© Pol J Radiol 2026; 91: e246-e247
DOI: https://doi.org/10.5114/pjr/219935

Received: 18.02.2026
Accepted: 24.03.2026
Published: 25.05.2026

Polish Journal of Radiology

http://www.polradiol.com

Letter to the Editor

Response to "On patient-level splitting, contrast-free claims and unsupported comparators in *Machine learning-based classification of multiple sclerosis lesion activity using multi-sequence MRI radiomics*"

Mohammadreza Elhaie[A,C,D,E,F], Daryoush Shahbazi-Gahrouei[A,C,D,E,F,G]

Department of Medical Physics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

Dear Editor,

We appreciate the thoughtful comments from Dr Galassi in the Letter to the Editor [1] on our article [2] and we welcome the opportunity to clarify several methodological aspects. These comments raise important considerations in radiomics research, which we address below.

Regarding the allocation of patients and lesions to training and test sets, we employed a stratified lesion-level split to preserve class balance (39% active lesions), resulting in 140 lesions for training (74.8%), 19 for validation (10.2%), and 28 for testing (15.0%). While this approach ensured proportional representation of active and non-active lesions, we acknowledge that lesions from the same patient (mean 6.0 ± 3.2 per patient) may have been distributed across splits, potentially introducing shared biological or acquisition-related correlations that could optimistically bias performance estimates [3]. To mitigate this, feature selection via recursive feature elimination with cross-validation (RFECV) was performed within the training set's cross-validation loop, and the test set was held out entirely. However, we concur that patient-level splitting or grouped cross-validation would further enhance robustness against intra-patient dependencies. In supplementary analyses (not reported in the original manuscript), a post-hoc patient-level split (training on 23 patients/140 lesions, testing on 8 patients/47 lesions) yielded a comparable area under the receiver operating characteristic curve (AUC-ROC) of 0.85 [95% confidence interval (CI): 0.79-0.91], suggesting limited inflation in our primary results. Future external validation will incorporate patient-level strategies to confirm generalisability.

The discrepancy in study dates between the abstract (November 2024 – February 2025) and methods (November 2023 – February 2024) was an inadvertent typographical error in the abstract; the correct period is November 2023 – February 2024, as detailed in the methods. Additionally, we note an error in the description of the magnetic resonance imaging (MRI) scanner: it was incorrectly listed as a Philips Ingenia 1.5 T MRI system (Netherlands), whereas the correct system used was a 1.5 T MRI system (Siemens Magnetom, Germany). All scans were acquired on this single system with consistent protocols (Table 1), minimising variability from scanner or software changes.

On the contrast-free claim, we confirm that radiomic features were extracted exclusively from pre-contrast T1-weighted (T1W) images, alongside T2-weighted (T2W), fluid-attenuated inversion recovery (FLAIR), diffusion-weighted imaging, and susceptibility-weighted imaging sequences. Post-contrast T1W images were used solely for lesion labelling as the gold standard, based on gadolinium enhancement and/or T2/FLAIR changes compared to prior scans. This design supports our emphasis on a contrast-free classification pipeline for prospective use, though we agree that explicit phrasing in the methods reinforces this distinction.

The statement in the abstract and conclusions regarding performance "comparable to radiologist performance" was intended to contextualise our AUC-ROC of 0.87 (95% CI: 0.82-0.92) against reported radiologist accuracies

Correspondence address:
Mohammadreza Elhaie, Department of Medical Physics, School of Medicine, Isfahan University of Medical Sciences, Isfahan 81746-73461, Iran,
e-mail: mrelhaie@gmail.com

Authors' contribution:
A Study design · **B** Data collection · **C** Statistical analysis · **D** Data interpretation · **E** Manuscript preparation · **F** Literature search · **G** Funds collection

in the literature (typically 0.80-0.90 for visual assessment of lesion activity without advanced tools [4]), rather than to imply a direct head-to-head comparison in our cohort. No human-reader benchmark was conducted in this study, and we appreciate the suggestion to remove or qualify this phrasing to avoid unsupported implications. Our model's sensitivity (0.85) and specificity (0.83) were derived at a default probability threshold of 0.5, optimised on the validation set and fixed for the test set. For additional transparency, we calculated the precision-recall AUC post hoc as 0.84 (95% CI: 0.78-0.90), with positive predictive value (PPV) of 0.79 and negative predictive value (NPV) of 0.88 at a 39% active lesion prevalence (matching our dataset). At a lower prevalence of 20%, as may occur in clinically stable multiple sclerosis (MS) cohorts, PPV would be 0.62 and NPV 0.93, informing deployment considerations [5].

Finally, preprocessing transformations, including Nyul's intensity normalisation and z-score standardisation, were fitted exclusively on the training data and applied unchanged to validation and test sets, preventing information leakage. A sensitivity analysis adding Gaussian noise ($\sigma = 0.1$) to the test set maintained an AUC-ROC of 0.86 (95% CI: 0.81-0.91), confirming robustness [5,6].

We thank Dr Galassi for these valuable insights, which strengthen the interpretation of our work and guide future refinements.

Disclosures

1. Institutional review board statement: Not applicable.
2. Assistance with the article: None.
3. Financial support and sponsorship: None.
4. Conflicts of interest: None.

References

1. Galassi SG. On patient-level splitting, contrast-free claims and unsupported comparators in "Machine learning-based classification of multiple sclerosis lesion activity using multi-sequence MRI radiomics". Pol J Radiol 2026; 91: e56-e57. DOI: https://doi.org/10.5114/pjr/213873.
2. Elhaie M, Etemadifar M, Adariani AR, Khorasani A, Shahbazi-Gahrouei D. Machine learning-based classification of multiple sclerosis lesion activity using multi-sequence MRI radiomics: a complete analysis of T1, T2, FLAIR, DWI, and SWI features. Pol J Radiol 2025; 90: e394-e403. DOI: 10.5114/pjr/206986.
3. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 2017; 40: 913-929.
4. Filippi M, Rocca MA, Ciccarelli O, De Stefano N, Evangelou N, Kap-pos L, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. Lancet Neurol 2016; 15: 292-303.
5. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS One 2015; 10: e0118432. DOI: 10.1371/journal.pone.0118432.
6. Collins GS, Moons KG, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024; 385: e078378. DOI: 10.1136/bmj-2023-078378.