Automatic Segmentation of Multiple Sclerosis Lesions in Clinical Quality MRI: a Benchmarking Study

Johanna Appleton[1], Reese Triana[2], Shaila Gopal[1], Quentin Finn[3], Paul Wolujewicz[1,2,†], Zaenab Dhari[1,4,5, †]

1. Frank H. Netter MD School of Medicine, Quinnipiac University, North Haven, CT
2. Department of Biomedical Sciences, Quinnipiac University, Hamden, CT
3. Nantz National Alzheimer Center, Houston Methodist Hospital, Houston, TX
4. Mandell MS Center, Trinity Health, Hartford, CT
5. UConn Health, Farmington, CT

† These authors contributed equally to the work

Background

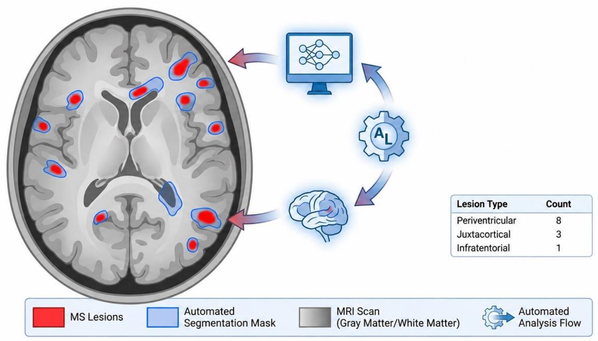
White matter lesions are key markers of multiple sclerosis (MS) disease activity and progression and are used to monitor treatment response. However, clinical MRI scans often have lower resolution and thicker slices than research scans, limiting their use in research pipelines. Several automated lesion segmentation tools have been developed to address these challenges, including WMH-SynthSeg[1] and MindGlide[2], which can segment lesions from routine clinical MRI scans.

Methods

We analyzed 145 MRI scans from two publicly available MS datasets: MSLesSeg[3] (n=115) and OpenMS[4] (n=30). Four automated lesion segmentation tools were evaluated: SAMSEG[5], WMH-SynthSeg, LST-AI[6-10], and MindGlide. WMH-SynthSeg and MindGlide used FLAIR images alone, whereas SAMSEG and LST-AI required both T1-weighted and FLAIR images.

The primary outcome was the Dice similarity coefficient, calculated against reference lesion masks from both datasets. Secondary outcomes included sensitivity and precision. For MSLesSeg, Dice scores were also correlated with reference lesion volume and MRI slice thickness.

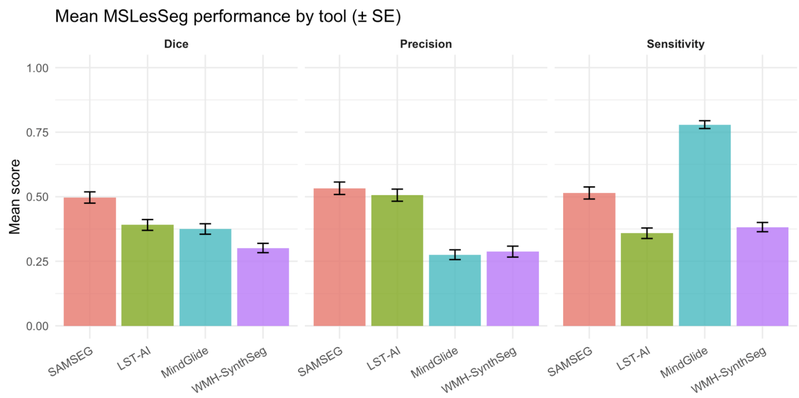
Brain MRI with MS Lesions and Automated Analysis

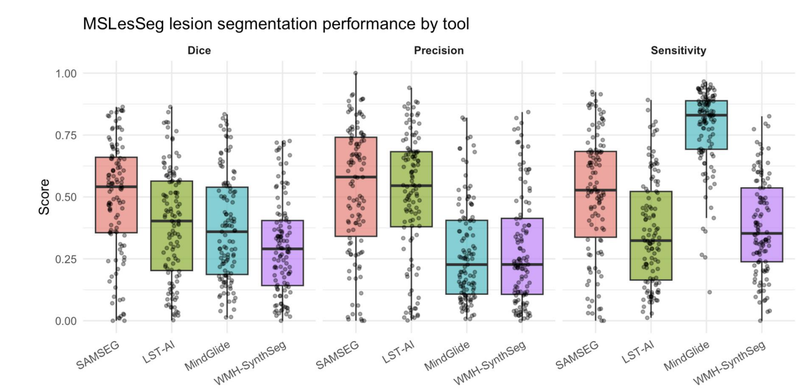


MSLesSeg
- 115 Scans
- 3T and 1.5 T
- Different protocols and scanners
- Diverse pool of MRIs performed for clinic use

OpenMS
- 30 Scans
- 3T scans
- Same scanner and protocol
- Scans from same hospital

Results





Method	MSLesSeg Dice Coefficeint N=115 †	MsLesSeg R_s, Lesion Volume and Dice	MSLesSeg R_s Slice Thickness and DICE	OpenMs Dice Coefficient N=30
SAMSEG	0.49	0.63**	-0.17	0.42
LST-AI	0.39	0.47**	-0.34*	0.65
WMH-SynthSeg	0.30	0.82**	0.24	0.32
MindGlide	0.37	0.94**	-0.15	0.48

†For SAMSEG, N=114. SAMSEG failed to run on one acquisition
*p<0.01 , **p<0.001

Conclusions and Limitations

Conclusions
- Across tools run on the MSLesSeg data set, lesion load correlated with Dice score, with higher lesion load prediciting better tool performance. MindGlide and WMH-SynthSeg had the highest correlations.
- Scanner quality most significantly impacted LST-AI:
  1) Only tool that had a significant correlation of slice thickness to Dice score
  2) OpenMS, a dataset consisting only 3T scans aquired on the same scanner, the dice score of LST-AI far outperformed other methods, and had the largest change of any method
- This emphasized the importance of manual review and correction of lesion masks, as well as choosing initial tools suited to specific MRIs, while also showing potential for automatic segmentation.

Limitations
- MSLesSeg does not provide scanner strength by acquisition.
- Slice thickness is used as a proxy of scan resolution and quality.

Future Aims

- This work is being completed as part of a study on BH4, a component of oxidative stress in Multiple Sclerosis, and findings will be used to select appropriate tool for MRI lesion segmentation.
- As work ongoing concordantly with benchmarking of automatic methods, we are training several raters on manual lesion counts, and inter-rater reliability is being evaluated.
- Once the appropriate tool is selected for lesion segmentation on our BH4 MRI dataset, then it will be used for initial segmentation.
- Trained raters will manually review and edit automatically generated lesion masks.

Ackowledgements

This work was done as part of the the Netter Summer Research Fellowship funded through the You've Got This, Kid! Leadership Foundation.

References