Letourneau-Guillon *European Radiology* (2026) 36:2531–2533
https://doi.org/10.1007/s00330-025-12317-2

EUROPEAN SOCIETY OF RADIOLOGY
European Radiology

COMMENTARY

Unlocking the full potential of MRI with artificial intelligence for pre-operative meningioma evaluation

Laurent Letourneau-Guillon[1,2*]

Meningiomas are the most common primary intracranial neoplasms and a common finding in day-to-day neuroradiological practice. Their reported prevalence is on the rise due to an aging population and increasing use of neuroimaging [1]. Despite some limitations, the current World Health Organization grading system strongly influences post-operative management [2]. While most meningiomas are benign (grade 1), grade 2 or 3 meningiomas exhibit more aggressive behavior, a higher risk of recurrence, and a poorer prognosis [3]. Another important prognostic marker is Ki-67, a marker of cellular proliferation [4]. However, a limitation of tumor grading and Ki-67 is that they are only available after surgery. Pre-operative imaging-based prediction of meningioma grading and proliferative index could inform prognosis and personalize management at earlier stages of the care pathway.

In their article, Cheng et al aim to advance the field by developing and testing models to predict meningioma grade (1 vs. 2–3) and proliferative index (Ki-67 ≥ 5%) before tissue is obtained for pathological tumor analysis [5]. This model development and evaluation study leveraged retrospective data from two Chinese institutions, including 638 patients used for training/validation and 314 patients from a different institution for external testing. Their work compared a clinical model combining demographic, radiological signs, and hematological parameters with two image-based approaches: radiomics and deep learning (DL) models, both using T1-weighted, T2-weighted, and contrast-enhanced T1-weighted sequences. To prepare the input images and ensure that image-based models focus on relevant meningioma regions, the authors developed and evaluated brain and tumor segmentation models using the widely adopted nnU-Net framework [6]. This enabled them to select the image slices containing the largest tumor area. Radiomics then involves extracting numerical image features for subsequent machine learning classification. To this end, the authors used support vector machines; other approaches, like gradient boosting, might have improved performance. Conversely, DL models learn patterns directly from image data. For this second strategy, the authors evaluated 2D and 2.5D inputs and found that the latter approach, i.e. using a stack of slices around the maximum tumor area, improved performance while remaining computationally efficient. They also compared convolutional neural networks with vision transformer backbones while integrating attention mechanisms to enhance feature representation. In the end, the best-performing DL model achieved areas under the receiver operating curve (AUC) of 0.90 and 0.89 for grade and Ki-67 prediction, respectively, on external testing, significantly outperforming clinical and radiomics-based models.

Overall, the results are promising, and future work could further improve the performance of similar prediction models. This includes expanding towards a multimodal model incorporating baseline clinical data, other imaging modalities, such as additional MRI sequences or positron emission tomography, and even emerging blood biomarkers [1]. Additionally, DL model refinements could include using 3D classifiers to capture the entire tumor volume, motivated by the varying spatial distribution of

This comment refers to the article available at https://doi.org/10.1007/s00330-025-11958-7.

*Correspondence:
Laurent Letourneau-Guillon
laurent.letourneau-guillon.1@umontreal.ca
[1]Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Université de Montréal, Montréal, QC, Canada
[2]Département de radiologie, Centre Hospitalier de l'Université de Montréal (CHUM); Département de radiologie, radio-oncologie et médecine nucléaire, faculté de médecine, Université de Montréal, Montréal, QC, Canada

© The Author(s), under exclusive licence to European Society of Radiology 2026

Springer

molecular mutations across morphologically distinct intra-tumoral regions [2]. Moreover, incorporating peri-tumoral regions could help identify aggressive features, such as invasion at the brain-to-tumor interface [7]. Finally, using a pre-trained foundation model encoder could boost performance and improve generalizability [8].

Beyond providing a model with high discrimination, as assessed by its AUC, a well-calibrated model provides reliable and trustworthy probability estimates, which are essential for safe clinical decision-making [9]. In this study, calibration was evaluated using two approaches: the Hosmer–Lemeshow test and calibration plots. The former was found non-significant, however, the Hosmer–Lemeshow test has recognized limitations that may account for this finding [10]. While no single calibration metric is perfect, using additional metrics such as the Brier score can provide a more comprehensive assessment of calibration. Furthermore, calibration can be assessed visually using calibration plots, which illustrate the agreement between predicted and observed event rates across the range of predicted probabilities. In this study, local miscalibrations were revealed by such plots, and the authors appropriately acknowledged that additional work is needed to obtain more reliable probability estimates.

Among the study's limitations, reliance on a retrospective dataset composed of a highly selected surgical cohort introduces the possibility of selection and spectrum biases that may limit external validity. Another limitation is the class imbalance regarding the outcomes to be predicted, a prevalent issue in medicine. To mitigate this issue, the authors employed common sampling strategies during training. Nevertheless, differences in class prevalence between the training and external test cohorts persist, and similar shifts are likely to occur if additional sites adopt this model in the future. This can affect model calibration and probability-based decision rules unless mitigation strategies, such as recalibration, are used.

