ORIGINAL RESEARCH

Cross-Sectional Validation of an Automated Lesion Segmentation Software in Multiple Sclerosis: Comparison with Radiologist Assessments

Maria Vittoria Spampinato, Heather R. Collins, Hannah Wells, William Dennis, Jordan H. Chamberlin, Emily Ye, Justin A. Chetta, Maria Gisele Matheus, Seth T. Stalcup, Donna R. Roberts

ABSTRACT

BACKGROUND AND PURPOSE: Magnetic Resonance Imaging is widely used to assess disease burden in multiple sclerosis (MS). This study aimed to evaluate the effectiveness of a commercially available k-nearest neighbors (k-NN) software in quantifying white matter lesion (WML) burden in MS. We compared the software's WML quantification to expert radiologists' assessments.

MATERIALS AND METHODS: We retrospectively reviewed brain MRI examinations of adult MS patients and of adult patients without MS and with a normal brain MRI referred from the neurology clinic. MRI images were processed using an AI-powered, cloud-based k-NN software, which generated a DICOM lesion distribution map and a report of WML count and volume in four brain regions (periventricular, deep, juxtacortical, and infratentorial white matter). Two blinded radiologists performed semi-quantitative assessments of WM lesion load and lesion segmentation accuracy. Additionally, four blinded neuroradiologists independently reviewed the data to determine if MRI findings supported an MS diagnosis. Results were considered significant when p < 0.05.

RESULTS: The study included 32 MS patients (35.4 years ± 9.1) and 19 patients without MS (33.5 years ± 12.1). The k-NN software demonstrated 94.1% and 84.3% accuracy in differentiating MS from non-MS subjects based respectively on WML count and WML volume, compared to radiologists' accuracy of 90.2% to 94.1%. Lesion segmentation was more accurate for the deep WM and infratentorial regions than for the juxtacortical region (both p <0.001).

CONCLUSIONS: k-NN-derived WML volume and WML count provide valuable quantitative metrics of disease burden in MS. AI-powered post-processing software may enhance the interpretation of brain MRIs in MS patients

ABBREVIATIONS: MS = multiple sclerosis; k-NN =k-Nearest Neighbors; WML =white matter lesion; MPRAGE = Magnetization-Prepared Rapid Acquisition Gradient Echo; SPACE = Sampling Perfection with Application-optimized Contrasts using a Different Flip Angle Evolution; EDSS = Expanded Disability Status Scale.

Received month day, year; accepted after revision month day, year.
From the Department of Radiology, Medical University of South Carolina, Charleston, SC, USA (MVS, HRC, WD, JHC, JAC, MGM, STS, DRR), College of Medicine, Medical University of South Carolina, Charleston, SC, USA (HW, EY).

Dr. Spampinato has been the recipient of research funding from Siemens Healthineers.

Please address correspondence to Maria V. Spampinato, MD, Department of Radiology and Radiological Science, Medical University of South Carolina, 96 Jonathan Lucas Street, MSC 323, Charleston, SC 29425, USA, Email address: spampin@musc.edu.

SUMMARY SECTION

PREVIOUS LITERATURE: Previous studies evaluating automated white matter lesion (WML) segmentation algorithms, including k-Nearest Neighbors (k-NN), primarily used 3T MRI datasets and focused on overlap-based metrics like the Dice Similarity Coefficient. Few studies explored the algorithm's real-world performance across varied MRI field strengths or compared outputs directly to radiologist assessments, an important factor for clinical applicability.

KEY FINDINGS: The k-NN algorithm demonstrated high accuracy in WML segmentation, particularly in infratentorial and deep white matter. The k-NN software demonstrated comparable accuracy to radiologists in differentiating MS from non-MS subjects, with better performance of k-NN-based WM lesion count than WM lesion volume.

KNOWLEDGE ADVANCEMENT: This study evaluates the use of a k-NN-based WML segmentation algorithm in an external cohort and highlights its potential to streamline radiology workflows by providing objective WML quantification, while identifying areas for refinement, particularly in juxtacortical lesion segmentation.

INTRODUCTION

Multiple sclerosis (MS) is an autoimmune disorder of the central nervous system characterized by recurrent episodes of inflammation, demyelination, and axonal loss, which can lead to a progressive accumulation of neurologic disability. The McDonald criteria for diagnosing MS rely on clinical evaluation, laboratory tests, and presence of lesions on T2-weighted images, particularly fluid-attenuated inversion recovery (FLAIR) MRI sequences1. MRI supports

the diagnosis by demonstrating whether the criteria for dissemination in time and space are met1, 2. Lesion segmentation can provide imaging biomarkers of disease burden3. However, manual delineation of MRI lesions requires considerable time, resources and technical expertise.

