

# Generic Contrast Agents

Our portfolio is growing to serve you better. Now you have a *choice*.



[VIEW CATALOG](#)

# AJNR

## **A Radiomic "Warning Sign" of Progression on Brain MRI in Individuals with MS**

Brendan S. Kelly, Prateek Mathur, Gerard McGuinness, Henry Dillon, Edward H. Lee, Kristen W. Yeom, Aonghus Lawlor and Ronan P. Killeen




This information is current as of May 25, 2025.

*AJNR Am J Neuroradiol* 2024, 45 (2) 236-243

doi: <https://doi.org/10.3174/ajnr.A8104>

<http://www.ajnr.org/content/45/2/236>

# A Radiomic “Warning Sign” of Progression on Brain MRI in Individuals with MS

 Brendan S. Kelly, Prateek Mathur, Gerard McGuinness, Henry Dillon,  Edward H. Lee,  Kristen W. Yeom, Aonghus Lawlor, and Ronan P. Killeen



## ABSTRACT

**BACKGROUND AND PURPOSE:** MS is a chronic progressive, idiopathic, demyelinating disorder whose diagnosis is contingent on the interpretation of MR imaging. New MR imaging lesions are an early biomarker of disease progression. We aimed to evaluate a machine learning model based on radiomics features in predicting progression on MR imaging of the brain in individuals with MS.

**MATERIALS AND METHODS:** This retrospective cohort study with external validation on open-access data obtained full ethics approval. Longitudinal MR imaging data for patients with MS were collected and processed for machine learning. Radiomics features were extracted at the future location of a new lesion in the patients' prior MR imaging (“prelesion”). Additionally, “control” samples were obtained from the normal-appearing white matter for each participant. Machine learning models for binary classification were trained and tested and then evaluated the external data of the model.

**RESULTS:** The total number of participants was 167. Of the 147 in the training/test set, 102 were women and 45 were men. The average age was 42 (range, 21–74 years). The best-performing radiomics-based model was XGBoost, with accuracy, precision, recall, and F1-score of 0.91, 0.91, 0.91, and 0.91 on the test set, and 0.74, 0.74, 0.74, and 0.70 on the external validation set. The 5 most important radiomics features to the XGBoost model were associated with the overall heterogeneity and low gray-level emphasis of the segmented regions. Probability maps were produced to illustrate potential future clinical applications.

**CONCLUSIONS:** Our machine learning model based on radiomics features successfully differentiated prelesions from normal-appearing white matter. This outcome suggests that radiomics features from normal-appearing white matter could serve as an imaging biomarker for progression of MS on MR imaging.

**ABBREVIATIONS:** AI = artificial intelligence; MSSEG2 = MS SEGmentation Challenge 2; NAWM = normal-appearing white matter

MS is a chronic, progressive, idiopathic, demyelinating disorder of the CNS.<sup>1</sup> Diagnosis is contingent on timely and precise application of the McDonald criteria, which rely on the interpretation of MR imaging.<sup>2</sup> It is unusual among many chronic conditions in that imaging features can often predate clinical manifestations of disease. Imaging remains a prominent tool in the diagnosis, progress-monitoring, and evaluation of treatment efficacy.<sup>1,2</sup> New MS lesions are an important imaging biomarker because they can signify both disease progression

and the efficacy of disease-modifying drugs.<sup>3</sup> Indeed, the absence of new T2-FLAIR lesions in the CNS is used as the solitary metric of disease-modifying drug effectiveness.<sup>4</sup> However, the monitoring of lesions can be a tedious or repetitive task for neuroradiologists,<sup>5</sup> and that issue, compounded by supply-demand issues in radiology,<sup>6</sup> has prompted research into automated lesion identification.<sup>7</sup>


Computer-assisted segmentation methodologies have been an important topic for scientific exploration for the past 2 decades.<sup>8</sup> More recently, a substantial increase in the incorporation of artificial intelligence (AI) into these methodologies has been observed.<sup>9</sup> Current research trends are shifting away from the simple identification of MS lesions on T2-FLAIR to comparing images captured

Received August 2, 2023; accepted after revision November 8.

From the Department of Radiology (B.S.K., G.M., H.D., R.P.K.), St. Vincent's University Hospital, Dublin, Ireland; Insight Centre for Data Analytics (B.S.K., P.M., A.L.), and School of Medicine (B.S.K.), University College Dublin, Dublin, Ireland; Wellcome Trust and Health Research Board (B.S.K.), Irish Clinical Academic Training, Dublin, Ireland; and Lucille Packard Children's Hospital at Stanford (E.H.L., K.W.Y.), Stanford, California.

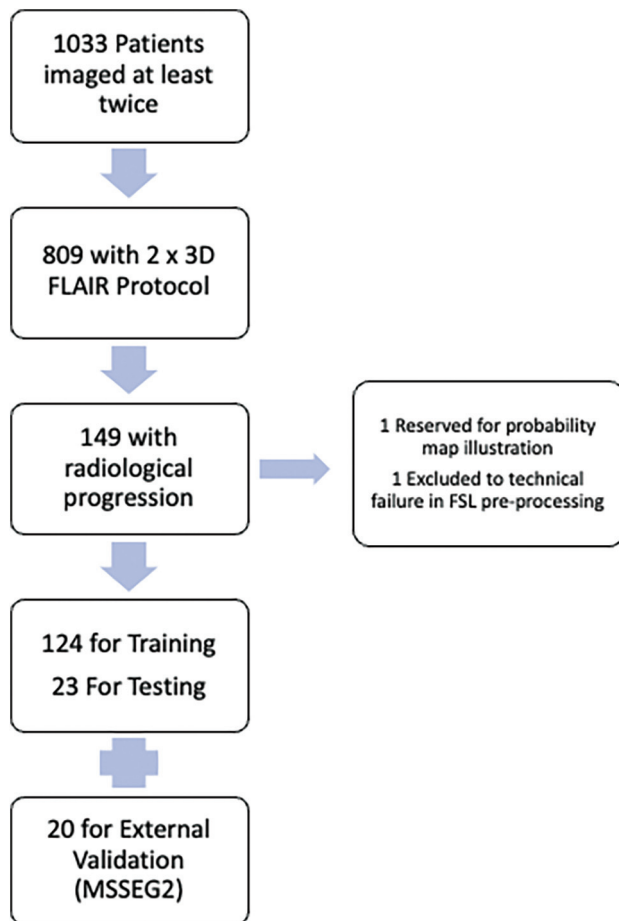
This work was performed within the Irish Clinical Academic Training Program, supported by the Wellcome Trust and the Health Research Board (grant No. 203930/B/16/Z); the Health Service Executive National Doctors Training and Planning and the Health and Social Care, Research and Development Division, Northern Ireland; the Faculty of Radiologists, Royal College of Surgeons in Ireland; and the Science Foundation Ireland (SFI) under grant No. SFI/12/RC/2289\_P2.

