Evaluating the Clinical Utility of Artificial Intelligence in Brain Tumour Diagnosis

A critical literature review of diagnostic performance, implementation barriers and human–AI collaboration

Gurung Amman

24076420S

Co- Author: Dr Xiangyan Fiona Chen

BSc (Hons) Scheme in Applied Sciences · School of Professional Education and Executive Development · The Hong Kong Polytechnic University

Artificial intelligence has already arrived in medicine

This is no longer a forecast. Regulators are clearing AI tools at pace — and imaging is where they have landed hardest.

1,451

AI-enabled medical devices authorised by the FDA since 1995

76%

of every one of those authorisations sits in radiology

295

authorised in 2025 alone — up from a handful a decade earlier

Brain tumour imaging sits squarely inside that 76%.

Source: U.S. FDA AI-Enabled Medical Device List (cumulative authorisations through December 2025), as reported by The Imaging Wire (2026).

Brain tumours are a genuinely hard diagnostic problem

300,000+

new brain and central nervous system cancers diagnosed worldwide every year, with especially high mortality in aggressive subtypes such as high-grade gliomas.

Filho et al. (2025)

Detection is only half the difficulty

Heterogeneous imaging appearances, subtle radiological features and symptoms that overlap with other neurological conditions push cases toward delay or error.

Experienced readers disagree with each other

Agreement between board-certified radiologists on glioma MRI was poor for diffusion characteristics, haemorrhage and eloquent-region involvement — even under standardised reporting criteria (Mohebbi et al., 2025).

Non-neoplastic conditions imitate tumours

Tuberculomas, abscesses and demyelinating lesions share radiological features with malignancy, sending clinicians down pathways that delay correct treatment (Godkhindi et al., 2022).

That variability is not academic — it feeds straight into treatment planning, prognosis and patient outcomes.

Filho et al. (2025); Mohebbi et al. (2025); Godkhindi et al. (2022).

Aim, objectives and method

AIM Critically evaluate the clinical utility of AI in enhancing diagnostic accuracy and efficiency for brain tumour patients.

01 Benchmark against current practice

Compare AI-enhanced diagnosis with conventional radiological assessment on sensitivity, specificity and clinical accuracy.

02 Map support across the pathway

Examine how AI supports radiologist expertise from detection through classification and grading.

03 Test it against reality

Analyse the benefits and the practical barriers clinicians meet when implementing AI in real-world settings.

04 Turn evidence into action

Develop evidence-based recommendations for integrating AI diagnostic tools into routine clinical practice.

METHOD PubMed, PolyU Library and targeted Google Scholar searches, 2019–2026. Studies prioritised where they reported diagnostic performance metrics, clinical implementation, workflow integration or human–AI collaboration. Reference lists of key articles and recent systematic reviews hand-searched.

From hand-crafted features to models that explain themselves

Traditional machine learning

Support vector machines, random forests and k-nearest neighbours. Powerful in principle, but every diagnostic feature had to be hand-engineered by a domain expert first — a bottleneck that capped how complex a task these methods could handle.

Avanzo et al. (2024)

Deep learning and CNNs

Convolutional neural networks learn hierarchical features directly from raw imaging data. Removing the manual feature-engineering step is what unlocked performance on genuinely complex diagnostic problems.

Sistaninejhad et al. (2023)

Explainable AI (XAI)

Tools and methods that make a model's decision-making transparent and human-readable — so a clinician can see which image regions drove the output, verify the reasoning, and fact-check it rather than trust it blindly.

Ali et al. (2023); Srinivas & Parvathi (2026)

Explainability matters clinically, not just technically: a radiologist who is professionally accountable for a report cannot defend an output they cannot interrogate.

Avanzo et al. (2024); Sistaninejhad et al. (2023); Ali et al. (2023); Srinivas & Parvathi (2026).

What AI does across the diagnostic pathway

01

Detection

Identifies tumours with high sensitivity, flagging suspicious regions that might escape initial radiological review.

Rahman et al. (2025)

02

Segmentation

Automatically delineates tumour boundaries — including necrotic cores and peritumoral oedema — with Dice coefficients approaching expert level. Fine-tuned models segment diffuse gliomas accurately enough for clinical use.

Bibault & Giraud (2023); Evangelou et al. (2025)

03

Classification & grading

Distinguishes glioma, meningioma and pituitary adenoma and predicts tumour grade. Deep learning consistently outperforms traditional machine learning on both tasks.