Automated lesion segmentation methods offer objectivity and speed for lesion detection and quantification and have the potential to improve MS patient care4. Artificial intelligence offers powerful solutions for medical image post-processing. Machine Learning can automate complex tasks as image segmentation and perform these analyses fast with high accuracy and reproducibility5. K-nearest Neighbors Networks (k-NN), a machine learning algorithm that classifies data points by analyzing the 'k' most similar neighbors in space, have been employed for lesion segmentation in MS and other disorders 6-10. Previous studies evaluating automated white matter lesion (WML) segmentation using k-NN algorithms, primarily used 3 Tesla MRI datasets and focused on overlap-based metrics like the Dice Similarity Coefficient. Few studies explored the algorithm's real-world performance across varied MRI field strengths or compared outputs directly to radiologist assessments, an important factor for clinical applicability.

Our purpose was to evaluate the performance of a commercially available k-NN algorithm in quantifying white matter lesion (WML) burden in brain MRI examinations of MS patients, compared to expert radiologists, using an external validation dataset

MATERIALS AND METHODS

Patient Cohort

This article follows the STARD reporting guidelines. Institutional review board approval was obtained for this retrospective single-center cross-sectional validation study, including a waiver of informed consent. We retrospectively reviewed all brain MRI examinations performed at our hospital system between July and September 2021, using a newly implemented MRI protocol for MS (described below) during the first three months of utilization of the protocol. Inclusion criteria were the following: 1) Age range = 18-65 years; 2) MS diagnosis1; 3) Brain MRI performed per our institution's MS protocol, including a 3D T2 FLAIR Sampling Perfection with Application-optimized Contrasts using a Different Flip Angle Evolution (SPACE) sequence11 and 3D Magnetization-Prepared Rapid Acquisition Gradient Echo (MPRAGE) without intravenous contrast administration; 4) Neurologic evaluation within 1 month of the MRI. We excluded patients with neurologic disorders other than MS, incomplete studies, or MRI examinations degraded by head motion or other artifacts. The degree of disability was assessed by using the Kurtzke Expanded Disability Status Scale (EDSS)12, ranging from 0 to 10, with higher scores more severe disability. As a comparison group, we included in the study consecutive patients without MS who were referred to our imaging facilities from the neurology clinic between July and September 2021 with the following inclusion criteria: 1) Age range = 18-65 years; 2) Brain MRI interpreted as normal for age by the interpreting radiologist and two senior co-authors; 3) Brain MR imaging including FLAIR SPACE and MPRAGE sequences; 4) No known neurologic disorder; 5) Normal neurologic evaluation within 1 month of the MR imaging.

Imaging Protocol and Post-processing

MRI examinations were performed on 1.5 Tesla and 3 Tesla MR scanners (Magnetom Lumina, Skyra, Aera, Sola, and Avanto, Siemens Healthineers, USA) using 16- and 20-channel head coils (see Supplementary Table 1 for scanning parameters). Preliminary mage quality assessments for study inclusion were performed by two board-certified neuroradiologists.

Post-processing was performed offline using AI-Rad Companion Brain MR White Matter Hyperintensities, an AI application that extends AI-Rad Companion Brain MR Brain Morphometry by performing segmentation and quantification of white matter hyperintensities. MPRAGE and FLAIR SPACE DICOM images were post-processed using cloud-based software. The framework combines two approaches: 1. A supervised method was used to obtain a map of potential lesional tissue using a k-nearest neighbors (k-NN) classifier trained on a set of features (signal intensities, spatial location coordinates, and tissue prior probabilities) obtained from atlas-based prior probability maps of gray matter, white matter, and CSF; 1. A Bayesian Partial Volume estimation algorithm6-8. The White Matter Hyperintensity segmentation method included the following steps (Supplementary Figure 1): (1) Pre-processing: Images were aligned, skull-stripped, corrected for bias field and intensity normalized; (2) Lesion segmentation was performed by a k-NN classifier, in which each voxel was labeled with a value representing the probability of containing lesion or healthy tissue13. Prior probability maps were included to estimate realistic concentration maps of normal and abnormal brain tissue. Probability maps are employed to guide tissue classification (gray matter, white matter, cerebrospinal fluid, and lesions) and to compute lesion volumes14. This algorithm automatically evaluates WMH lesion load (count and volume) while accounting for the mixing of normal and lesional tissue within voxels due to partial volume effects7.

The software generated: 1. A color-coded lesion distribution maps overlaid on the sagittal FLAIR SPACE images; 2. A segmentation report of WML count (lesion number) and WML volume (lesion volume [mL]) for four brain subregions (periventricular, deep, juxta-cortical, and infratentorial WM).

