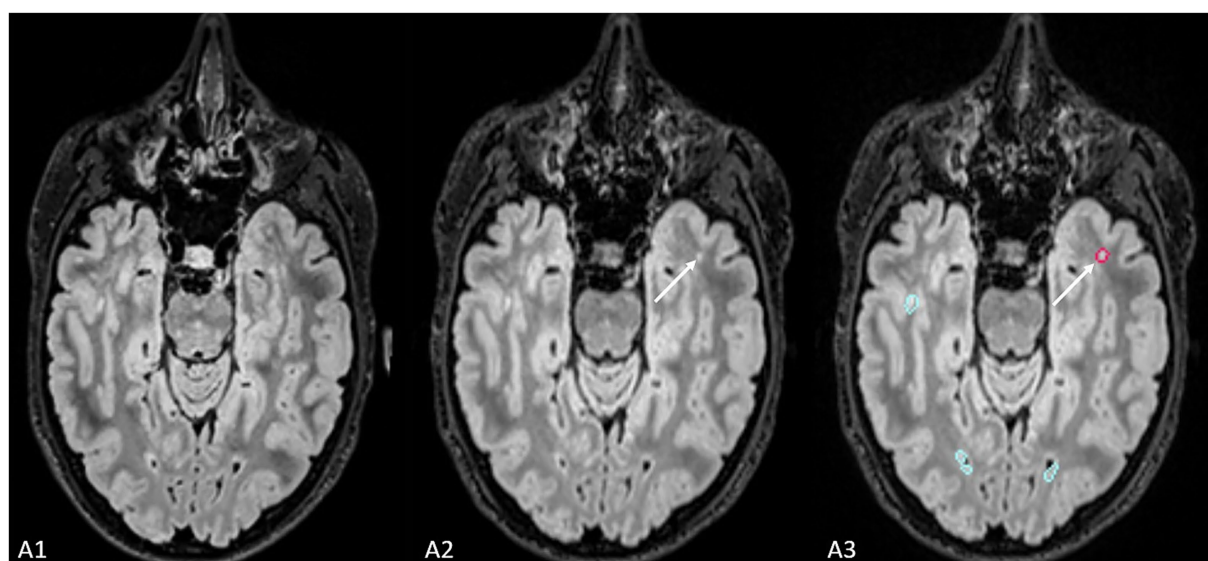


FIGURE 3

The figure shows an example of the lesion (indicated by the white arrow) missed by radiological evaluation by Jazz® software, while it was detected by AI-assisted radiological report made using Pixyl.Neuro.MS® software.

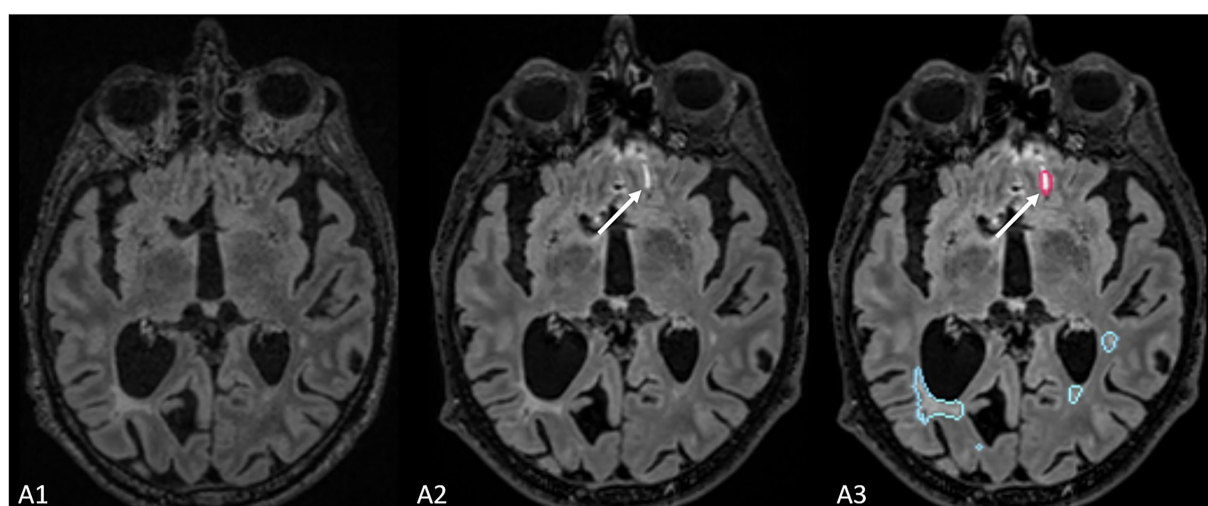


FIGURE 4

The figure represents a false positive lesion (indicated by the white arrow) which was segmented as a new lesion but actually represents an artifact located at the interface between cerebrospinal fluid and brain parenchyma.

and PPV (31.2%). These results suggest that both software solutions enhance lesion detection compared to standard practice, with Jazz® favoring specificity and precision, while Pixyl® maximizes sensitivity but requires radiologist oversight to filter out false positives (Table 2).