Al-Rumaihi et al. (2025)

04

Molecular prediction

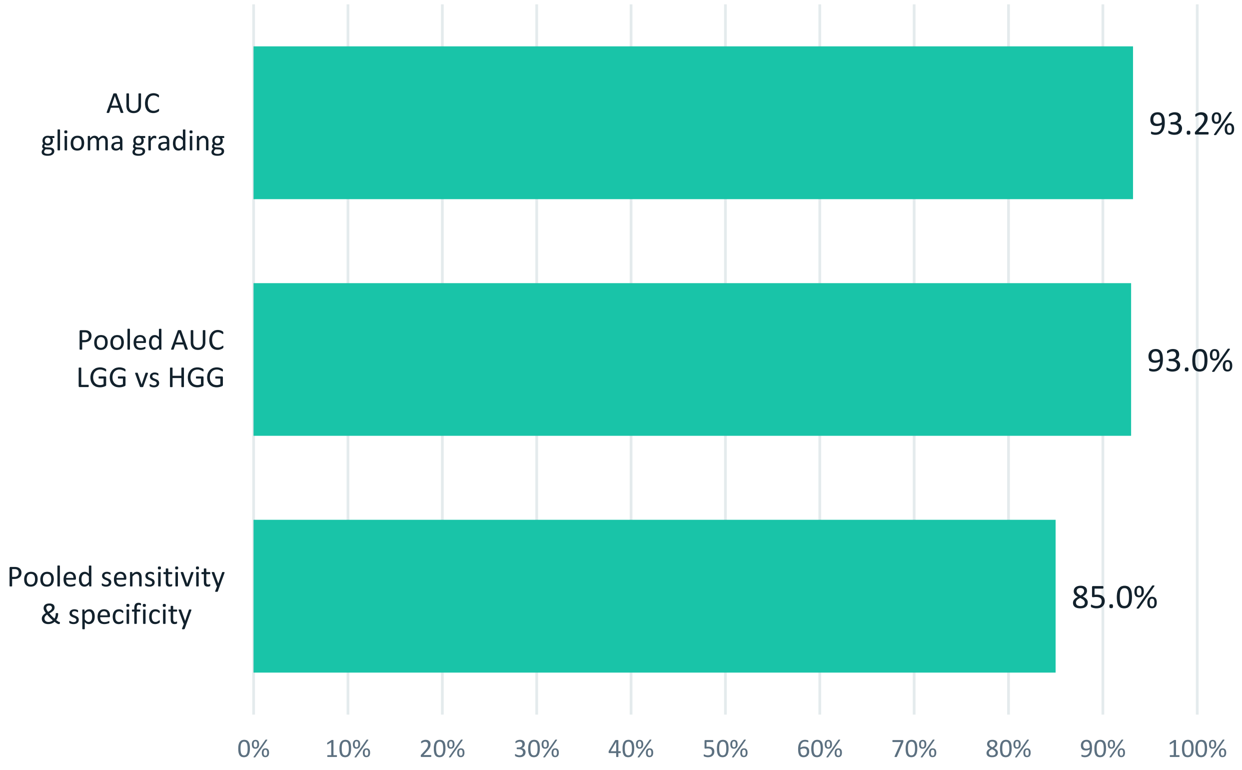
Infers molecular markers and genetic subtypes from imaging alone — with the potential to reduce reliance on invasive biopsy.

Khalighi et al. (2024)

Capability now spans the whole workflow — not a single point task.

Rahman et al. (2025); Bibault & Giraud (2023); Evangelou et al. (2025); Al-Rumaihi et al. (2025); Khalighi et al. (2024).

On paper, the performance is strong



Pooled sensitivity and specificity exceeded 85% across CNNs, support vector machines and ensemble methods in a systematic review and meta-analysis.

Al-Rumaihi et al. (2025)

A separate meta-analysis reported pooled AUC above 93% for separating low-grade from high-grade gliomas; a single-study AUC of 93.2% was reported for grading.

Sun et al. (2023); Bhandari (2024)

In neuroradiology specifically, AI systems have shown particular effectiveness in detecting intracranial haemorrhage and assessing brain tumours.

Pitarch et al. (2023)

Read the small print: these are pooled figures, largely from retrospective, often single-institution datasets.