Radiologist Imaging Interpretation

After a training session, two blinded board-certified radiologists, a neuroradiology attending and a neuroradiology fellow (20 and 4 years of experience) independently reviewed MPRAGE, FLAIR SPACE, and color-coded lesion maps superimposed on sagittal FLAIR SPACE images on Syngo.via (Siemens Healthineers, Malvern, PA, USA) in a randomized manner. During two separate sessions the neuroradiologists performed the following assessments:

1. Radiologist-assessed Lesion Load: Neuroradiologists independently evaluated multiplanar 3D FLAIR images of each subject in a random order and rated lesion load of each brain region (periventricular, deep, juxta-cortical, and infratentorial WM) on a scale ranging from 0 to 3 (0 = none, 1 = mild, 2 = moderate, 3 = severe) (Supplementary Figure 2). From these ratings, we calculated the radiologist-assessed WML load, as the sum of the lesion load for each region.

2. Radiologist Assessment of color-coded lesion distribution maps: In a separate session, neuroradiologists evaluated the software performance by reviewing the native FLAIR images along with the model's color-coded lesion distribution maps, which were superimposed on the sagittal FLAIR SPACE images. Reader discrepancies were adjudicated during a joint reading session. Radiologists performed the following assessments.

a. Semi-quantitative assessment of false positive and false negative lesions by brain subregion: Neuroradiologists provided a semiquantitative rating of the number of false positive (FP) and false negative (FN) lesions on the k-NN-based color-coded lesion distribution map using the following scale: 0 = no FP/FN, 1 = 1-5 FP/FN lesions, 2 = 6-10 FP/FN lesions, 3 = at least 11 FP/FN lesions.

b. Semi-quantitative evaluation of lesion size by brain subregion: Neuroradiologists assessed how well the borders of the segmented color-coded lesions matched the lesion borders on the native FLAIR images: 1 = underestimation of lesion size, 2 = slight underestimation, 3 = accurate, 4 = slight overestimation, 5 = overestimation.

c. Semi-quantitative ratings of overall WMH segmentation accuracy by brain subregion: Neuroradiologists rated the overall WMH segmentation accuracy on a scale of 1 to 5: 1 = very poor, 2 = poor, 3 = fair, 4 = good, 5 = excellent.

3. Radiologist Diagnosis of MS. Three blinded board-certified neuroradiologists and a neuroradiology fellow (20, 15, 10, and 4 years of experience), not involved in the previous evaluations, reviewed multiplanar FLAIR SPACE images in a randomized order, and determined whether findings supported a diagnosis of MS.

Statistical Analysis

Age, disease duration, lesion volume, and lesion count were summarized using means, and standard deviations. Independent Mann-Whitney U tests were used to compare MS and non-MS on the k-NN WML volume and count. Sensitivity, specificity, and ROC curves for the WML volume and count, with area under the curve (AUC) and 95% confidence interval (CI), were also reported. Radiologist-assessed WML load ratings were summarized using medians, median absolute deviations, and ranges. Comparisons across brain regions for FP lesions, FN lesions, lesion size accuracy, and overall WML segmentation accuracy were performed using Friedman's two-way analysis of variance by ranks, with Bonferroni correction. These data were summarized with medians and median absolute deviations. The associations between radiologist-rated WML load and CNN model WML volume and count were evaluated with Spearman's rank order correlation coefficient (rho) because these variables were not normally distributed. Reader agreement in MS diagnosis was evaluated with a two-way random effect, single measures, absolute agreement intraclass correlation coefficient (ICC) and 95% CI. An ROC curve was constructed using the mean confidence values of all four readers, with AUC and 95% CI. Two-sided p-values were reported, statistical significance was set at the $\alpha < .05$ threshold. Analyses were conducted with SPSS (version 28, IBM: Armonk, NY).

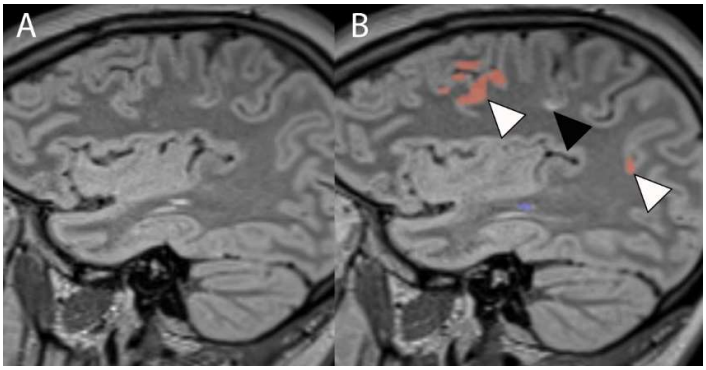

FIG 1. Sagittal native T2 FLAIR SPACE image (A) and color-coded lesion map superimposed on the T2 FLAIR SPACE image (B) demonstrate false positive (white arrowhead) and false negative results (black arrowhead) for the juxtacortical region.

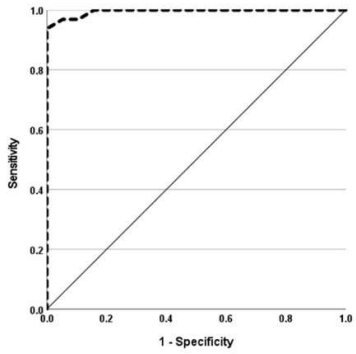

FIG 2. The ROC curve for all four readers yielded an AUC of .995, 95% CI = .921 - 1.000, $p < .001$.

