Evaluating the Impact of Manual Editing on Automated MRI Lesion Segmentation in Multiple Sclerosis

F. X. Faust[1], **K. M. Leyden[1],** L Chance[1], D. Peterson[1], A. Keshavan, PhD[1]
[1]Octave, Menlo Park, CA

P262

Introduction

Multiple Sclerosis (MS) lesion quantification through Magnetic Resonance Imaging (MRI) can improve disease monitoring and treatment efficacy evaluation. The accuracy of lesion quantification derived from commercially available automated algorithms is impacted by image quality and protocol variation. To ensure the outputs provided by the algorithm are accurate, software tools recommend specific scanner and series parameters that may require modification of standard clinical and research protocols. Historical clinical exams and retrospective datasets may be incompatible with these recommended parameters, resulting in inaccurate segmentations. Manual edits of lesion segmentations can ensure accuracy of lesion counts and volumes, but the added time investment can hinder clinical adoption.

Objectives

This analysis explores the extent to which manual corrections of automated lesion segmentation masks (using a semi-automated editing tool) significantly impacts measures of lesion burden in patients with MS (pwMS).

Methods

- 86 pwMS were selected from a retrospective database to obtain 2 sessions of 3D T1 and 3D T2 FLAIR MRI series, for a total of 172 scans.
- Each scan was processed using a commercially available, FDA cleared software for automated lesion detection.
- A proprietary, semi-automated lesion segmentation editing tool was used to correct and refine lesion delineation.
  - Lesion segmentation masks from the automated algorithm were loaded into the tool, allowing the expert rater to correct the lesion segmentation. (Figure 1)
- A 2nd expert rater performed a final review of the lesion segmentation to confirm accuracy before updated lesion counts and volumes were extracted. (Figure 2)
- To compare the lesion volumes and lesion counts before and after manual editing, a two sample T-test was used to test differences in total lesion volumes and total lesion counts across 4 anatomical regions that correspond to the 2017 McDonald diagnostic criteria.

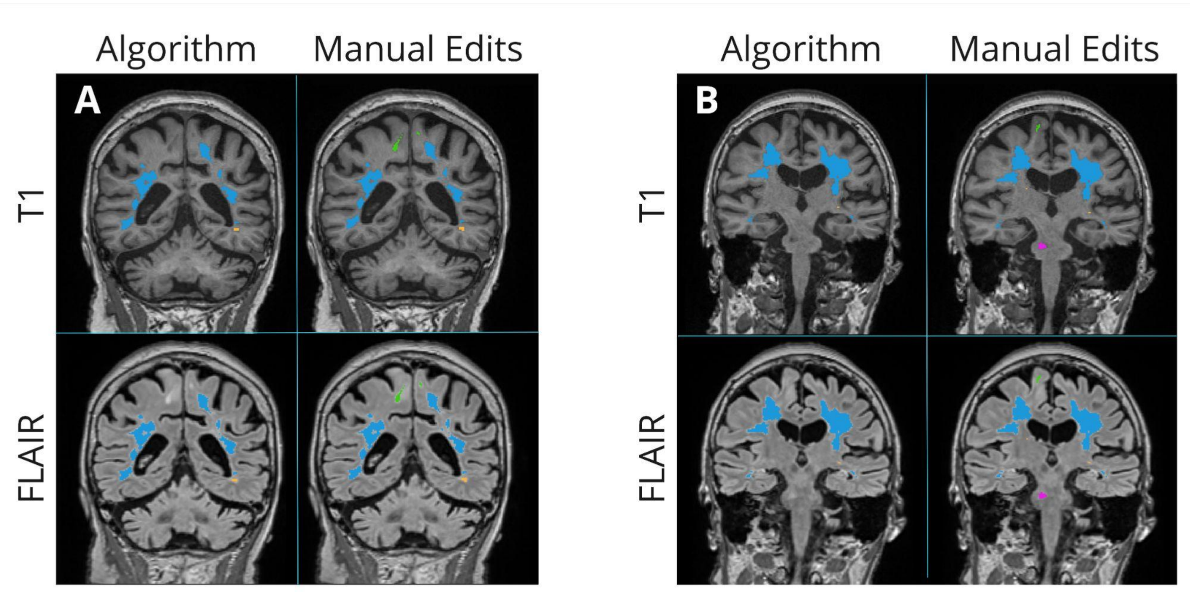

Figure 1: Case example of lesion segmentation mask before and after semi-automated manual edits. Examples A and B are 2 different slices from a single subject.