Al-Rumaihi et al. (2025); Sun et al. (2023); Bhandari (2024); Pitarch et al. (2023). Metrics differ in kind and are not directly comparable with one another.

But the laboratory is not the clinic

The central tension in this literature: pooled figures look consistently strong, yet external-validation and case-level evidence suggest they overstate how reliably AI performs in practice.

External validation lowers the numbers

A systematic review of 19 brain tumour AI studies with external validation datasets reported a pooled lesion-wise Dice score of 84% and sensitivities of roughly 86–87% — notably below the figures typically reported on internal validation.

Wang et al. (2024)

The scanner itself predicts failure

Meta-regression found MRI scanner manufacturer diversity to be a statistically significant predictor of generalisability. Performance degrades predictably when a model meets scanner hardware unrepresented in its training data.

Wang et al. (2024)

Approved tools still get it wrong

An FDA-approved neuroimaging tool misinterpreted CT head angiograms in multiple documented cases against experienced neuroradiologists — failing on post-surgical anatomy, image artifacts, and pathologies needing clinical context.

Young et al. (2024)

Domain shift — degraded performance when a model meets data from different scanners, protocols or populations — remains the single largest barrier to clinical translation (Oyeniyi, 2025).

Four barriers standing between the model and the ward

Workflow and technical integration

AI must slot into existing PACS and RIS without disrupting how radiologists already work — demanding real-time processing, usable interfaces and interoperability across diverse healthcare IT.

Dikici et al. (2020)

Alert fatigue and over-reliance

Too many false positives and clinicians start ignoring alerts. Too much trust and automation bias sets in — favouring the machine's suggestion even when it is wrong. Verifying critical findings protects safety but erodes the speed gains.

Abdelwanis et al. (2024); Cross et al. (2024)

Regulatory burden

Approval demands extensive validation of safety and efficacy — slow and expensive. Devices such as Pixyl Neuro and Rapid ASPECTS have cleared it, but adaptive systems that keep learning fit awkwardly into traditional device rules.

Sivakumar et al. (2025)

Ethics, consent and liability

Patient privacy under data-hungry models; whether patients understand when AI has shaped their diagnosis; and who is liable when an AI-assisted diagnosis proves wrong. Professional guidance is being drafted, but much is unresolved.

Cestonaro et al. (2023)

Only a minority of institutions have reached the point where clinicians can correct AI outputs and feed those corrections back into the model.

The question is partnership, not replacement

AI is fast, tireless and consistent. Radiologists bring clinical judgement, contextual understanding, adaptability to unusual presentations and ethical oversight. Two frameworks have emerged for combining them.

	Human-in-the-loop	Human-on-the-loop
Radiologist's role	Reviews, modifies, and accepts or rejects every AI output before it becomes final.	Monitors performance and supplies feedback for improvement, without intervening at every decision.
Best suited to	Outputs entering the diagnostic record — exemplified by AI-assisted lung nodule identification inside PACS.	Automated image quality optimisation and protocol selection.
Trade-off	Preserves human oversight at critical decision points, but diminishes the potential speed gains.	Faster and lighter-touch, but oversight is less granular.

Worklist prioritisation

AI identifies urgent findings and reorders the radiologist's queue, enabling faster response to critical cases. But a standalone system cannot dictate priorities on its own — it has to sit inside broader worklist drivers balancing study availability, ordering department and service-level agreements.

Tejani et al. (2024); Baltruschat et al. (2020).

Integrate in three phases, not one leap

PHASE 1

Research maturity

AI results are visualised for radiologists but never generate a patient record. Clinical validation happens in a low-risk environment where performance can be assessed without affecting care decisions — and feedback establishes whether the tool meets a real clinical need.

PHASE 2

Production maturity

AI-generated results enter the official record and are stored in institutional PACS. Radiologists review and validate every output before finalising reports. Requires robust technical integration, workflows that state clearly who is responsible for what, and comprehensive training.

PHASE 3

Feedback maturity

Radiologists edit AI results and those corrections feed back into retraining — a virtuous cycle where clinical use improves the model over time. This demands sophisticated infrastructure and sustained institutional commitment; few departments have reached it.

Choose the first application carefully. Start with high-volume, well-defined tasks that have objective evaluation standards — detection, segmentation, tumour volume quantification — and especially where inter-observer variability is already high. Avoid rare conditions with thin training data and tasks needing nuanced clinical judgement. Credibility built on simpler applications is what earns the more ambitious ones.