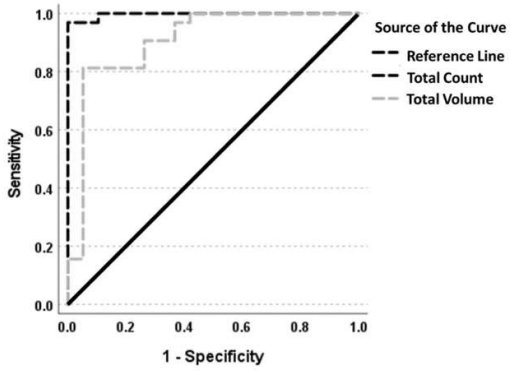

FIG 3. ROC Curves for the total WMH count (AUC = .997; 95% CI= .924 - 1.000; $p < .001$) and total WMH volume (AUC = .905; 95% CI = .789 - .969; $p < .001$).

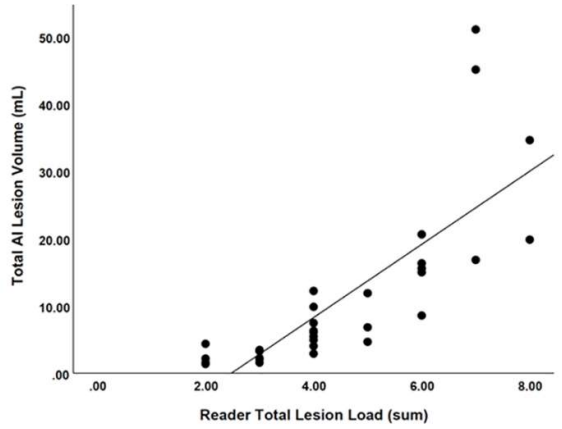

FIG 4. The scatterplot shows the k-NN model Total Lesion Volume (mL) and Reader Total Lesion Load values, rho = .91, p < .001.

RESULTS

Patient population: 32 patients with MS and 19 non-MS subjects were included in the study. See Table 1 for clinical characteristics of the study subjects. EDSS ranged between 0 and 6 with a median of 2.5 (MAD = 1.25).

1. **Automated Segmentation Report and Radiologist-assessed Lesion Load:**
   Lesion segmentation was carried out using a commercially available tool, with guidance from priori probability maps to inform tissue classification. The resulting WML counts and WML volumes (mL) for each brain subregions, and total WML counts and volumes (mL), as well as sensitivity, specificity, and accuracy are reported in Table 2. Radiologist-assessed WML load is reported in Table 3. There was no significant difference in k-NN lesion count and k-NN lesion volume between MS patients who underwent brain MRI at 1.5T and those scanned at 3T, as shown by an independent-samples Mann-Whitney U test (lesion count: U = 78, p = 0.192; lesion volume: U = 64, p = 0.061). While lesion volume appeared slightly higher in the 1.5T subgroup, this difference was not statistically significant (Supplementary Table 2).

2. **Radiologist Assessment of Color-coded Lesion Distribution Map Accuracy**:

a. Semi-Quantitative Assessment of False Positive and False Negative Lesions: In MS patients, the juxtacortical region exhibited significantly more FP lesions than the infratentorial region (p < .001) (Figure 1), and the periventricular region more FP lesions than the infratentorial (p < .001) and deep WM (p < .001). A similar pattern was observed in the non-MS group, with more FP lesions in the juxtacortical (p = .02) and periventricular (p = .005) than the infratentorial regions (Table 4).

A greater number of FN lesions were found in the juxtacortical region compared to the infratentorial (p = .02) and periventricular (p = .02) regions in MS patients. No FN lesions were observed in the non-MS group (Table 4).

b. Semi-Quantitative Evaluation of Lesion Size by Brain Subregion: In the MS group, lesion size estimation was more accurate in the infratentorial (p = 0.01), and deep WM regions (p = .006) compared to the periventricular WM. The k-NN model tended to overestimate lesion size in the periventricular and juxtacortical regions (Table 4).

c. Semi-Quantitative Ratings of Overall WMH Segmentation Accuracy: Radiology readers found that MS lesion segmentation accuracy was significantly worse in the juxtacortical area compared to the infratentorial (p = .001) and deep WM (p < .001) regions. For the non-MS group, readers rated lesion segmentation accuracy as being significantly worse in the periventricular compared to the infratentorial region, p = .03 (Table 4). Additionally, no significant differences were observed in lesion segmentation accuracy by region between MS patients who underwent brain MRI at 1.5T and those scanned at 3T (Supplementary Table 2).