Please address correspondence to Brendan S. Kelly, MB, Department of Radiology, St Vincent's University Hospital, Dublin, Ireland; e-mail: brendanskelly@me.com; @BSKRad

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

 Indicates article with online supplemental data.

<http://dx.doi.org/10.3174/ajnr.A8104>



**FIG 1.** Flow chart of included patients.

at distinct time intervals.<sup>7</sup> In response to this escalating interest, the Medical Image Computing and Computer Assisted Interventions society initiated a challenge concentrating on the detection of new lesions (MS SEGmentation Challenge 2 [MSSEG2]), further galvanizing research interest in this domain.<sup>3</sup>

Predictive models in MS have also been the subject of research; however, their methods and external generalizability vary.<sup>10</sup> Many of these models use “systems biology” approaches (Omics data) to predict various disease markers and outcomes.<sup>11</sup> Recent imaging research into predictive modeling for MS includes applications in cognitive decline<sup>12</sup> and deep learning approaches to disease progression.<sup>13</sup> Radiomics is a field of research in which higher-order image features such as texture or intensity distribution are investigated (often using AI and machine learning models to glean deeper information from radiologic images).<sup>14,15</sup> The methodology is widely used in medical imaging research, including applications in MS.<sup>16-19</sup> By transforming simple images into mineable high-dimensional data, it allows in-depth characterization of MS lesions. Thus, radiomics has this potential to augment diagnostic accuracy and individualize patient management.<sup>7</sup>

While previously radiology AI research was focused on a narrow range-of-use cases<sup>6</sup> aligned to the above advances, there has been a significant increase in interest in the concept of temporality in the field.<sup>20</sup> The interpretation of medical images is not a static process, and recent research has shown the importance

both of incorporating prior imaging<sup>21</sup> into the pipeline and also considering medical images as part of a time-series and in making predictions.<sup>22,23</sup>

In this study, we aimed to evaluate the potential of radiomics features in predicting progression on MR imaging of the brain in individuals with MS. In our approach, we propose using a machine learning model and radiomics features to differentiate a “prelesion” from a control sample in the normal-appearing white matter (NAWM).

## MATERIALS AND METHODS

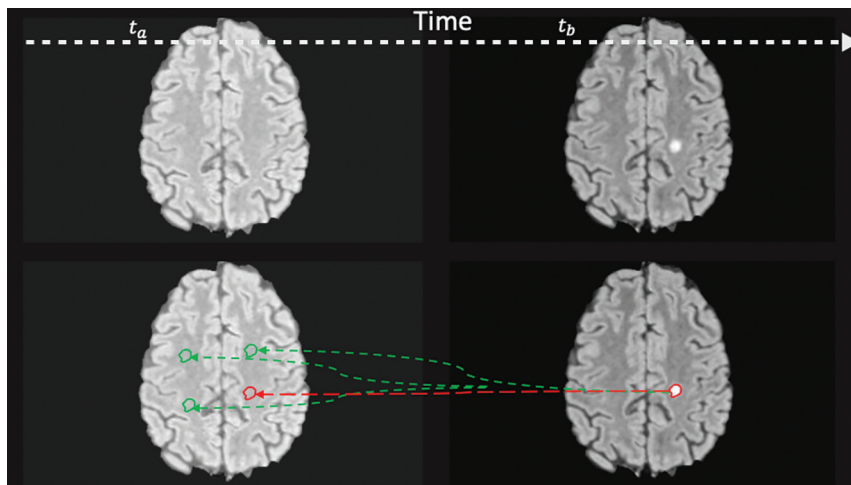
This retrospective study with external validation was designed according to both the Radiological Society of North America and the European Society of Radiology published principles,<sup>24,25</sup> with patient expert involvement.<sup>26</sup> We obtained independent statistical advice. The article was prepared using the Checklist for Artificial Intelligence in Medical Imaging<sup>27</sup> and conforms to the Assessment of Radiomics rEsearch (ARISE) guideline.<sup>28</sup> It received full institutional review board approval (St. Vincent’s University Hospital), and the requirement for prospective consent was waived. This research constitutes Level 5A evidence (data quality and AI model development with external testing) because it represents 1 retrospective study with internal and external data used for final performance reporting.<sup>29</sup>