4 Discussion

In this study, AI-assisted radiological assessments significantly improved lesion detection in MS patients with a high lesion load.

Compared to standard MRI reading, both Pixyl.Neuro.MS®, which automates lesion segmentation, and Jazz®, which optimizes follow-up image display, enhanced radiologists' ability to identify new lesions ($z = -2.646$, $p = 0.01$ for both software). This resulted in additional lesion detection in 20 and 22% of cases (reader 1 and reader 2, respectively), highlighting the potential of AI in challenging diagnostic scenarios. The early and accurate detection of new FLAIR lesions is one of the most important parameters for therapeutic decision-making since it helps estimate actual disease activity (22, 23). Changes in the lesion count were found to influence therapeutic decisions in various ways, potentially leading to treatment escalation in cases of breakthrough disease. However, a manual follow-up

TABLE 2 Diagnostic testing accuracy table provides sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) based on the patient-level.

| Radiological report | Sensitivity | Specificity | PPV | NPV |
|---------------------|-------------|-------------|-------|-------|
| Standard report | 22.2% | 100% | 100% | 78.8% |
| Reader 1 + Jazz | 88.9% | 100% | 100% | 96.3% |
| Reader 2 + Jazz | 77.8% | 100% | 100% | 92.9% |
| Reader 1 + Pixyl | 100% | 100% | 100% | 100% |
| Reader 2 + Pixyl | 100% | 96.2% | 90% | 100% |
| Pixyl | 100% | 26.9% | 31.2% | 100% |

evaluation of MRI exams remains challenging for radiologists, especially in advanced stages that have a high lesion burden.

Several factors contribute to these difficulties. Initially the detection of new lesions is complicated by the multiple pre-existing lesions, especially when they are small or when their localization is in areas of tissue remodeling. Second, brain atrophy, a common feature in MS, alters brain morphology and can introduce interpretation errors. Finally, the interrater variability in neuroradiology still represents an issue which can be responsible for inconsistencies in lesion identification.

In this context, AI provides valuable assistance by improving objectivity and reproducibility in image analysis. Pixyl.Neuro.MS® and Jazz® exemplify two complementary approaches: segmentation-based versus co-registration/display-based. Pixyl.Neuro.MS® provides fully automatic quantitative outputs (lesion count, lesion volume, activity maps), offering a standardized method to track disease evolution, but at the cost of occasional false positives that require validation. Jazz®, on the other hand, provides a quantitative information but not the lesion masks, improving visual comparison between time points, thereby reducing the risk of missing subtle changes while keeping the final decision fully dependent on the radiologist.

These operational differences also imply distinct workflow consequences. In clinical practice, Pixyl.Neuro.MS® integrates with PACS and produces structured DICOM reports and overlays, which can facilitate multicenter standardization and longitudinal monitoring. Jazz® functions more as a reading accelerator, streamlining the toggling between prior and follow-up scans. The choice between the two may thus depend on whether the clinical priority is quantitative lesion tracking or rapid and confident visual assessment.

Our results align with a previous study that has explored AI's role in MS imaging. The study of Combes et al. (24) shows the improved detection of new lesions when utilizing segmentation masks during radiological reading, whereas the study of Federau et al. (25) shows the improved detection of new lesions using AI-based software for the longitudinal MRI assessments of MS patients.

However, the impact of AI on diagnostic efficiency depends mainly on data quality and algorithm training. Conversely to the studies whose focus was on moderate lesion load cases, our study specifically investigated patients with a high lesion burden, where AI's added value is particularly relevant. Our cohort was intentionally biased toward patients with a high lesion load, as the clinical advantage of AI-assisted detection is most pronounced when manual identification of new lesions becomes more difficult due to the sheer

number of pre-existing lesions. While previous AI studies in inflammatory conditions have typically included mixed cohorts (26), the greatest detection challenges, and thus the most potential for AI benefit occur in cases with extensive lesion burden. This specificity could explain why the improvement in detection rate observed in our study was statistically significant.