3. **Radiologist Diagnosis of MS**:
   The four neuroradiologists demonstrated excellent agreement in classifying subjects as either MS or non-MS, ICC = .86, 95% CI = .79 - .91. Accuracy ranged from 90.2% to 94.1%, sensitivity ranged from 84.4% to 93.8%, and specificity ranged from 94.7% to 100.0% (Supplementary Table 3). The ROC curve for all four readers yielded an AUC of .995, 95% CI = .984 – 1.00, p < .001 (Figure 2).

4. **Classification of MS and non-MS Subjects Using the K-NN Model Report:**
   The k-NN-based total WML count, and volumes (mL) were significantly greater in MS compared to non-MS subjects (both p<.001). The total WML count in the deep and infratentorial WM showed respectively excellent (94.7%) and very good sensitivity (89.5%) in the classification of MS versus non-MS subjects. The sensitivity of periventricular and juxtacortical WML count were respectively 78.9% and 73.7%. The specificity of WML count was above 80% across all brain regions (periventricular: 81.3%, juxtacortical: 87.5%, deep WM: 96.9%) (Table 2).

With respect to WML volume (mL), the sensitivity of the deep WM lesion volume was 85.7%. The sensitivity for periventricular, juxtacortical, and infratentorial WML volume were respectively 58.8%, 46.2%, and 50.0%. Specificity was above 80% for the periventricular (90.6%) and juxtacortical (90.0%) areas (Table 2). The ROC curve for WML count and volume respectively yielded an AUC of .997, 95% CI = .924 - 1.000, $p < .001$, and an AUC of .905, 95% CI = .789 - .969, $p < .001$ (Figure 3).

There was a significant association between the k-NN model Total Lesion Volume (mL) and Reader Total Lesion Load values (rho = .91, $p < .001$) (Figure 4).

DISCUSSION

We evaluated the performance of a commercially available k-NN algorithm in the segmentation of WML and in the differentiation of MS patients from non-MS patients. We found that the accuracy of lesion segmentation varied across brain regions and was most accurate for the deep WM and infratentorial regions. The k-NN software demonstrated comparable accuracy to radiologists in differentiating MS from non-MS subjects, with better performance of k-NN-based WM lesion count than WM lesion volume.

K-NN models have been applied to MS lesion segmentation6, 7, 9, 10. The accuracy of k-NN classification of MS WM lesions can be improved by optimizing signal intensity normalization and by adding in the model GM, WM, and CSF tissue priors from healthy controls10. Fartaria et al. tested a k-NN lesion segmentation method using 3D FLAIR, MPRAGE, Magnetization Prepared Two Inversion-contrast Rapid Gradient-echo (MP2RAGE), and 3D Double-inversion Recovery (3D DIR) in MS patients with low disability status 6. The combination of routine clinical sequences as MPRAGE and 3D FLAIR was adequate to detect small white matter lesions. However, the addition of more advanced sequences such as MP2RAGE and 3D DIR improved the automated detection of juxta-cortical lesions. MP2RAGE and 3D DIR have relatively long acquisition times and are not routinely included in clinical MRI protocols for MS. Farfaria et al. also compared the k-NN lesion segmentation method using 3D FLAIR and MPRAGE with manual lesion segmentation and two open-access software applications, and found that the k-NN method outperformed the others due to improved segmentation of small lesions, prone to partial volume effects7. The performance of commercially available k-NN applications for lesion quantification in MS requires further investigation.

Our approach differed from previous studies by having neuroradiologists evaluate the k-NN model's performance. Additionally, while most prior research was conducted using 3T MRI6, 7, 10, our study included scans acquired at both 1.5T and 3T, providing a broader evaluation of the model's performance in clinical settings, where patients are commonly scanned on MRI systems at different field strengths. The k-NN model segmentation accuracy varied across different regions. Specifically, the model was accurate in segmenting infratentorial and deep WM lesions, with few FP and FN WML reported in these regions. The evaluation of WML size was more accurate for infratentorial and deep WML than for periventricular WML. In contrast, juxtacortical and periventricular WML segmentation was less accurate, with FP WM lesions reported in both MS and non-MS patients. In addition, FN WM lesions were reported for the juxtacortical region of MS patients. Accurate segmentation of the juxtacortical and periventricular lesions is clinically important for MS diagnosis, as dissemination in space of white matter lesions is a key criterion but also for monitoring disease progression1. The lower accuracy in juxtacortical and periventricular segmentation could be attributed to several factors. The proximity and the similarity in signal intensity of the juxtacortical WML and cortex coupled with the inherent complexity and variability of gyral anatomy may have resulted in incorrect segmentation (see Figure 1). A potential solution would be to train the software to filter juxtacortical WML based on their shape, typically round, ovoid, irregular, or U-shaped, in contrast with the curvilinear cortical signal. Refining the algorithm to recognize these shape and location characteristics could improve its ability to distinguish juxtacortical lesions from the surrounding anatomical structures. Additional training of the algorithm with a dataset enriched in juxtacortical lesions, might also improve segmentation accuracy in this region. Furthermore, the use of MP2RAGE and 3D DIR sequences may improve the automated detection of juxta-cortical lesions6. While these sequences are relatively too long to be routinely included in clinical protocols at many institutions, faster acquisition protocols with AI-powered image reconstruction technology may soon overcome this limit15. Additionally, occurrences of incorrect periventricular segmentation were due to WML size overestimation, mislabeling of the caudate nucleus and normal T2 FLAIR hyperintensity along the lateral ventricles. In summary, the k-NN software labeled a higher number of WML than the radiologists, with FP results especially in the juxta-cortical and periventricular region. This issue may arise because the algorithm might be overly sensitive to certain features leading to incorrect lesion identification. Finding a balance between sensitivity and specificity remains a key challenge in automated segmentation tools for clinical use.