### Participants and Data

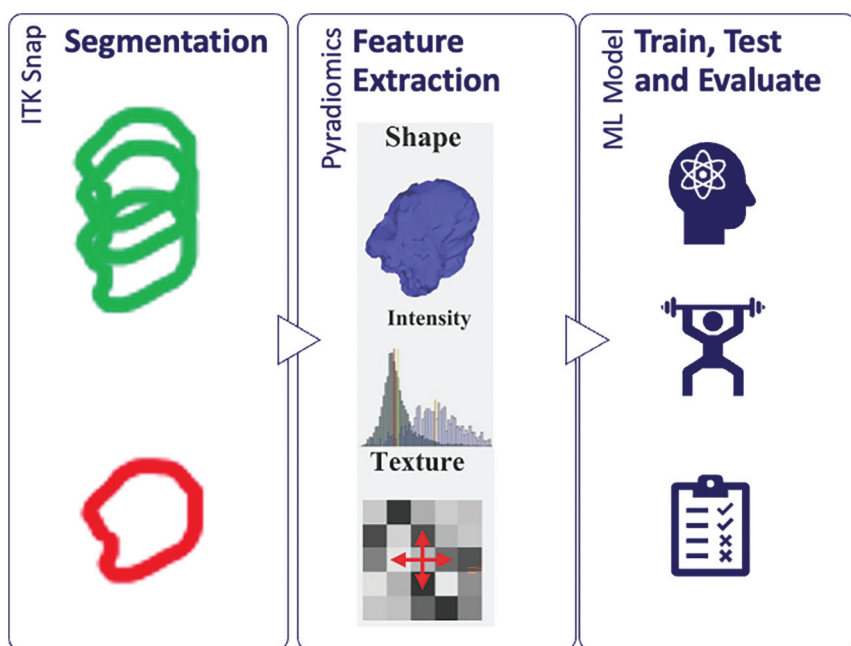
Consecutive patients who had at least 2 MR imaging brain studies for MS at our institution between January 2019 and December 2022 were reviewed (Fig 1). Those with a new lesion on follow-up imaging were included in this study. There is no overlap of the internal cohort with prior studies. The external cohort is publicly available and has been previously described.<sup>3</sup>

Images were acquired on a 1.5T system (Magnetom Avanto syngo MR B19; Siemens). Imaging sequences included a 3D T2-FLAIR sequence using the following parameters: acquired voxel size,  $1.1 \times 1.1 \times 1.1$  mm; TR, 6000 ms; TE, 413 ms; TI, 2030 ms; acquisition time, 6 minutes 44 seconds; orientation, sagittal. All scans were defaced using the FSL Brain Extraction Tool 10 (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET/UserGuide>) to preserve participant anonymity.<sup>30</sup> All images were coregistered to the first time point also using FSL (<http://www.fmrib.ox.ac.uk/fsl>).

A baseline automated segmentation of MS lesions was generated using DeepMedic (<https://github.com/deepmedic/deepmedic>).<sup>31</sup> These baseline segmentations were then manually corrected by 1 of 2 certified radiologists in their first year post-board examination using ITK-Snap Version 3.8.0 ([www.itksnap.org](http://www.itksnap.org)).<sup>32</sup> Radiologic progression (new or enlarging lesions) was defined according to the Magnetic Resonance Imaging in MS (MAGNIMS) consensus guidelines (<https://www.magnims.eu/>).<sup>33</sup> Cases with progression were initially identified from the radiologic report and confirmed at a dedicated research re-read. In the case of enlarging lesions, subtractions of coregistered intensity-normalized images were used to confirm that the lesion had unequivocally enlarged. Only the newly segmented area (ie, the area that changed) was included. Cases with progression were first segmented and manually corrected as detailed above and then additionally verified by a third radiologist who is a subspecialist neuroradiologist with >10 years postfellowship experience.



**FIG 2.** Prelesion and control masks. Expert segmentation mask at time  $b$  ( $t_b$ ) is projected backward to time  $a$  ( $t_a$ ) to the location where a lesion will occur (prelesion, red) and the other random areas in the NAWM (control, green). Note that this 2D representation is for illustrative purposes only, and for the experiments, the random translation was in 3D.



**FIG 3.** Radiomics workflow in which features are extracted from the segmented regions and passed to the machine learning models.

New lesion segmentations were projected backward in time to the same location (dubbed the prelesion) on the prior MR imaging study. This same segmentation was then randomly translated 3 times to other locations within the NAWM to produce 3 control segmentations (Fig 2). The NAWM mask was obtained by subtracting the lesion mask from the white matter mask generated by FSL. By means of PyRadiomics (Version 3.01; <https://pypi.org/project/pyradiomics/>),<sup>34</sup> high-order image features were extracted from the prelesion and control segmentations (Fig 3). The 3 control features were averaged to gain a fairer representation of the NAWM. Because the shapes of the control segmentations were dependent on the original lesion segmentations, shape-related radiomic features were excluded.

## Model, Training, and Evaluation

A baseline model using just the mean intensity value of the segmented regions was first tested to ensure that there was justification to proceed to use higher-order features;<sup>28</sup> control and prelesion regions were compared on both internal and external data.

Several machine learning models were trained and tested (eXtreme Gradient Boosting [XGBoost; <https://xgboost.readthedocs.io/en/stable/>], Support Vector Classifier, K Nearest Neighbor, and logistic regression). Hyperparameter optimization was performed using a grid search. Because the classes were balanced 50:50 in a binary prediction problem, we chose overall prediction accuracy as our primary evaluation metric.<sup>35</sup> Feature importance was extracted from the best-performing model. This was used for feature selection and to give a level of explainability. The optimal number of features is a trade-off between model complexity and performance. The number of features that yields the best performance across the different classification models would be chosen as the optimal number of features. Failure analysis was performed on incorrectly classified cases.<sup>27</sup>

To demonstrate the potential clinical usefulness of our approach,<sup>27,28</sup> we produced a probability map using a previously unseen case. A section with a new lesion was selected, and the image was divided into patches. Radiomics features were extracted from each patch and passed to the best-performing model to predict whether that patch was a prelesion. The absolute and relative probabilities given by the model were calculated and used in a data visualization to illustrate a potential future clinical application.

The code necessary to reproduce these experiments is available here (<https://github.com/insight-ucd/insightmri/tree/main/MSOmics>).