One major consideration for integrating AI into neuroradiology is its effect on workflow efficiency. In our study, the average reading time with AI assistance was 2 min 46 s for Pixyl.Neuro.MS® and 3 min 33 s for Jazz®, compared to an estimated 4 min for neuroradiologists and 8 min for radiology residents in conventional reading reported in the literature (27). A study by Sima et al. (32) showed that the use of AI software resulted in statistically shorter reading times compared to standard radiological reading. Similar results were shown by Peters et al. (28) and Combes et al. (24), who demonstrated shorter reading times when using AI compared to reading times without AI. Even though we did not compare reading times without and with the use of AI, our results suggest a potential time-saving benefit, though this advantage may be offset by the need to verify AI outputs. While AI identifies more lesions, it can also produce false positive results, which need to be validated by radiologists (29). The majority of false positive results in our study were associated with co-registration errors, either only co-registration or linked to the presence of severe atrophy. Additionally, some SELs that were considered new in new exams were reclassified as SELs upon further radiological evaluation. We also encountered artifacts that were mistakenly identified as new lesions. These artifacts typically appear at the boundary where fluid meets the air and the bone, which cannot be completely suppressed in 3D FLAIR sequences (30). This aspect of AI validation may explain why the reading times might even increase in clinical practice.

Several limitations should be considered when interpreting our results. Firstly, our study was conducted on a relatively small cohort of 35 patients, with total of 21 new lesions, which limits the generalizability of the findings but nevertheless provides an important first step toward understanding how such software solutions may perform in clinical practice. Patients which were included in our study mostly had high lesion loads (>20 lesions), which prevented us from assessing AI's performance in early-stage or low-burden cases. Secondly, MRI readings were performed in an already pre-prepared environment, which may not fully reflect everyday clinical workflow. A limitation of our study is that AI-assisted lesion detection was not applied under routine clinical conditions. Standard clinical reports were generated in daily practice, often under time constraints and other pressures, whereas the AI-assisted assessments were conducted in a controlled, experimental environment. This difference may partly contribute to the improved lesion detection observed with AI support. Future studies should investigate AI-assisted detection directly in clinical routine to better understand its practical impact. We evaluated only FLAIR sequences, whereas the integration of other sequences, such as DIR, DWI, and T1 post-contrast, could have an impact on the enhanced lesion characterization. Lastly, our study did not assess whether AI-enhanced lesion detection can be translatable to improve long-term clinical outcomes.

It should also be noted that since the time of this study, several new releases of Pixyl.Neuro.MS® and Jazz® have become available, with improvements specifically aimed at reducing false positives. Our results therefore reflect the performance of the evaluated version at the time of analysis, and may not fully capture the capabilities of more recent software iterations.

The use of AI tool which improves lesion detection opens a promising avenue in neuroradiology and aligns with the broader evolution toward augmented radiology, where AI assists rather than replaces radiologists. Several future directions should be explored. Firstly, AI needs to be validated in a clinical environment with a diverse patient cohort and radiologists of varying expertise levels. Secondly, in the study it was evaluated only FLAIR sequence using AI, but integration of other sequences, such as DIR, SWI or post contrast T1 sequences, could not only enhance lesion detection and characterization, but also improve generalizability of the findings. Thirdly, how AI can be better integrated into the clinical workflow in order to minimize validation time and maximize efficiency needs to be investigated. A hybrid AI-radiologist approach needs to be explored, mainly the fusion models between human expertise and AI, and incorporated into optimized double-reading protocols. Last but not least, a further investigation of the impact of AI-assisted radiological report on the clinical management of MS patients should be explored, especially concerning the treatment decision changes and patient outcomes, given the fact that number of new lesions and relapses in patients treated with interferon-beta are a useful tool for predicting the clinical disease activity in MS patients (31).

In conclusion, our study highlights the significant contribution of AI in detecting new lesions on follow-up MRI in MS patients with high lesion loads. Both Pixyl.Neuro.MS® and Jazz® significantly improved detection of new lesions compared to conventional readings. However, AI tools still require radiological validation to ensure that the results are accurate and also to eliminate the false positive results. Future AI integration must consider its impact on workflow efficiency and adaptation to clinical constraints.

AI represents a promising advancement in neuroradiology, but its clinical adoption requires further validation to optimize its performance and ensure meaningful contributions to patient care.

Data availability statement

The datasets presented in this article are not readily available because ethical considerations. Requests to access the datasets should be directed to myriam.edjlali@aphp.fr.

