Journal of Clinical Neuroscience 138 (2025) 111360

Contents lists available at ScienceDirect

Journal of Clinical Neuroscience

journal homepage: www.journals.elsevier.com/journal-of-clinical-neuroscience

Correspondence

AI for disease prediction: Performance insights and key limitations

Dear Editor,

We have read with great interest the article by Zhang et al. titled "Development of Hybrid radiomic Machine learning models for preoperative prediction of meningioma grade on multiparametric MRI" published in a recent issue of the Journal of Clinical Neuroscience [1]. The authors are to be commended for their work on developing and comparing machine learning models for distinguishing low and high-grade meningiomas using multiparametric MRI, which provides valuable insights into the potential of radiomics and deep learning in this domain. This study makes a significant contribution to the field of neuro-oncology and medical imaging. While this study is insightful, we wish to discuss certain aspects that warrant further consideration, particularly concerning the study's sample size, the clinical applicability of the reported predictive performance, and the lack of external validation.

A primary concern is the study's reliance on a relatively small and imbalanced dataset. The analysis included 97 low-grade and only 18 high-grade meningiomas. This small sample size for high-grade tumors significantly impacts the generalizability and stability of the developed models, a well-recognized challenge in machine learning applications [2]. While the authors acknowledge this limitation and employed oversampling techniques, the robustness of models trained on such a limited number of high-grade cases remains questionable. This is particularly critical as the accurate identification of high-grade meningiomas is paramount for appropriate treatment planning [3].

Furthermore, while the reported Receiver Operating Characteristic Area Under the Curve (ROC AUC) values for the Handcrafted Radiomics Only (HRO) and Handcrafted with Deep Learning Radiomics (HDLR) models (0.825 and 0.794, respectively) are promising, the reported sensitivity and Positive Predictive Value (PPV) raise concerns about the models' clinical utility. The sensitivities (0.499 for HRO and 0.509 for HDLR) and the wide confidence intervals for PPV (0.529 [0.238,0.924] for HRO and 0.465 [0.263,0.846] for HDLR) suggest that a substantial proportion of high-grade meningiomas might be misclassified as low-grade, and the reliability of a positive prediction for a high-grade tumor is highly variable. In a clinical setting, high sensitivity is crucial to avoid under-treatment of aggressive tumors, and a more consistent PPV is needed for confident decision-making [4].

The study also acknowledges the lack of external validation, stating that "Larger, multi-centre studies are warranted to confirm our findings." This is a crucial point, as single-center retrospective analyses, while valuable for initial exploration, often suffer from limited generalizability due to institution-specific patient populations, imaging protocols, and scanner characteristics [5]. Without external validation, the true performance of these models in diverse clinical settings remains unknown. Additionally, the manual segmentation of tumors introduces potential operator-dependent variability, which could affect feature extraction and, consequently, model performance [6]. Future studies might consider exploring automated segmentation methods or assessing inter-rater reliability for manual segmentations to enhance reproducibility [7].

In summary, while the study by Zhang et al. is a valuable contribution to the preoperative grading of meningiomas using advanced imaging analytics, the limitations regarding small sample size, particularly for high-grade tumors, the modest sensitivity and PPV, and the absence of external validation temper the immediate clinical applicability of the proposed models. We believe that discussion of these points will be beneficial for the ongoing research in this area. We look forward to further research that builds upon these important findings, ideally through larger, multi-center prospective studies with robust external validation to develop more reliable and clinically translatable predictive tools.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge the support and guidance provided by our respective institutions, including Xinjiang University, Shenzhen Bay Laboratory, Hong Kong Polytechnic University.

References

[1] Zhang S, Richter J, Veale J, Phan VMH, Candy N, Poonnoose S, et al. Development of Hybrid radiomic Machine learning models for preoperative prediction of meningioma grade on multiparametric MRI. *J Clin Neurosci* 2025;135:111118.
[2] Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One* 2019;14(11):e0224365.
[3] Goldbrunner R, Stavrinou P, Jenkinson MD, Sahm F, Mawrin C, Weber DC, et al. EANO guideline on the diagnosis and management of meningiomas. *Neuro Oncol* 2021;23(11):1821–34.
[4] Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Front Public Health* 2017;5:307.
[5] Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130(6):515–24.
[6] Pavic M, Bogowicz M, Würms X, Glatz S, Finazzi T, Riesterer O, et al. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol* 2018;57(8):1070–4.
[7] Krüger J, Opfer R, Gessert N, Ostwaldt AC, Manogaran P, Kitzler HH, et al. Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks. *NeuroImage: Clinical* 2020;28:102445.

https://doi.org/10.1016/j.jocn.2025.111360
Received 27 May 2025; Accepted 29 May 2025
Available online 3 June 2025

Yiquan Wang[a,b,*], Minnuo Cai[c], Tin-Yeh Huang[d]
[a] College of Mathematics and System Science, Xinjiang University, Urumqi, Xinjiang 830046, China
[b] Institute of Neurological and Psychiatric Disorders, Shenzhen Bay Laboratory, Shenzhen, Guangdong 518132, China
[c] Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Urumqi, Xinjiang 830049, China
[d] Department of Industrial and Systems Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong SAR 999077, China

[*] Corresponding author at: College of Mathematics and System Science, Xinjiang University, Urumqi, Xinjiang 830046, China.
E-mail address: ethan@stu.xju.edu.cn (Y. Wang).