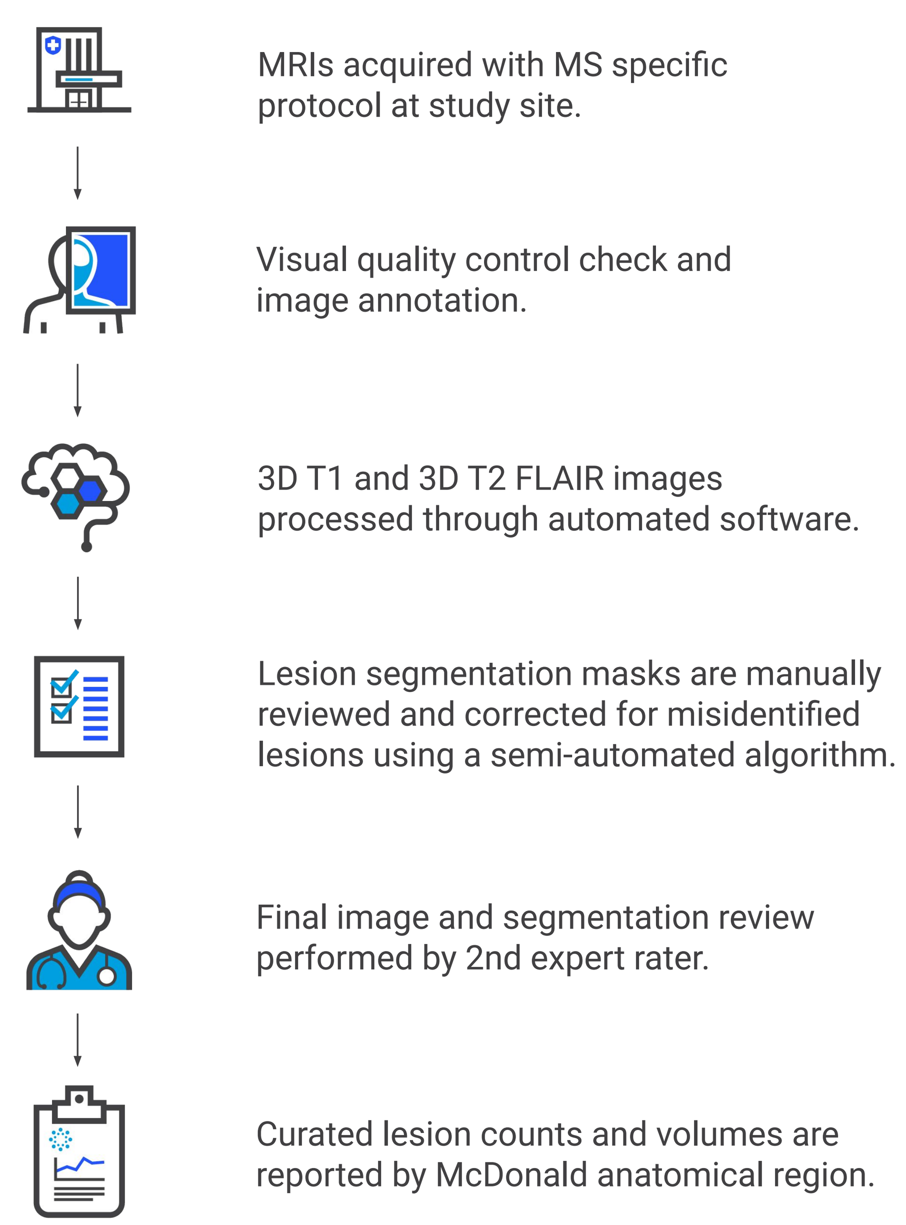

Figure 2: Illustration of the workflow performed by an expert rater to process MS MRIs through automated software, then correct the lesion segmentation mask outputs using semi-automated lesion detection tools.