Ethics statement

The studies involving humans were approved by the Institutional Review Board of CPP Ile-de-France VI (ID RCB: 2019-A03066-51). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

MM: Investigation, Conceptualization, Writing – review & editing, Writing – original draft, Visualization, Data curation, Formal analysis, Methodology. OH: Investigation, Writing – review & editing, Methodology, Conceptualization, Resources. CF: Methodology, Conceptualization, Resources, Writing – review & editing. VM-R: Resources, Methodology, Conceptualization, Data curation, Writing – review & editing. MB: Methodology, Investigation, Writing – review & editing, Conceptualization, Resources. JB: Conceptualization, Writing – review & editing, Methodology. FC: Conceptualization, Writing – review & editing, Methodology. ME: Investigation, Conceptualization, Writing – original draft, Writing – review & editing, Methodology, Data curation, Supervision, Formal analysis.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

Vincent Lebon. This work was performed on a platform member of France Life Imaging network (grant ANR-11-INBS-0006).

Conflict of interest

Verónica Muños-Ramírez is an employee at Pixyl SA. Christian Federau is the founder and CEO of AI Medical AG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Filippi M, Preziosa P, Banwell BL, Barkhof F, Ciccarelli O, de Stefano N, et al. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*. (2019) 142:1858–75. doi: 10.1093/brain/awz144
- Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. (2018) 17:162–73. doi: 10.1016/S1474-4422(17)30470-2
- Baneke Peer (2020). Atlas of MS 3rd Edition The Multiple Sclerosis International Federation. Available online at: <https://www.msif.org/wp-content/uploads/2020/10/Atlas-3rd-Edition-Epidemiology-report-EN-updated-30-9-20.pdf> (Accessed June 25, 2025).
- Montalban X, Gold R, Thompson AJ, Otero-Romero S, Amato MP, Chandraratna D, et al.ECTRIMS/EAN guideline on the pharmacological treatment of people with multiple sclerosis. *Mult Scler J*. (2018) 24:96–120. doi: 10.1177/1352458517751049
- Baskaran AB, Grebenciucova E, Shoemaker T, Graham EL. Current updates on the diagnosis and management of multiple sclerosis for the general neurologist. *J Clin Neurol*. (2023) 19:217–29. doi: 10.3988/jcn.2022.0208
- Brown JWL, Coles A, Horakova D, Havrdova E, Izquierdo G, Prat A, et al. Association of Initial Disease-Modifying Therapy with Later Conversion to secondary progressive multiple sclerosis. *JAMA*. (2019) 321:175–87. doi: 10.1001/jama.2018.20588
- Lamb YN. Ocrelizumab: a review in multiple sclerosis. *Drugs*. (2022) 82:323–34. doi: 10.1007/s40265-022-01672-9
- Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, et al. 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *Lancet Neurol*. (2021) 20:653–70. doi: 10.1016/S1474-4422(21)00095-8
- Brisset J-C, Kremer S, Hannoun S, Bonneville F, Durand-Dubief F, Tourdias T, et al. New OFSEP recommendations for MRI assessment of multiple sclerosis patients: special consideration for gadolinium deposition and frequent acquisitions. *J Neuroradiol*. (2020) 47:250–8. doi: 10.1016/j.neurad.2020.01.083
- Kaunzner UW, Gauthier SA. MRI in the assessment and monitoring of multiple sclerosis: an update on best practice. *Ther Adv Neurol Disord*. (2017) 10:247–61. doi: 10.1177/1756285617708911
- Kamraoui RA, Mansencal B, Manjon JV, Coupé P. Longitudinal detection of new MS lesions using deep learning. *Front Neuroimaging*. (2022) 1:948235. doi: 10.3389/fnimg.2022.948235
- Fagotti C, Pertici L, Mancini S. Comparative analysis of three commercially available AI software for brain volumetry and lesion segmentation in multiple sclerosis. *ECR EPOS*. (2023). doi: 10.26044/ecr2023/C-20095
- Zheng D, He X, Jing J. Overview of artificial intelligence in breast Cancer medical imaging. *J Clin Med*. (2023) 12:419. doi: 10.3390/jcm12020419
- Shmatko A, Ghaffari Laleh N, Gerstung M, Kather JN. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat Can*. (2022) 3:1026–38. doi: 10.1038/s43018-022-00436-4