Framework adapted from Dikici et al. (2020); application selection informed by Mohebbi et al. (2025).

Where the field needs to go next

Standardisation

The AI-RANO initiative is establishing guidelines for standardisation, validation and good clinical practice in neuro-oncology. Standardised AI assessment could replace the current reliance on individual radiologist measurements to dictate treatment options, giving more objective and reproducible evaluation.

Villanueva-Meyer et al. (2024); Jeong et al. (2025)

Multi-institutional evidence

Most models are still built and validated at single academic centres with standardised protocols. Collaborations pooling diverse datasets remain uncommon, and federated learning — which could enable privacy-preserving collaboration — is still largely experimental.

Baduwal et al. (2026); Oyeniyi (2025)

Beyond the first scan

Brain tumour care is longitudinal, but research concentrates on single time points. Tracking tumour volume and treatment response over time is an underexplored, high-value need — as is real-time intraoperative guidance to maximise resection while preserving critical tissue.

Nensa (2025); Mansour et al. (2025)

The long-term vision is comprehensive, individualised planning — systems integrating patient-specific imaging, genetics and population-level outcomes to recommend management strategies. Not replacing clinical judgement, but providing evidence-based decision support that raises the quality and consistency of care.

CONCLUSION

Cautious optimism, not uncritical enthusiasm

The capability is real.

AI detects, segments and classifies brain tumours at levels rivalling experienced radiologists — reducing inter-observer variability, offering objective second opinions and catching findings that might be missed.

The gap is translation, not technology.

Generalisability across settings is only partly solved, workflow integration needs clinical input, regulation is still evolving, and trust has yet to be fully earned.

The framing should be partnership.

AI will not replace radiologists; it will shift their role toward diagnostic orchestration and complex interpretative work — with efficiency gains that may help address workforce shortages.

The question has changed.

Not whether to engage with AI, but how to do so thoughtfully and strategically, through prospective validation, standardised evaluation and collaborative human–AI models.

Thank you — I'm happy to take questions.

References

Abdelwanis, M., Alarafati, H. K., Tammam, M. M. S., & Simsekler, M. C. E. (2024). Exploring the risks of automation bias in healthcare artificial intelligence applications: A Bowtie analysis. Journal of Safety Science and Resilience, 5(4), 460–469. https://doi.org/10.1016/j.jnlssr.2024.06.001

Al-Rumaihi, G., Khan, M. M., Saleh, A., Ali, A., Al-Romaihi, L., Al-Jaber, N., Al-Suliaiti, G., Chowdhury, M. E., Babu, G. R., & Pedersen, S. (2025). Performance evaluation of artificial intelligence techniques in the diagnosis of brain tumors: A systematic review and meta-analysis. Cureus, 17(7), e88915. https://doi.org/10.7759/cureus.88915

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. Information Fusion, 99, 101805. https://doi.org/10.1016/j.inffus.2023.101805

Avanzo, M., Stancanello, J., Pirrone, G., Drigo, A., & Retico, A. (2024). The evolution of artificial intelligence in medical imaging: From computer science to machine and deep learning. Cancers, 16(21), 3702. https://doi.org/10.3390/cancers16213702

Baduwal, M., Paudel, P., & Chaudhary, V. (2026). Federated learning: A survey of core challenges, current methods, and opportunities. Computers, 15(3), 155. https://doi.org/10.3390/computers15030155

Baltruschat, I., Steinmeister, L., Nickisch, H., Saalbach, A., Grass, M., Adam, G., Knopp, T., & Ittrich, H. (2020). Smart chest X-ray worklist prioritization using artificial intelligence: A clinical workflow simulation. European Radiology, 31(6), 3837–3845. https://doi.org/10.1007/s00330-020-07480-7

Bhandari, A. (2024). Revolutionizing radiology with artificial intelligence. Cureus, 16(10), e72646. https://doi.org/10.7759/cureus.72646

Bibault, J., & Giraud, P. (2023). Deep learning for automated segmentation in radiotherapy: A narrative review. British Journal of Radiology, 97(1153), 13–20. https://doi.org/10.1093/bjr/tqad018