## RESULTS

The total number of participants in the training, test, and external validation sets was 167 (124, 23 and 20); 1 additional participant was randomly chosen for production of the probability map only. Failure of FSL registration and brain extraction led to the exclusion of 1 patient (Fig 1). Of the 147 participants in the training/test set, 102 were women and 45 were men. Overall, the average age was 42 years (range, 21–74 years) (Table 1). Two hundred ninety-four volumetric MR imaging FLAIR images were used in the process of training

and internal testing, with 40 heterogeneous FLAIR acquisitions for external validation.

Results of different models for testing and validation are displayed in Tables 2 and 3. The preliminary model based on first-order image metrics (mean pixel intensity) had modest performance on internal data (accuracy, precision, recall, and F1-score of 0.77, 0.76, 0.77, and 0.75). There was a statistically significant difference in the overall mean pixel intensity between the prelesion samples and controls on internal data with a paired *t* test ( $P < .05$ ). However, this method did not generalize to external data (accuracy, precision, recall, and F1-score of 0.5, 0.25, 0.5, and 0.33), justifying the use of higher-order features.

The best-performing radiomics-based model was XGBoost, with accuracy, precision, recall, and F1-score of 0.91, 0.91, 0.91,

and 0.91 on the test set, and 0.74, 0.74, 0.74, and 0.70 on the external validation set. The 5 most important radiomics features (Fig 4) to the XGBoost model were associated with the overall heterogeneity (RunEntropy, Variance, and GrayLevelNonUniformityNormalized) and low gray-level emphasis (LowGrayLevelEmphasis and LongRunLowGrayLevelEmphasis) of the segmented regions. Complete feature importance of the best-performing model is given in the Online Supplemental Data. The Online Supplemental Data show the performance for a model trained on only 14 features (including only 1 feature for each 10 participants<sup>28</sup>), which was inferior to the chosen model.

Probability maps were produced for illustrative purposes of potential future clinical application and are shown in Fig 5A, -B. These show the probability of a new lesion in absolute terms (A) and relative to the overall risk (B). The relative probability is defined in Equations 1 and 2.

$$1) \quad P(x, y) = P(N) \times (NLV/WMV).$$

$P(x, y)$  represents the general probability of a lesion at location coordinates  $x, y$ , where  $P(N)$  represents the probability of a new lesion anywhere. NLV represents the average new lesion volume. WMV represents the NAWM volume.

$$2) \quad RP(\bar{x}, \bar{y}) = P(x, y)/P'(\bar{x}, \bar{y}).$$

The relative probability of a new lesion in a specific location ( $\bar{x}, \bar{y}$ ) is represented by  $RP(\bar{x}, \bar{y})$ , where the predicted probability of a lesion by the model in the specific location  $\bar{x}, \bar{y}$  is  $P'(\bar{x}, \bar{y})$ , and  $P(x, y)$  represents the baseline probability of a lesion in any location.

For our data,  $P(x, y) = 0.5(0.02) = 0.01$

$$\text{and } RP(\bar{x}, \bar{y}) = 0.01/P'(\bar{x}, \bar{y}).$$

Our sample probability map demonstrated 1 true-positive and 1 false-positive using a cutoff of relative increased probability of a new white matter lesion of 4 compared with background (Fig 5).

Failure analysis on external validation showed that in many instances, a potential cause for misclassified cases was misregistration of images rather than the machine learning model. For example, of the 3 false-negatives, 2 were likely related to misregistration (Fig 6).

## DISCUSSION

In this study, we have demonstrated the potential of radiomics features in predicting progression on MR imaging of

**Table 1: Patient demographics**

Demographics	
Total participants	147
Average age (yr)	42.19
Min	21
Max	74
Sex	
Male	45
Female	102

**Note:**—Max indicates maximum; Min, minimum.

**Table 2: Internal test cohort results**

Model	Precision	Recall	F1-Score	Accuracy	Best Parameters
XGBoost	0.91	0.91	0.91	0.91	{'classifier__colsample_bylevel': 0.8, 'classifier__gamma': 0, 'classifier__learning_rate': 0.2, 'classifier__max_depth': 4, 'classifier__min_child_weight': 1, 'classifier__n_estimators': 100, 'classifier__subsample': 0.5}
SVC	0.90	0.89	0.89	0.89	{'classifier__C': 10, 'classifier__kernel': 'rbf'}
Logistic regression	0.81	0.78	0.78	0.78	{'classifier__C': 1, 'classifier__penalty': 'l1', 'classifier__solver': 'liblinear'}
KNN	0.83	0.78	0.78	0.78	{'classifier__n_neighbors': 7}
Intensity baseline	0.76	0.77	0.75	0.77	NA

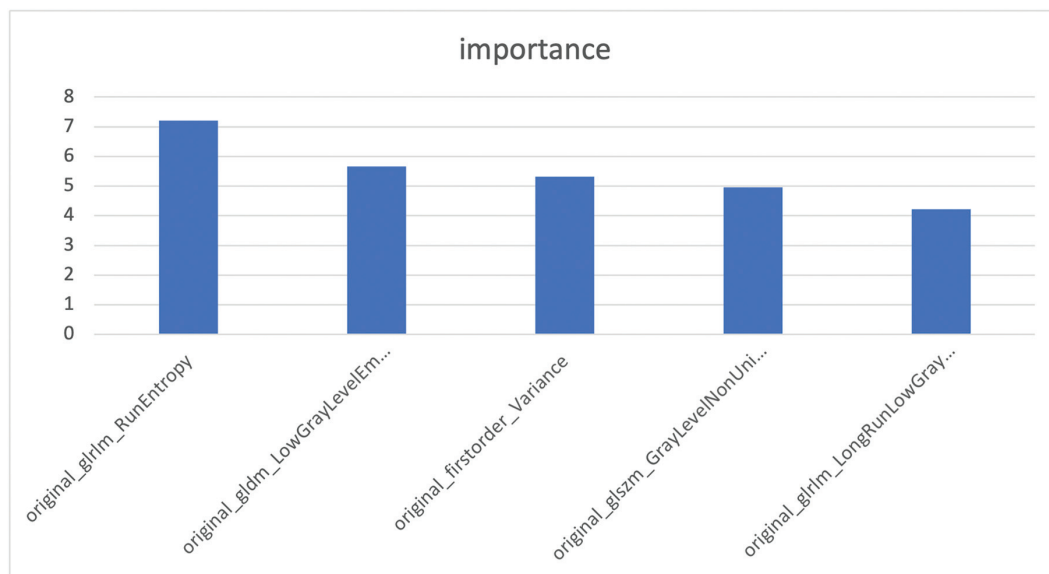
**Note:**—SVC indicates support vector classifier; KNN, K nearest neighbor; NA, not applicable.