Disclosure: FXF, KML, DP, LC and AK are employees of Octave Bioscience, Inc.

Results

- When combining all lesion categories, the overall impact of semi-automated manual lesion edits is evident in the significant increase in total lesion count (p=2.564e-10) (Table 1). The average manual editing time per scan was **188 minutes** (SD=256).

Lesion Source	Mean Total Lesion Count	Mean Lesion Volume (cm3, SD)	Minimum Lesion Size (cm3)	Median Lesion Size (cm3)	Maximum Lesion Size (cm3)
Algorithm	37.093	0.21 (1.16)	0.001	0.015	23.631
Manual Edits	55.889	0.17 (1.09)	0.001	0.016	23.733

Table 1: The distribution of total lesion counts, highlighting the increase of detected lesions while the distribution of lesion volumes remains stable.

- A significant increase in lesion counts (Figure 3) and volumes (Figure 4) in the Deep White, Juxtacortical, and Infratentorial regions was observed after manual editing. The Infratentorial region had the highest increase in lesion counts and volumes, followed by Deep White then Juxtacortical regions. Manual edits did not produce significant changes in lesion count or volume in the Periventricular region. (Table 2)

Region	Lesion Count	Lesion Volume
Deep White	**+83.3%** (p=6.037e-13)	+46.2% (p=1.420e-5)
Juxtacortical	+38.4% (p=0.009)	+59.6% (p=1.257e-13)
Infratentorial	+172.4% (p=3.115e-10)	+498.4% (p=1.257e-13)
Periventricular	-3.1% (p=0.509)	+14.2% (p=0.305)

Table 2: The percentage change and p-values of lesion counts and lesion volumes after performing manual edits upon the automated algorithm lesion segmentation masks.

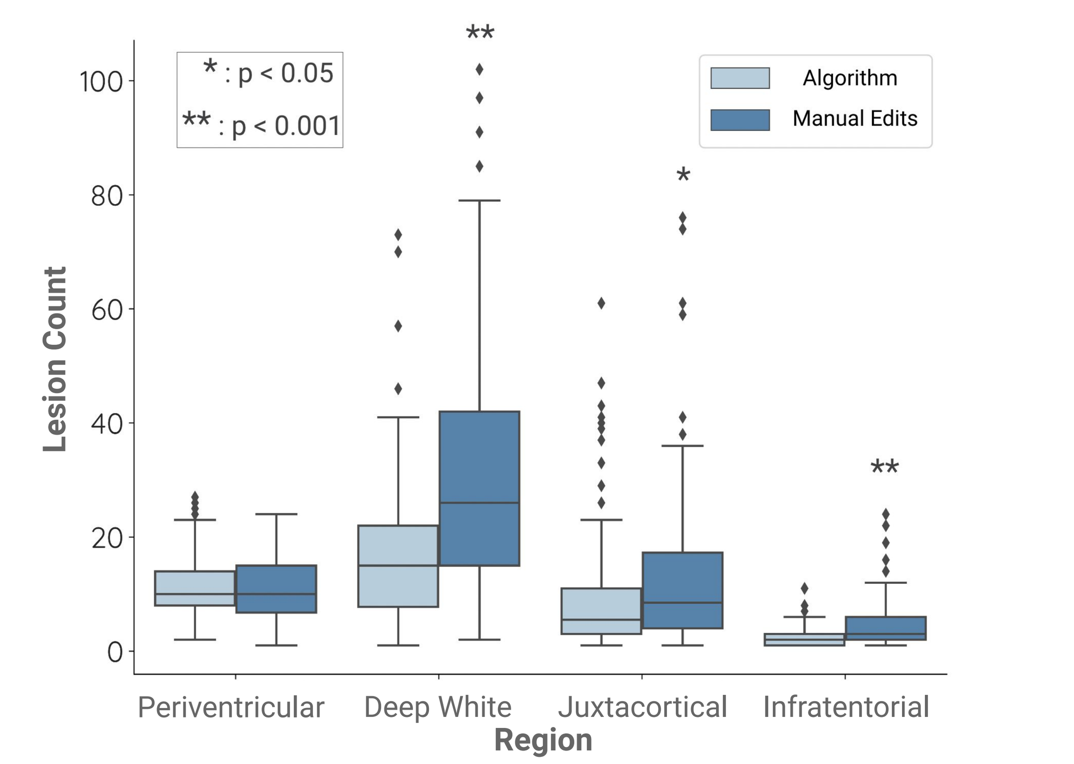

Figure 3: The distribution of lesion count results from the automated algorithm (light blue) and semi-automated manual lesion edits (dark blue).