Cestonaro, C., Delicati, A., Marcante, B., Caenazzo, L., & Tozzo, P. (2023). Defining medical liability when artificial intelligence is applied on diagnostic algorithms: A systematic review. Frontiers in Medicine, 10, 1305756. https://doi.org/10.3389/fmed.2023.1305756

Cross, J. L., Choma, M. A., & Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. PLOS Digital Health, 3(11), e0000651. https://doi.org/10.1371/journal.pdig.0000651

Dikici, E., Bigelow, M., Prevedello, L. M., White, R. D., & Erdal, B. S. (2020). Integrating AI into radiology workflow: Levels of research, production, and feedback maturity. Journal of Medical Imaging, 7(01), 1. https://doi.org/10.1117/1.jmi.7.1.016502

Evangelou, K., Kotsantis, I., Kalyvas, A., Kyriazoglou, A., Economopoulou, P., Velonakis, G., Gavra, M., Psyrri, A., Boviatsis, E. J., & Stavrinou, L. C. (2025). Artificial intelligence in the diagnosis and treatment of brain gliomas. Biomedicines, 13(9), 2285. https://doi.org/10.3390/biomedicines13092285

References (cont. 2/3)

Filho, A. M., Znaor, A., Sunguc, C., Zahwe, M., Marcos-Gragera, R., Figueroa, J. D., & Bray, F. (2025). Cancers of the brain and central nervous system: Global patterns and trends in incidence. Journal of Neuro-Oncology, 172(3), 567–578. https://doi.org/10.1007/s11060-025-04944-y

Godkhindi, V. M., Monappa, V., Kairanna, N. V., Sharma, S., Vasudevan, G., & Hebbar, K. D. (2022). Brain infections that mimic malignancy. Diagnostic Histopathology, 28(10), 456–466. https://doi.org/10.1016/j.mpdhp.2022.08.009

Jeong, J., Kim, S., Pan, L., Hwang, D., Kim, D., Choi, J., Kwon, Y., Yi, P., Jeong, J., & Yoo, S. (2025). Reducing the workload of medical diagnosis through artificial intelligence: A narrative review. Medicine, 104(6), e41470. https://doi.org/10.1097/md.0000000000041470

Khalighi, S., Reddy, K., Midya, A., Pandav, K. B., Madabhushi, A., & Abedalthagafi, M. (2024). Artificial intelligence in neuro-oncology: Advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. NPJ Precision Oncology, 8(1), 80. https://doi.org/10.1038/s41698-024-00575-0

Mahmood, U., Shukla-Dave, A., Chan, H., Drukker, K., Samala, R. K., Chen, Q., Vergara, D., Greenspan, H., Petrick, N., Sahiner, B., Huo, Z., Summers, R. M., Cha, K. H., Tourassi, G., Deserno, T. M., Grizzard, K. T., Näppi, J. J., Yoshida, H., Regge, D., … Hadjiiski, L. (2024). Artificial intelligence in medicine: Mitigating risks and maximizing benefits via quality assurance, quality control, and acceptance testing. BJR|Artificial Intelligence, 1(1), ubae003. https://doi.org/10.1093/bjrai/ubae003

Mansour, M. A., El-Salamoni, M. A., Zohney, M., Elshaer, A. M., Abdelwahab, M., Aziz, M. M., Ayoub, B., Mostafa, H. N., & El-Samman, A. (2025). Neurosurgery reimagined: How AI is redefining patient care and surgical excellence. Intelligent Surgery, 8, 84–91. https://doi.org/10.1016/j.isurg.2025.06.002

Mohebbi, A., Mohammadzadeh, S., Zare, A. H., Moradi, Z., Ardakani, A. A., & Mohammadi, A. (2025). Assessing inter-rater reliability of MRI features in glioma: A multi-radiologist agreement study. BMC Medical Imaging, 25(1), 480. https://doi.org/10.1186/s12880-025-01941-5

Najjar, R. (2023). Redefining radiology: A review of artificial intelligence integration in medical imaging. Diagnostics, 13(17), 2760. https://doi.org/10.3390/diagnostics13172760

Nensa, F. (2025). The future of radiology: The path towards multimodal AI and superdiagnostics. European Journal of Radiology Artificial Intelligence, 2, 100014. https://doi.org/10.1016/j.ejrai.2025.100014