The k-NN model demonstrated high accuracy in differentiating MS from non-MS subjects, with an overall greater accuracy for WM lesion count (sensitivity and specificity of 89.5% and 96.9%) compared to WM lesion volume (sensitivity and specificity of 83.8% and 81.3%). Among the four readers, accuracy ranged from 90.2% to 94.1% with sensitivity ranging from 84.4% to 93.8% and specificity ranging from 94.7% to 100.0%. Notably, the k-NN model's performance in terms of sensitivity and specificity is comparable to the range observed among human readers, indicating its potential reliability as an adjunct tool in clinical settings.

The primary benefit of an automated WM segmentation method is its potential to streamline the radiologist interpretation workflow by providing consistent and rapid lesion quantification. Annotated FLAIR images and a report of WML count and volume would enable a rapid assessment of disease burden, potentially saving time during study interpretation. In addition, quantitative measurements could be included in the radiology report as objective markers of disease severity and progression, which in the future could aid clinicians in monitoring lesion burden and treatment response over time. Automated tools to improve workflow efficiency would be particularly valuable in high-volume clinical settings and for managing follow-up imaging in MS. While we have not specifically tested the utilization of the algorithm in the longitudinal assessment of MS lesion burden, this application represents an area where workflow efficiency would be highly beneficial.

Our study has several limitations, including the retrospective study design and small convenience sample size, consisting of all patients meeting the inclusion and exclusion criteria during the first three months of implementation of an MS MRI protocol including 3D FLAIR and MPRAGE sequences. We acknowledge that the relatively small size of the patient and control cohorts limits the generalizability of

the results. Additionally, the MRI scanners used varied across the patient and control cohorts, which introduced variability. However, it is important to note that this algorithm has already been tested by the developers on a larger, internally controlled dataset. The aim of our study was to evaluate the algorithm in an external cohort to better understand its strengths and weaknesses when applied to clinical cases, where variability in imaging protocols and equipment is expected. The focus of this work is not on developing a novel method, but rather on evaluating a commercially available machine learning algorithm. While deep learning segmentation methods are advancing rapidly, the purpose of this study is to assess the performance and generalizability of a commercially available machine learning tool when applied outside the original training set. Since our MS sample included predominantly mild/moderate cases (median EDSS = 2.5, range 0-6), software performance should be evaluated in a sample including MS patients with more severe disability. Our comparison group consisted of patients referred from the neurology clinic with a brain MRI normal for age and neurological examinations, rather than typical healthy controls. While this may introduce a bias, as this group does not represent the general population, these patients were deemed adequate as a control group for our study purposes. The k-NN model produced color-coded lesion distribution maps overlaid only on sagittal FLAIR images. The reviewers noted that additional coronal and axial images would have been beneficial, especially for infratentorial assessment. We evaluated the accuracy of the model in differentiating MS from non-MS subjects as a secondary measure of the segmentation algorithm's accuracy. Our aim was to evaluate how well the algorithm's segmentation aligns with MS-related patterns identified by radiologists, not to imply that the algorithm alone could serve as a stand-alone diagnostic tool. We did not perform manual segmentation of MS lesions, nor did we evaluate the overlap between lesion masks generated by the automated algorithm and manually defined ground truth using the Dice Similarity Coefficient, as has been done in previous studies assessing algorithm performance 6, 7 . Instead, our goal was to compare the software's output with the type of semi-quantitative assessments of white matter hyperintensity burden that are commonly employed in clinical practice. This study is structured to assess the practice utility of the software in a clinical setting where radiologists typically provide a semi-quantitative assessment. We evaluated the performance of the algorithm in the detection of WMH in MS patients, and we did not test its application in the segmentation of WMH secondary to other pathologies. As a result, our reported findings of regional performance differences are relevant only to the detection of lesions patterns in MS, not those associated with other conditions. The control group consisted of non-MS patients with brain MRI scans that were normal for age. This study design does not capture the more complex diagnostic scenarios encountered in clinical practice, such as distinguishing MS from other conditions with T2 hyperintense lesions (e.g., chronic small vessel disease, other demyelinating processes). While this design allows for a controlled evaluation of the algorithm's performance, it may limit the general applicability of the results. Future studies should focus on testing the algorithm in cohorts with more diagnostically challenging differential diagnoses to better assess its potential utility in clinical workflows. In this study, we did not integrate the tool into our clinical workflow or evaluate its impact on the speed of study interpretation. Future studies could explore how the tool might influence workflow efficiency, diagnostic accuracy, and overall clinical utility when integrated into routine practice. Despite these limitations, we believe that our study provides valuable insight into the effectiveness of the software in the evaluation of MS patients.