While the authors made a notable contribution, several steps remain to establish clinical utility and bridge the translational gap in this specific domain. Replication in external, multicenter datasets and across various MRI scanners and acquisition protocols is required to confirm generalizability and explore potential biases, including demographic or scanner-related factors. The authors addressed sex imbalance, an important consideration given the higher prevalence of meningiomas in females, but evaluation across diverse patient subgroups remains essential. For instance, stratified analysis according to age group, meningioma localization, contrast agent type and dosing could identify other patient or institutional factors that may limit generalizability. As the study was conducted in an Asian population, caution is warranted when extrapolating findings to other demographic groups. This is particularly concerning for Black patients, who have higher meningioma incidence [1, 11], but may be underrepresented in training data, potentially leading to performance degradation and exacerbating healthcare inequities. Prospective evaluation and continued monitoring are also of paramount importance, given the expected changes in data distribution that are likely to worsen model performance over time. Finally, deploying models in clinical settings and evaluating the human-AI interaction will ultimately determine whether such models enhance patient management and outcomes.

In summary, Cheng et al provide further evidence that DL can outperform radiomics and clinical models in predicting histological features associated with worse prognosis in meningiomas. Their study exemplifies the shift from descriptive imaging toward automated predictive tools. The next challenge lies not in developing yet other promising models, but in bridging the translational gap; only then can the field move from bytes to bedside.

Acknowledgements
Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CHUM), Département de radiologie, CHUM and Département de Radiologie, Radio-Oncologie et Médecine Nucléaire, Faculté de Médecine, Université de Montréal.

Funding
Laurent Letourneau-Guillon is supported by a Fonds de Recherche Québec Santé (FRQ-S)/Fondation de L'Association des Radiologistes du Québec (FARQ) Junior 1 salary award (https://doi.org/10.69777/299979).

Compliance with ethical standards

Guarantor
The scientific guarantor of this publication is Laurent Letourneau-Guillon.

Conflict of interest
The author of this manuscript declares no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry
No complex statistical methods were necessary for this paper.

Informed consent
Written informed consent was not required for this study because it is an invited commentary.

Ethical approval
Institutional Review Board approval was not required because it is an invited commentary.

Study subjects or cohorts overlap
Not applicable

Methodology

- Commentary

Received: 9 September 2025 Revised: 10 December 2025 Accepted: 29 December 2025
Published online: 9 February 2026

References

1. Wang JZ, Landry AP, Raleigh DR et al (2024) Meningioma: International Consortium on Meningiomas consensus review on scientific advances and treatment paradigms for clinicians, researchers, and patients. Neuro-Oncol 26:1742–1780. https://doi.org/10.1093/neuonc/noae082
2. Goldbrunner R, Stavrinou P, Jenkinson MD et al (2021) EANO guideline on the diagnosis and management of meningiomas. Neuro-Oncol 23:1821–1834. https://doi.org/10.1093/neuonc/noab150
3. Soni N, Ora M, Bathla G et al (2025) Meningioma: Molecular Updates from the 2021 World Health Organization Classification of CNS Tumors and Imaging Correlates. AJNR Am J Neuroradiol 46:240–250. https://doi.org/10.3174/ajnr.A8368
4. Yarabarla V, Mylarapu A, Han TJ et al (2023) Intracranial meningiomas: an update of the 2021 World Health Organization classifications and review of management with a focus on radiation therapy. Front Oncol 13:1137849. https://doi.org/10.3389/fonc.2023.1137849
5. Cheng X, Li H, Li C et al (2025) Attention-based deep learning network for predicting World Health Organization meningioma grade and Ki-67 expression based on magnetic resonance imaging. Eur Radiol. Published online 20 August 2025. https://doi.org/10.1007/s00330-025-11958-7
6. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18:203–211. https://doi.org/10.1038/s41592-020-01008-z
7. Joo L, Park JE, Park SY et al (2021) Extensive peritumoral edema and brain-to-tumor interface MRI features enable prediction of brain invasion in meningioma: development and validation. Neuro-Oncol 23:324–333. https://doi.org/10.1093/neuonc/noaa190
8. Chen M, Zhang M, Yin L et al (2024) Medical image foundation models in assisting diagnosis of brain tumors: a pilot study. Eur Radiol 34:6667–6679. https://doi.org/10.1007/s00330-024-10728-1
9. On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative, Van Calster B, McLernon DJ et al (2019) Calibration: the Achilles heel of predictive analytics. BMC Med 17:230. https://doi.org/10.1186/s12916-019-1466-7
10. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L (2020) A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc 27:621–633. https://doi.org/10.1093/jamia/ocz228
11. Jackson HN, Hadley CC, Khan AB et al (2022) Racial and socioeconomic disparities in patients with meningioma: a retrospective cohort study. Neurosurgery 90:114–123. https://doi.org/10.1227/NEU.0000000000001751

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.