**Table 3: External validation cohort results**

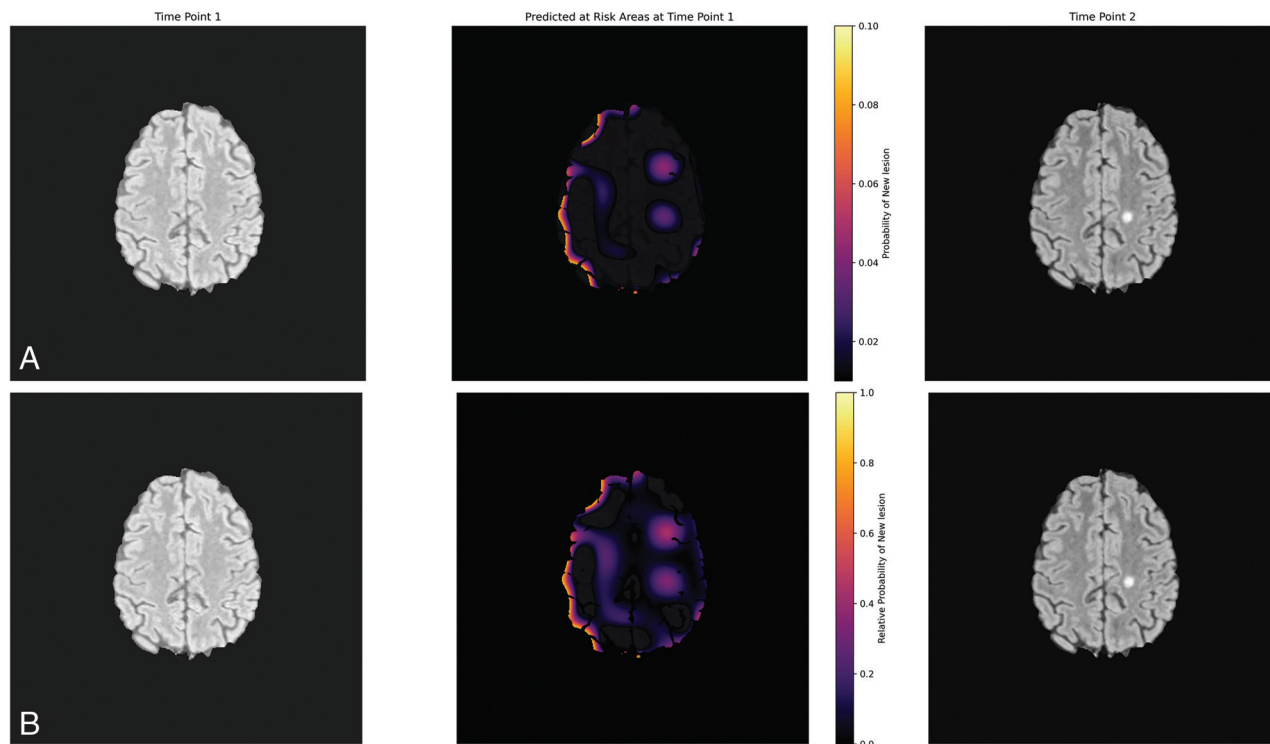
Model	Precision	Recall	F1-Score	Accuracy	Best Parameters
XGBoost	0.74	0.74	0.70	0.74	{'classifier__colsample_bylevel': 0.8, 'classifier__gamma': 0, 'classifier__learning_rate': 0.2, 'classifier__max_depth': 4, 'classifier__min_child_weight': 1, 'classifier__n_estimators': 100, 'classifier__subsample': 0.5}
SVC	0.69	0.71	0.68	0.71	{'classifier__C': 10, 'classifier__kernel': 'rbf'}
Logistic regression	0.62	0.55	0.56	0.55	{'classifier__C': 1, 'classifier__penalty': 'l1', 'classifier__solver': 'liblinear'}
KNN	0.51	0.42	0.43	0.42	{'classifier__n_neighbors': 7}
Intensity baseline	0.25	0.50	0.33	0.50	NA

