



Letter to the Editor: Comparing commercial artificial intelligence for intracranial hemorrhage: Why operating points matter for procurement

Alexander Reyes Lobo

Universidad Industrial de Santander School of
Medicine, Bucaramanga, Colombia

Dear Editor,

We read with great interest the study by Kim et al.¹ comparing four commercially available artificial intelligence (AI) solutions for acute intracranial hemorrhage detection on non-contrast computed tomography (CT). Their head-to-head, real-world evaluation is particularly relevant because hospitals increasingly need comparative evidence to guide procurement and work-flow integration.

The authors appropriately evaluated each solution as deployed using the product-reported binary classification to calculate sensitivity, specificity, precision, and F1 score. This answers an important practical question: How does each product perform in its current configuration? A complementary question becomes important, however, when these results are used for procurement: How much of the observed difference reflects discriminatory capability, and how much reflects the product-specific operating point?

This distinction matters because changing the decision threshold of the same AI model can materially alter the sensitivity–specificity trade-off. Developers can select different operating points to favor either sensitivity or specificity.² In the present study, solutions A–C provided continuous probability scores, whereas solution D provided only binary classifications.¹ Accordingly, higher specificity, precision, or F1 score at native binary outputs should be interpreted as performance of the deployed product configuration rather than as threshold-independent superiority of the underlying algorithm.

Both perspectives are clinically useful. Native-output performance estimates what an institution may experience after deployment. For products providing continuous scores, a second, threshold-standardized analysis could compare specificity and false-positive burden at a pre-specified matched sensitivity, or sensitivity at matched specificity. Operating-point matching has been used in large diagnostic AI comparisons to improve clinical interpretability.³ Reporting false-positive notifications per 1,000 CT examinations at a clinically acceptable sensitivity could further translate statistical differences into expected work-flow burden.

Because solution D did not provide a continuous score, threshold-standardized comparison could not be performed for that product from the reported outputs. Whether its operating point can be locally configured is therefore a separate product characteristic that may also matter for procurement.

Kim et al.¹ provide a valuable framework for evidence-based AI selection. Future head-to-head evaluations could strengthen this framework by reporting both native-product and threshold-standardized performance, thereby separating two clinically distinct questions: How does the product perform as deployed, and how does it compare when institutions require the same diagnostic safety target?

Corresponding author: Alexander Reyes Lobo

E-mail: reylobo1976@gmail.com

Received 05 September 2026; accepted 07 September 2026.



Epub: 14.10.2026

DOI: 10.4274/dir.2026.264450

Acknowledgments

Generative AI (ChatGPT, OpenAI) was used under the author's supervision for literature organization, brainstorming, and language refinement. The author independently verified the scientific claims, interpretations, and references and takes full responsibility for the final content.

Conflict of interest disclosure

The author declared no conflicts of interest.

References

1. Kim J, Lee HY, Eun J, Im S, Min EJ, Oh SW. Real-world multimetric comparison of four commercial artificial intelligence solutions for intracranial hemorrhage detection. *Diagn Interv Radiol*. 2026. [\[Crossref\]](#)
2. Agarwal S, Wood D, Grzeda M, et al. Systematic review of artificial intelligence for abnormality detection in high-volume neuroimaging and subgroup meta-analysis for intracranial hemorrhage detection. *Clin Neuroradiol*. 2023;33(4):943-956. [\[Crossref\]](#)
3. Saha A, Bosma JS, Twilt JJ, et al. Artificial intelligence and radiologists in prostate cancer detection on MRI (PI-CAI): an international, paired, non-inferiority, confirmatory study. *Lancet Oncol*. 2024;25(7):879-887. [\[Crossref\]](#)