- Rezazade Mehrizi MH, Mol F, Peter M, Ranschaert E, dos Santos DP, Shahidi R, et al. The impact of AI suggestions on radiologists' decisions: a pilot study of explainability and attitudinal priming interventions in mammography examination. *Sci Rep*. (2023) 13:9230. doi: 10.1038/s41598-023-36435-3
- Dahlblom V, Andersson I, Lång K, Tingberg A, Zackrisson S, Dustler M. Artificial intelligence detection of missed cancers at digital mammography that were detected at digital breast Tomosynthesis. *Radiol Artif Intell*. (2021) 3:e200299. doi: 10.1148/ryai.2021200299
- Pacilè S, Lopez J, Chone P, Bertinotti T, Grouin JM, Fillard P. Improving breast Cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. *Radiol Artif Intell*. (2020) 2:e190208. doi: 10.1148/ryai.2020190208
- Commowick O, Istace A, Kain M, Laurent B, Leray F, Simon M, et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci Rep*. (2018) 8:13650. doi: 10.1038/s41598-018-31911-7
- Commowick O, Kain M, Casey R, Ameli R, Ferré JC, Kerbrat A, et al. Multiple sclerosis lesions segmentation from multiple experts: the MICCAI 2016 challenge dataset. *NeuroImage*. (2021) 244:118589. doi: 10.1016/j.neuroimage.2021.118589
- AI Medical. (2023). Available online at: <https://www.ai-medical.ch> (Accessed June 26, 2025).
- Pixyl. AI Insight for Improved Patient Care. (2024). Available online at: <https://pixyl.ai/> (Accessed June 26, 2025).
- Sormani M, Rio J, Tintorè M, Sormani MP, Signori A, Li D, et al. Scoring treatment response in patients with relapsing multiple sclerosis. *Mult Scler*. (2013) 19:605–12. doi: 10.1177/1352458512460605
- Sormani MP, Gasperini C, Romeo M, Rio J, Calabrese M, Cocco E, et al. Assessing response to interferon- β in a multicenter dataset of patients with MS. *Neurology*. (2016) 87:134–40. doi: 10.1212/WNL.0000000000002830
- Combès B, Kerbrat A, Pasquier G, Commowick O, le Bon B, Galassi F, et al. A clinically-compatible workflow for computer-aided assessment of brain disease activity in multiple sclerosis patients. *Front Med*. (2021) 8:740248. doi: 10.3389/fmed.2021.740248
- Federau C, Hainc N, Edjlali M, Zhu G, Mastilovic M, Nierobisch N, et al. Evaluation of the quality and the productivity of neuroradiological reading of multiple sclerosis follow-up MRI scans using an intelligent automation software. *Neuroradiology*. (2024) 66:361–9. doi: 10.1007/s00234-024-03293-3
- Peters S, Schmill L, Gless CA, Stürner K, Jansen O, Seehafer S. AI-based assessment of longitudinal multiple sclerosis MRI: strengths and weaknesses in clinical practice. *Eur J Radiol*. (2025) 183:111941. doi: 10.1016/j.ejrad.2025.111941
- Al Yassin A, Salehi Sadaghiani M, Mohan S. It is about “time”: academic Neuroradiologist time distribution for interpreting brain MRIs. *Acad Radiol*. (2018) 25:1521–5. doi: 10.1016/j.acra.2018.04.014
- Peters S, Kellermann G, Watkinson J, Gärtner F, Huhndorf M, Stürner K, et al. AI supported detection of cerebral multiple sclerosis lesions decreases radiologic reporting times. *Eur J Radiol*. (2024) 178:111638. doi: 10.1016/j.ejrad.2024.111638
- Haller S, Van Cauter S, Federau C. The R-AI-DIOLOGY checklist: a practical checklist for evaluation of artificial intelligence tools in clinical neuroradiology. *Neuroradiology*. (2022) 64:851–64. doi: 10.1007/s00234-021-02890-w
- Naganawa S, Kato Y, Yoshida T, Sone M. Fluid signal suppression characteristics of 3D-FLAIR with a T2 selective inversion pulse in the skull base. *Nat Commun*. (2023) 14:4915. doi: 10.1038/s41467-023-40507-3
- Sormani MP, Freedman MS, Aldridge J, Marhardt K, Kappos L, De Stefano N. MAGNIMS score predicts long-term clinical disease activity-free status and confirmed disability progression in patients treated with subcutaneous interferon beta-1a. *Mult Scler Relat Disord*. (2021) 49:102790. doi: 10.1016/j.msard.2021.102790
- Sima D, Wilms G, Vyvère TV (2020). On the use of icobrain's prepopulated radiology reporting template for multiple sclerosis follow-up. In: ECR 2020 EPOS. Available online at: <https://epos.mysr.org/poster/esr/ecr2020/C-11342>. (Accessed 19 Feb 2024)