**Note:**—SVC indicates support vector classifier; KNN, K nearest neighbor; NA, not applicable.





**FIG 4.** Top 5 radiomic features identified by the top performing XGBoost model.

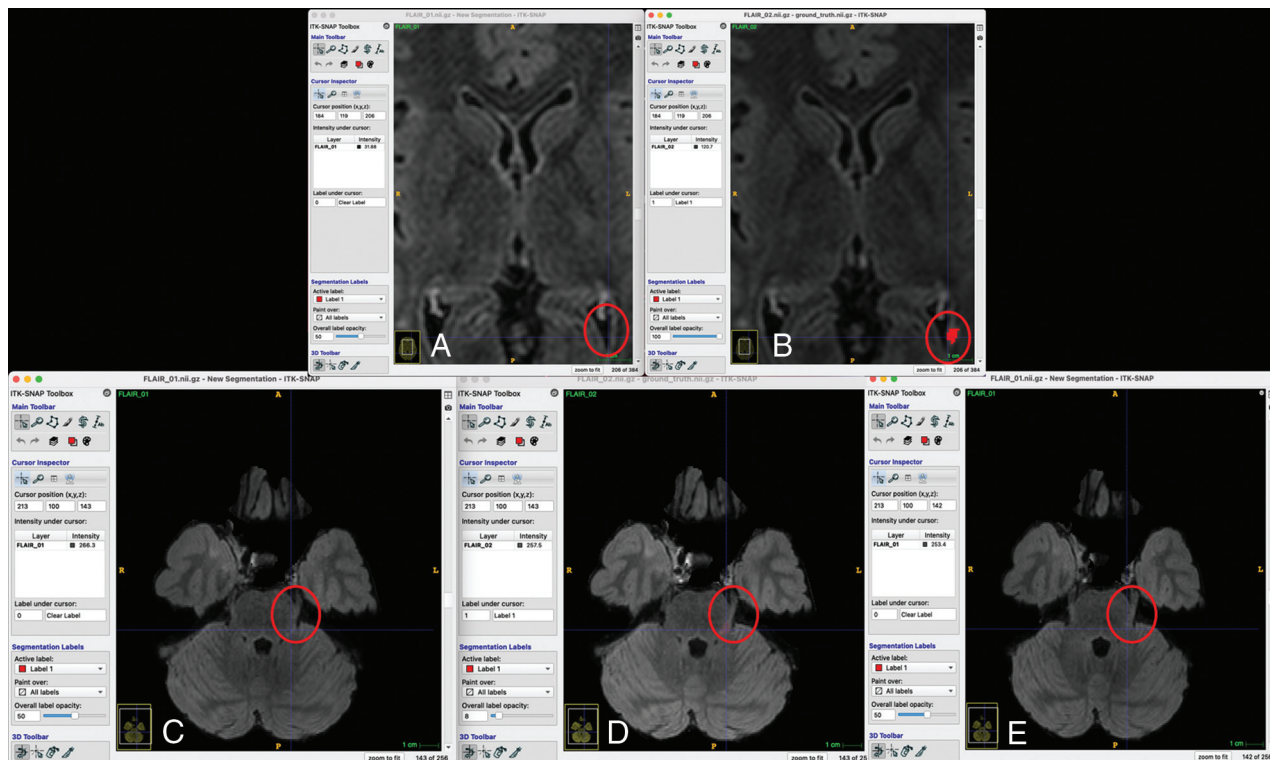


**FIG 5.** Illustrative probability maps showing the absolute (A, *Upper row*) and relative (B, *Lower row*) probability of a new lesion occurring in each patch.

the brain in individuals with MS. We used a machine learning model and the extracted features at the location of future progression to predict the occurrence of a new lesion.

AI is revolutionizing the field of MS imaging research, promising to improve diagnostic accuracy, reduce the time taken for image interpretation, and facilitate personalized patient management.<sup>7</sup> Radiomics, the extraction of high-order features from radiologic images using data-characterization algorithms, plays

an increasingly important role in MS research.<sup>16-19</sup> Building on this body of evidence and understanding that the white matter of individuals with MS is quantifiably different from that of healthy controls,<sup>9,36</sup> our hypothesis was that radiomics features could potentially reveal subtle alterations before they become visually discernible on MR imaging. As in previous studies,<sup>37,38</sup> interpretation of the radiomics features that were most important to our final model showed that more homogeneous and lower gray-level



**FIG 6.** False-negative analysis (A and B, Upper row, C–E, Lower row). Two of the 3 false-negative cases in the external validation set are shown. A and B, The new lesion map falls within the ventricle on the prior image, making a negative prediction more likely because the low gray levels were associated with negative predictions (Fig 4). C, D, and E, The orientation of the proximal left trigeminal nerve is different so that the segmentation is cast onto the normal brainstem instead.

regions were more likely to be classified as “control,” yielding a level of intuition about model behavior and suggesting that the model performed as expected.

Temporality or the consideration of time-based changes is increasingly recognized as a critical component in radiology AI research.<sup>21</sup> It allows longitudinal analysis of imaging data, potentially enabling the AI to capture disease progression and treatment responses and even predict future outcomes.<sup>22,23,39</sup> Integrating temporality into AI models enhances their capacity to detect subtle, time-dependent changes in patient imaging data, which may remain undetected by conventional image analysis.<sup>23</sup> We aimed to build on this research because in conditions such as MS, in which the temporal evolution of lesions is a critical aspect of disease-monitoring and management, they could lead to earlier intervention and better patient outcomes.<sup>40,41</sup> The MSSEG2 was a clear recognition of the importance of temporality in research. We demonstrate generalizability through external validation of our results on MSSEG2. While there was a drop in performance at external validation, it was in keeping with what was expected from the literature.<sup>42</sup> Furthermore, the MSSEG2 data are known to be very heterogeneous, including several different institutions, 15 scanners, and scanning protocols with a mix of 1.5T and 3T, meaning that MSSEG2 was a robust test of performance.<sup>3</sup> Diagnostic accuracy results from MSSEG2 have been modest overall, putting our performance in context. Even the best-performing published models from the challenge demonstrate only modest accuracy for new-lesion detection, indicating both the robust challenge represented

by the data set and the potential for improvement in the field.<sup>3</sup> Furthermore, false-negative analysis also shows that some of the misclassifications were technical (Fig 6).

A relative probability map was produced on an unseen case to demonstrate a possible future clinical application. There are some artifacts in the cortical gray matter on the right of the image and lower-level increased probability in a band of subcortical white matter on the right. These could be removed with anatomic filtering or further thresholding, but we chose to present the map “as is” rather than overly “tuning it.” One false-positive was seen, even with thresholding, but this was at the site of a lesion that regressed, so there was a change at this location, albeit in the opposite direction. Furthermore, the purpose of the map is to illustrate a potential clinical application for future work rather than it being directly related to the research question of this study. Prospective analysis would be needed before the relative probability map could be implemented clinically.