Table 1: XYZ.

Column A	
Total A	X
Total B	YZD
C	A (XY.Z%)
D	B (CD.E%)
Age (yr)	
Z	54
D	22–82
Total Z	243
Median	5
Range	234-235
ABCDE	
A	B
B	A
C	C
D	D
E	E
Z	E

CONCLUSIONS

AI-powered post processing software may be a useful adjunct to the interpretation of brain MRIs in MS. The information obtained regarding cumulative WML volume and count may serve as quantitative metrics of disease burden. K-NN tools for longitudinal WM lesion analyses will improve the efficiency of the radiologist workflow and serve as a valuable adjunct to radiologist interpretation.

ACKNOWLEDGMENTS

The study was supported by a research grant from Siemens Healthineers. The funding source had no role in data collection, analysis, interpretation of the study results, and manuscript preparation and editing.

REFERENCES

1. 1. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. The Lancet Neurology 2018;17:162-173
2. 2. Filippi M, Rocca MA. MR Imaging of Multiple Sclerosis. Radiology 2011;259:659-681
3. 3. Diaz-Hurtado M, Martínez-Heras E, Solana E, et al. Recent advances in the longitudinal segmentation of multiple sclerosis lesions on magnetic resonance imaging: a review. Neuroradiology 2022;64:2103-2117
4. 4. Egger C, Opfer R, Wang C, et al. MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation? Neuroimage Clin 2017;13:264-270
5. 5. Bonacchi R, Filippi M, Rocca MA. Role of artificial intelligence in MS clinical practice. Neuroimage Clin 2022;35:103065
6. 6. Fartaria MJ, Bonnier G, Roche A, et al. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis. J Magn Reson Imaging 2016;43:1445-1454
7. 7. Fartaria MJ, Todea A, Kober T, et al. Partial volume-aware assessment of multiple sclerosis lesions. Neuroimage Clin 2018;18:245-253
8. 8. Ayrignac X, Le Bars E, Duflos C, et al. Serum GFAP in multiple sclerosis: correlation with disease type and MRI markers of disease severity. Sci Rep 2020;10:10923
9. 9. Griffanti L, Zamboni G, Khan A, et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. Neuroimage 2016;141:191-205
10. 10. Steenwijk MD, Pouwels PJ, Daams M, et al. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). Neuroimage Clin 2013;3:462-469
11. 11. Wattjes MP, Ciccarelli O, Reich DS, et al. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol 2021;20:653-670
12. 12. Kurtzke JF. Rating neurologic impairment in multiple sclerosis. An expanded disability status scale (EDSS) 1983;33:1444-1444
13. 13. Wu Y, Warfield SK, Tan IL, et al. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. Neuroimage 2006;32:1205-1215
14. 14. Maldjian JA, Laurienti PJ, Kraft RA, et al. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 2003;19:1233-1239
15. 15. Kiryu S, Akai H, Yasaka K, et al. Clinical Impact of Deep Learning Reconstruction in MRI. Radiographics 2023;43:e220133

SUPPLEMENTAL FILES

Supplementary Table 1.

Representative scanning parameters of the sagittal MPRAGE and T2 FLAIR SPACE sequences at 1.5 Tesla and 3 Tesla (1.5T/3T).

	MPRAGE 1.5T/3T	**T2 FLAIR SPACE 1.5T/3T**
TR	1900/1900 ms	7000/5000 ms
TE	2.91/2.54 ms	406/386 ms
TI	1100/900 ms	2050/1800 ms
Flip angle	15°/ 9°	120°/120°
Bandwidth	130/220 Hz/Px	651/750 Hz/Px
Voxel size	1 x 1 x 1 mm^3	1 × 1 × 1 mm^3

Supplementary Table 2.