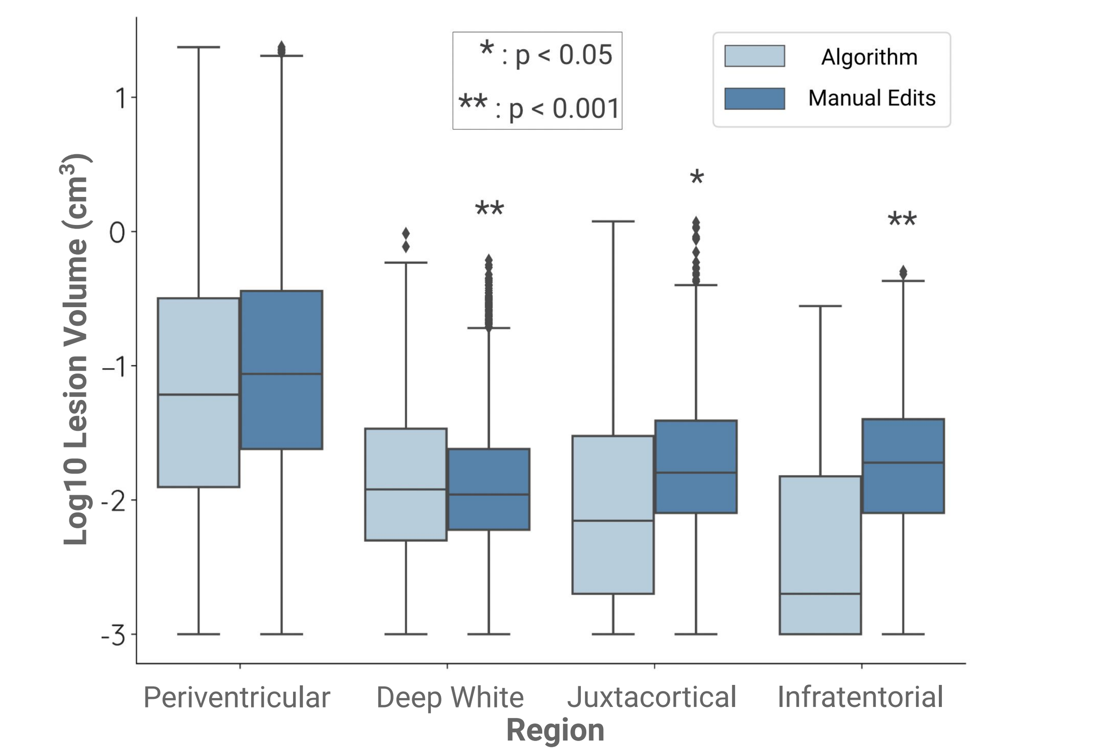

Figure 4: The distribution of lesion volume results from the automated algorithm (light blue) and semi-automated manual lesion edits (dark blue).

Conclusions

Inaccurate measures of lesion burden may hinder the adoption of quantitative MRI algorithms in clinical standard of care. Our results show that manual editing of automated MS lesion detection in pwMS yields an increase in Infratentorial, Deep White, and Juxtacortical lesion burden, areas that may influence clinical decision making. The manual edits in this analysis required considerable time investment and specialized resources. A focus on lesion editing in the most salient regions may maximize time-efficiency and boost the clinical utility of lesion quantification in MS patient care.

Presented at the Ninth Congress of the Americas Committee for Treatment and Research in Multiple Sclerosis (ACTRIMS), February 29 - March 2, 2024, West Palm Beach USA