Oyeniyi, J. G. (2025). From lab to clinic: Addressing bias and generalizability in AI diagnostic systems. World Journal of Advanced Research and Reviews, 28(3), 2134–2179. https://doi.org/10.30574/wjarr.2025.28.3.4249

Pitarch, C., Ribas, V., & Vellido, A. (2023). AI-based glioma grading for a trustworthy diagnosis: An analytical pipeline for improved reliability. Cancers, 15(13), 3369. https://doi.org/10.3390/cancers15133369

Rahman, A., Hayat, M., Iqbal, N., Alarfaj, F. K., Alkhalaf, S., & Alturise, F. (2025). Enhanced MRI brain tumor detection using deep learning in conjunction with explainable AI SHAP based diverse and multi feature analysis. Scientific Reports, 15(1), 29411. https://doi.org/10.1038/s41598-025-14901-4

References (cont. 3/3)

Reddy, S. (2024). Global harmonization of artificial intelligence-enabled software as a medical device regulation: Addressing challenges and unifying standards. Mayo Clinic Proceedings: Digital Health, 3(1), 100191. https://doi.org/10.1016/j.mcpdig.2024.100191

Sistaninejhad, B., Rasi, H., & Nayeri, P. (2023). A review paper about deep learning for medical image analysis. Computational and Mathematical Methods in Medicine, 2023, 7091301. https://doi.org/10.1155/2023/7091301

Sivakumar, R., Lue, B., & Kundu, S. (2025). FDA approval of artificial intelligence and machine learning devices in radiology. JAMA Network Open, 8(11), e2542338. https://doi.org/10.1001/jamanetworkopen.2025.42338

Srinivas, V. R., & Parvathi, R. (2026). Explainable AI-driven MRI-based brain tumor classification: A novel deep learning approach. Frontiers in Artificial Intelligence, 8, 1700214. https://doi.org/10.3389/frai.2025.1700214

Sun, W., Song, C., Tang, C., Pan, C., Xue, P., Fan, J., & Qiao, Y. (2023). Performance of deep learning algorithms to distinguish high-grade glioma from low-grade glioma: A systematic review and meta-analysis. iScience, 26(6), 106815. https://doi.org/10.1016/j.isci.2023.106815

Tejani, A. S., Cook, T. S., Hussain, M., Schmidt, T. S., & O'Donnell, K. P. (2024). Integrating and adopting AI in the radiology workflow: A primer for standards and integrating the healthcare enterprise (IHE) profiles. Radiology, 311(3), e232653. https://doi.org/10.1148/radiol.232653

Varnosfaderani, S. M., & Forouzanfar, M. (2024). The role of AI in hospitals and clinics: Transforming healthcare in the 21st century. Bioengineering, 11(4), 337. https://doi.org/10.3390/bioengineering11040337

Villanueva-Meyer, J. E., Bakas, S., Tiwari, P., Lupo, J. M., Calabrese, E., Davatzikos, C., Bi, W. L., Ismail, M., Akbari, H., Lohmann, P., Booth, T. C., Wiestler, B., Aerts, H. J. W. L., Rasool, G., Tonn, J. C., Nowosielski, M., Jain, R., Colen, R. R., Pati, S., … Galldiks, N. (2024). Artificial intelligence for response assessment in neuro oncology (AI-RANO), part 1: Review of current advancements. The Lancet Oncology, 25(11), e581–e588. https://doi.org/10.1016/S1470-2045(24)00316-4

Wang, T.-W., Shiao, Y.-C., Hong, J.-S., Lee, W.-K., Hsu, M.-S., Cheng, H.-M., Yang, H.-C., Lee, C.-C., Pan, H.-C., You, W. C., Lirng, J.-F., Guo, W.-Y., & Wu, Y.-T. (2024). Artificial intelligence detection and segmentation models: A systematic review and meta-analysis of brain tumors in magnetic resonance imaging. Mayo Clinic Proceedings: Digital Health, 2(1), 75–91. https://doi.org/10.1016/j.mcpdig.2024.01.002

Young, A., Tan, K., Tariq, F., Jin, M. X., & Bluestone, A. Y. (2024). Rogue AI: Cautionary cases in neuroradiology and what we can learn from them. Cureus, 16(3), e56317. https://doi.org/10.7759/cureus.56317

Slide 2 statistics: U.S. FDA AI-Enabled Medical Device List (authorisations through December 2025), as reported by The Imaging Wire (2026).