	Multiple Sclerosis (3 Tesla MRI)	**Multiple Sclerosis (1.5 Tesla MRI)**	***p*-value[4]**
Number of subjects	22	10	
Sex (M/F)	3/19	2/10	
Age (yr.)[1]	35.2 ± 8.9	35.8 ± 10.0	
Subtype of MS [2]RR/PP/SP/unknown	21/0/0/1	8/0/0/2	
Years since initial MS diagnosis (yr.)[1]	6.5 ± 5.5	8.0 ± 7.9	
EDSS[3]	2.5, 0-3	2, 1.13-2.85	
k-NN total WM lesion volume[1]	9.94± 9.46	16.3 ± 16.7	.061
k-NN total WM lesion count[1]	37.1 ± 31.4	29.2 ± 5.0	.192
Periventricular WM Lesion Segmentation Accuracy	3.5, 3 - 4	3.75, 3.38 - 4	.684
Juxta-cortical WM Lesion Segmentation Accuracy	3, 2.5 - 3.5	2.75, 1.38 - 3.13	.171
Infratentorial Lesion Segmentation Accuracy	5, 3 - 5	4, 3 – 4.63	.455
Deep WM Lesion Segmentation Accuracy	4, 3.88 - 4.63	4.5, 3.88 – 4.5	.660

[1] Mean ± Standard Deviation
[2] Subtypes of MS: Relapsing remitting (RR), primary progressive (PP), secondary progressive (SP), and unknown subtype.
[3] Median, interquartile range
[4] *p*-values not corrected for multiple comparisons

Supplementary Table 3.

Accuracy, sensitivity, and specificity for all radiologists.

Readers	**Accuracy**	**Sensitivity**	**Specificity**
Reader 1[1]	94.1%	93.8%	94.7%
Reader 2[1]	92.2%	87.5%	100.0%
Reader 3[1]	94.1%	93.8%	94.7%
Reader 4[2]	90.2%	84.4%	100.0%

[1]Neuroradiology Attendings

[2]Neuroradiology Fellow

Supplementary Figure 1.

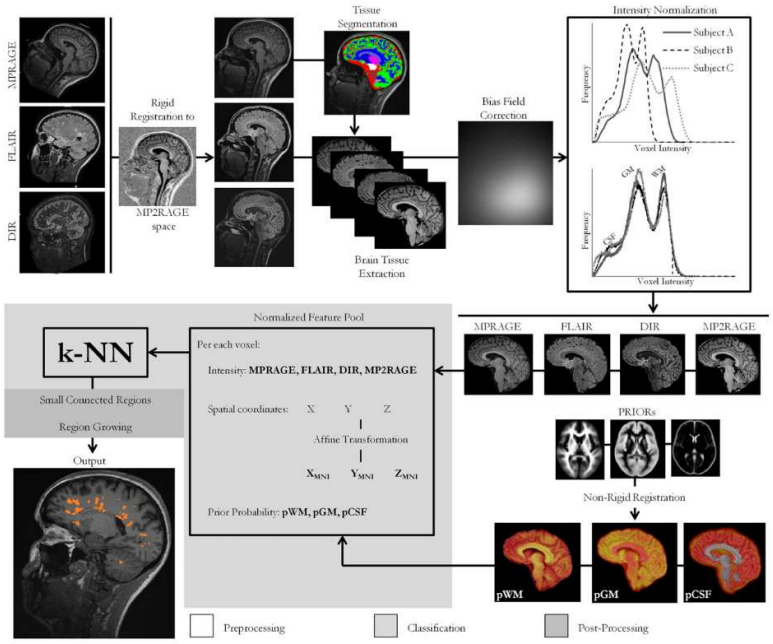

Figure Legend: Schematic diagram of the lesion segmentation algorithm, divided into three main steps: preprocessing (white background), classification (light gray background), and postprocessing (dark gray background). MPRAGE, FLAIR, DIR, and MP2RAGE are the abbreviations for the used image contrasts. Please note that in our study we employed only MPRAGE and FLAIR images. pWM, pGM, and pCSF are representative of the prior probabilities (PRIORs) of the respective brain tissues: white matter, gray matter, and cerebrospinal fluid. Reproduced with permission from reference 7 (Fartaria MJ, Bonnier G, Roche A, et al. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis. *J Magn Reson Imaging* 2016;43:1445-1454).

Supplementary Figure 2.

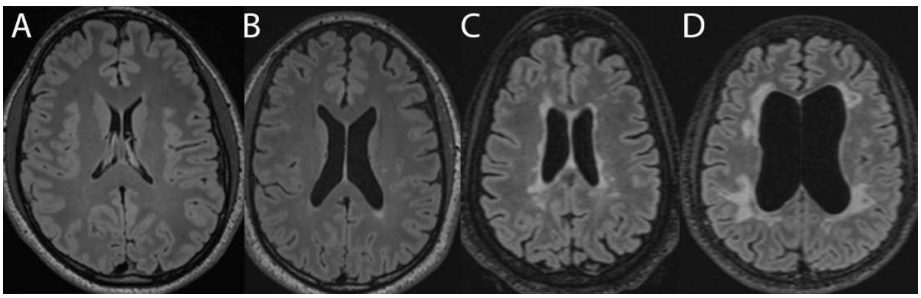

Figure Legend: Neuroradiologists evaluated multiplanar 3D FLAIR images of each subject in a random order and rated lesion load of each brain region (periventricular, deep, juxta-cortical, and infratentorial WM) using a 3-point Likert scale. The figure illustrates examples of normal brain (A), mild (B), moderate (C), and severe (D) periventricular white matter lesion burden.