### Limitations

The retrospective study design limits the level of evidence. Furthermore, because our experiments only involved those patients with progression, there was a strong selection bias. While this is a common issue in clinical radiology research,<sup>43</sup> it remains a clear limitation. The purpose of the study, to examine the potential for radiomics markers to predict new lesion occurrence, informed our research design; however, this purpose would need to be addressed prospectively before any implementation. Our sample size was modest, but having >120 pairs with change for

training compares favorably with MSSEG2,<sup>3</sup> which comprised only 100 pairs of patients in total, of whom one-half were stable.

### Future Directions and Implications for Practice

Our findings indicate that radiomics features have the potential to serve as an imaging biomarker for predicting radiologic progression of MS. The extracted features could be used to develop a probability map for future lesion occurrence. This map and the relative probabilities of progression have many potential clinical applications. These include making personalized decisions around the timing for follow-up imaging or even interventions. The use of imaging biomarkers in this way is a topic of prospective evaluation in cancer imaging.<sup>44</sup> Furthermore, stratification of patients based on imaging biomarkers to identify individuals suited for preventive intervention is highlighted as a key role of medical imaging in personalized medicine by the European Society of Radiology.<sup>45</sup> Another possible application would be to highlight ROIs during radiologist interpretation<sup>46</sup> or to direct attention<sup>47</sup> in a lesion-detection algorithm.

### CONCLUSIONS

Our machine learning model based on radiomics features successfully differentiated prelesions from NAWM. This result suggests that radiomics features from NAWM could serve as an imaging biomarker for progression of MS on MR imaging.

**Disclosure forms** provided by the authors are available with the full text and PDF of this article at [www.ajnr.org](http://www.ajnr.org).

### REFERENCES

- McNamara C, Sugrue G, Murray B, et al. **Current and emerging therapies in multiple sclerosis: implications for the radiologist, Part 1: mechanisms, efficacy, and safety.** *AJNR Am J Neuroradiol* 2017;38:1664–71 [CrossRef Medline](#)
- McNamara C, Sugrue G, Murray B, et al. **Current and emerging therapies in multiple sclerosis: implications for the radiologist, Part 2: surveillance for treatment complications and disease progression.** *AJNR Am J Neuroradiol* 2017;38:1672–80 [CrossRef Medline](#)
- Commowick O, Cervenansky F, Cotton F, et al. **MSSEG-2 challenge proceedings: multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure.** In: *International Conference on Medical Image Computing and Computer Assisted Intervention*. September 27 to October 1, 2021; Virtual
- Yang J, Hamade M, Wu Q, et al. **Current and future biomarkers in multiple sclerosis.** *Int J Mol Sci* 2022;23:5877 [CrossRef Medline](#)
- Martin D, Tong E, Kelly B, et al. **Current perspectives of artificial intelligence in pediatric neuroradiology: an overview.** *Front Radiol* 2021;1:713681 [CrossRef Medline](#)
- Kelly BS, Judge C, Bollard SM, et al. **Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE).** *Eur Radiol* 2022;32:8054 [CrossRef Medline](#)
- 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;64:2103–17 [CrossRef Medline](#)
- Lladó X, Ganiler O, Oliver A, et al. **Automated detection of multiple sclerosis lesions in serial brain MRI.** *Neuroradiology* 2012;54:787–807 [CrossRef Medline](#)
- Filippi M, Preziosa P, Arnold DL, et al. **Present and future of the diagnostic work-up of multiple sclerosis: the imaging perspective.** *J Neurol* 2023;270:1286–99 [CrossRef Medline](#)
- Havas J, Leray E, Rollot F, et al. **Predictive medicine in multiple sclerosis: a systematic review.** *Mult Scler Relat Disord* 2020;40:101928 [CrossRef Medline](#)
- Lorefice L, Pitzalis M, Murgia F, et al. **Omics approaches to understanding the efficacy and safety of disease-modifying treatments in multiple sclerosis.** *Front Genet* 2023;14:1076421 [CrossRef Medline](#)
- Brummer T, Muthuraman M, Steffen F, et al. **Improved prediction of early cognitive impairment in multiple sclerosis combining blood and imaging biomarkers.** *Brain Commun* 2022;4:fcac153 [CrossRef Medline](#)
- Storelli L, Azzimonti M, Gueye M, et al. **A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging.** *Invest Radiol* 2022;57:423–32 [CrossRef Medline](#)
- Santos DP, dos Dietzel M, Baessler B. **A decade of radiomics research: are images really data or just patterns in the noise?** *Eur Radiol* 2021;31:1–4 [CrossRef Medline](#)
- Huang EP, O'Connor JP, McShane LM, et al. **Criteria for the translation of radiomics into clinically useful tests.** *Nat Rev Clin Oncol* 2023;20:69–82 [CrossRef Medline](#)
- Liu Y, Dong D, Zhang L, et al. **Radiomics in multiple sclerosis and neuromyelitis optica spectrum disorder.** *Eur Radiol* 2019;29:4670–77 [CrossRef Medline](#)
- Peng Y, Zheng Y, Tan Z, et al. **Prediction of unenhanced lesion evolution in multiple sclerosis using radiomics-based models: a machine learning approach.** *Mult Scler Relat Disord* 2021;53:102989 [CrossRef Medline](#)
- Luo X, Piao S, Li H, et al. **Multi-lesion radiomics model for discrimination of relapsing-remitting multiple sclerosis and neuropsychiatric systemic lupus erythematosus.** *Eur Radiol* 2022;32:5700–10 [CrossRef Medline](#)
- Pontillo G, Tommasin S, Cuocolo R, et al. **A combined radiomics and machine learning approach to overcome the clinicroadiologic paradox in multiple sclerosis.** *AJNR Am J Neuroradiol* 2021;42:1927–33 [CrossRef Medline](#)
- Rajpurkar P, Lungren MP. **The current and future state of AI interpretation of medical images.** *N Engl J Med* 2023;388:1981–90 [CrossRef Medline](#)
- Acosta JN, Falcone GJ, Rajpurkar P. **The need for medical artificial intelligence that incorporates prior images.** *Radiology* 2022;304:283–88 [CrossRef Medline](#)
- Sushentsev N, Rundo L, Abrego L, et al. **Time series radiomics for the prediction of prostate cancer progression in patients on active surveillance.** *Eur Radiol* 2023;33:3792–800 [CrossRef Medline](#)
- Kelly BS, Mathur P, Plesniar J, et al. **Using deep learning-derived image features in radiologic time series to make personalised predictions: proof of concept in colonic transit data.** *Eur Radiol* 2023;33:8376–86 [CrossRef Medline](#)
- Moskowitz CS, Welch ML, Jacobs MA, et al. **Radiomic analysis: study design, statistical analysis, and other bias mitigation strategies.** *Radiology* 2022;304:265–73 [CrossRef Medline](#)
- Kocak B, Baessler B, Bakas S, et al. **CheckList for Evaluation of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII.** *Insights Imaging* 2023;14:75 [CrossRef Medline](#)
- Kelly B, Kirwan A, Quinn M, et al. **The ethical matrix as a method for involving people living with disease and the wider public (PPI) in near-term artificial intelligence research.** *Radiography (Lond)* 2023;29(Suppl 1):S103–11 [CrossRef Medline](#)
- Mongan J, Moy L, Kahn CE. **Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers.** *Radiol Artif Intell* 2020;2:e200029 [CrossRef Medline](#)
- Kocak B, Chepelev LL, Chu LC, et al. **Assessment of Radiomics Research (ARISE): a brief guide for authors, reviewers, and readers from the Scientific Editorial Board of European Radiology.** *Eur Radiol* 2023;33:7556–60 [CrossRef Medline](#)
- Pham N, Hill V, Rauschecker A, et al. **Critical appraisal of artificial intelligence-enabled imaging tools using the levels of evidence system.** *AJNR Am J Neuroradiol* 2023;44:E21–28 [CrossRef Medline](#)



30. Jenkinson M, Beckmann CF, Behrens TEJ, et al. **FSL**. *Neuroimage* 2012;62:782–90 [CrossRef Medline](#)
31. Kamnitsas K, Ferrante E, Parisot S, et al. **Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries**. In: *Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction with MICCAI 2016*. Athens, Greece. October 17, 2016
32. Yushkevich PA, Gao Y, Gerig G, et al. **ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images**. *Annu Int Conf IEEE Eng Med Biol Soc* 2016;2016:3342–45 [CrossRef Medline](#)
33. Rovira À, Wattjes MP, Tintoré M, et al; MAGNIMS study group. **MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis: clinical implementation in the diagnostic process**. *Nat Rev Neurol* 2015;11:471–82 [CrossRef Medline](#)
34. Griethuysen JJM, van Fedorov A, Parmar C, et al. **Computational radiomics system to decode the radiographic phenotype**. *Cancer Res* 2017;77:e104–07 [CrossRef Medline](#)
35. Maier-Hein L, Reinke A, Christodoulou E, et al. **Metrics reloaded: pitfalls and recommendations for image analysis validation**. *ArXiv* 2023. <https://arxiv.org/abs/2206.01653v1>. Accessed November 1, 2023
36. Filippi M, Preziosa P, Rocca MA. **Brain mapping in multiple sclerosis: lessons learned about the human brain**. *Neuroimage* 2019;190:32–45 [CrossRef Medline](#)
37. Wang L, Kelly B, Lee EH, et al. **Multi-classifier-based identification of COVID-19 from chest computed tomography using generalizable and interpretable radiomics features**. *Eur J Radiol* 2021;136:109552 [CrossRef Medline](#)
38. Ardakani AA, Bureau NJ, Ciaccio EJ, et al. **Interpretation of radiomics features: a pictorial review**. *Comput Methods Programs Biomed* 2022;215:106609 [CrossRef Medline](#)
39. Kelly B, Martinez M, Do H, et al. **DEEP MOVEMENT: deep learning of movie files for management of endovascular thrombectomy**. *Eur Radiol* 2023;33:5728–39 [CrossRef Medline](#)
40. Brex PA, Ciccarelli O, O'Riordan JI, et al. **A longitudinal study of abnormalities on MRI and disability from multiple sclerosis**. *N Engl J Med* 2002;346:158–64 [CrossRef Medline](#)
41. Todea AR, Melie-Garcia L, Barakovic M, et al; Swiss MS Cohort Study. **A multicenter longitudinal MRI study assessing LeMan-PV Software accuracy in the detection of white matter lesions in multiple sclerosis patients**. *J Magn Reson Imaging* 2023;58:864–76 [CrossRef Medline](#)
42. Yu AC, Mohajer B, Eng J. **External validation of deep learning algorithms for radiologic diagnosis: a systematic review**. *Radiol Artif Intell* 2022;4:e210064 [CrossRef Medline](#)
43. Sica GT. **Bias in research studies**. *Radiology* 2006;238:780–89 [CrossRef Medline](#)
44. Martí-Bonmatí L, Alberich-Bayarri Á, Ladenstein R, et al. **PRIMAGE project: predictive in silico multiscale analytics to support childhood cancer personalised evaluation empowered by imaging biomarkers**. *Eur Radiol Exp* 2020;4:22 [CrossRef Medline](#)
45. European Society of Radiology (ESR). **Medical imaging in personalised medicine: a white paper of the research committee of the European Society of Radiology (ESR)**. *Insights Imaging* 2015;6:141–55 [CrossRef Medline](#)
46. Yap MH, Goyal M, Osman F, et al. **Breast ultrasound region of interest detection and lesion localisation**. *Artif Intell Med* 2020;107:101880 [CrossRef Medline](#)
47. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. **Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions**. *Diagn Interv Radiol* 2023 October 3. [Epub ahead of print] [CrossRef Medline](#)