JHWCR JOURNAL OF HEALTH, WELLNESS AND COMMUNITY RESEARCH 3007-0570

https://doi.org/10.61919/zstegs75

Original Article

AI-Assisted MRI for Brain Tumor Segmentation: A Cross-Sectional Diagnostic Accuracy Study

Maryam Noor[1], Sania Naveed[2], Iqra Kanwal[3], Mahum[4], Minahil Qasam[5], Alaina Asghar[6], Noor Ul Ain[7]

[1] MPhil Chemistry, Specialization in Inorganic Chemistry, School of Chemistry, University of the Punjab, Lahore, Pakistan
[2] Resident Radiology, INMOL Cancer Hospital, Lahore, Pakistan
[3] BS Medical Imaging, MPhil Forensic Science, Senior Lecturer, Riphah International University, Gujranwala, Pakistan
[4] MBBS, King Edward Medical University, Lahore, Pakistan
[5] Physiotherapist, Health Services Academy, Islamabad, Pakistan
[6] BS Medical Technology Specialization in Nuclear Medicine and Radiation Therapy, Institute of Medical Technology, Jinnah Sindh Medical University, Karachi, Pakistan
[7] PhD, Postdoc, Translational Brain Research, State Key Laboratory of Medical Neurobiology, MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China

***Corresponding author: Noor Ul Ain, noorulain22@yahoo.com**

"Cite this Article" Received: 23 December 2025; **Accepted**: 19 February 2026; **Published**: 15 March 2026

Author Contributions: Concept: MN, NS; Design: IK, M; Data Collection: MQ, NS; Analysis: NUA, IK; Drafting: MN, NUA. **Ethical Approval**: Hameed Latif Hospital, Lahore, Pakistan. **Informed Consent**: Written informed consent was obtained from all participants; **Conflict of Interest**: The authors declare no conflict of interest. **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: N/A.

ABSTRACT

Background: Brain tumor segmentation on MRI is essential for diagnosis, treatment planning, radiotherapy guidance, and follow-up assessment, but manual delineation is time-consuming and subject to reader variability. Artificial intelligence may improve segmentation consistency, although local validation is required before clinical use. Objective: To evaluate the diagnostic accuracy and segmentation agreement of AI-assisted MRI for brain tumor delineation compared with expert radiologist consensus assessment. Methods: This prospective cross-sectional diagnostic accuracy study included 124 adult patients undergoing MRI brain for suspected or diagnosed intracranial tumor in Lahore, Pakistan. MRI sequences included routine brain tumor imaging sequences, with contrast-enhanced assessment used where clinically indicated. AI-generated tumor masks were compared with expert radiologist consensus segmentation as the reference standard. Sensitivity, specificity, positive predictive value, negative predictive value, overall accuracy, Dice similarity coefficient, and absolute tumor volume difference were calculated. Results: AI-assisted MRI demonstrated sensitivity of 91.8%, specificity of 88.6%, positive predictive value of 90.4%, negative predictive value of 89.9%, and overall accuracy of 90.2%. The mean Dice similarity coefficient for whole tumor segmentation was 0.84 ± 0.09, with highest agreement in enhancing tumors (0.89 ± 0.06) and lowest agreement in post-treatment lesions (0.73 ± 0.11). The mean absolute AI-expert tumor volume difference was 5.9 ± 4.7 cm^3. Conclusion: AI-assisted MRI showed promising segmentation accuracy and may support radiologists by providing an initial tumor outline, particularly in enhancing lesions. Expert review remains necessary in non-enhancing and post-treatment cases, and multicenter validation is required before routine clinical adoption. Keywords: artificial intelligence, MRI, brain tumor, segmentation, diagnostic accuracy, glioma, Pakistan.

INTRODUCTION

Brain tumors represent a clinically important group of intracranial disorders because even small lesions can affect neurological function through mass effect, infiltration, edema, hemorrhage, or disruption of eloquent cortical and subcortical pathways. Magnetic resonance imaging is central to neuro-oncology because it provides superior soft-tissue contrast and enables assessment of tumor location, enhancement pattern, edema, necrosis, mass effect, treatment response, and surgical or radiotherapy planning margins. In Pakistan, the burden of brain tumors is increasingly recognized through multicentre epidemiological work, including the Pakistan Brain Tumour Epidemiology Study, which reported 2750 cases and showed that MRI was used more frequently than CT for diagnostic evaluation, with gliomas

forming the most common tumor group in the national dataset (1). Local epidemiological studies have also indicated that the distribution of central nervous system tumors in Pakistan may differ from international datasets, reinforcing the need for locally relevant imaging evidence rather than direct reliance on external populations alone (2). These local data are clinically important because Pakistani neuro-oncology services continue to face challenges related to documentation, referral pathways, multidisciplinary coordination, follow-up, and access to standardized treatment resources, making accurate and reproducible imaging assessment particularly valuable in tertiary care settings (3,4).

Brain tumor segmentation is the process of delineating tumor-related regions on MRI, including the whole tumor, enhancing tumor, tumor core, necrotic component, non-enhancing infiltrative disease, and peritumoral edema depending on the clinical and imaging context. This step is not merely a technical exercise; it directly influences baseline tumor burden assessment, operative planning, radiotherapy field design, volumetric follow-up, and response evaluation. However, tumor boundaries are often irregular and biologically heterogeneous. Enhancing margins may be visually distinct on post-contrast T1-weighted images, whereas infiltrative non-enhancing tumor and edema may overlap on T2-weighted and FLAIR sequences. Postoperative and post-radiotherapy scans introduce further complexity because residual or recurrent tumor may resemble treatment-related enhancement, gliosis, necrosis, or inflammatory change. The 2021 WHO classification has further emphasized the biological heterogeneity of central nervous system tumors, while response assessment frameworks such as RANO highlight the continuing difficulty of standardizing radiological interpretation in treated gliomas (5,6).

Manual segmentation by expert radiologists is commonly treated as the reference approach in imaging validation studies, but it is time-consuming and vulnerable to inter-reader variability, particularly at indistinct tumor margins. Variation in manual delineation can affect tumor volume estimates, treatment planning, progression assessment, and research reproducibility. This limitation has encouraged the development of artificial intelligence approaches, particularly deep learning-based segmentation models, to provide faster and more consistent tumor masks from multimodal MRI. Public benchmark datasets and expert-labelled imaging repositories, including BraTS and TCGA glioma MRI segmentation resources, have accelerated algorithm development by providing standardized multimodal MRI data and expert annotations for model training and comparison (7,8). Foundational convolutional network architectures such as U-Net, 3D U-Net, V-Net, and multi-scale 3D convolutional models established the technical basis for biomedical and volumetric segmentation, while later adaptive frameworks such as nnU-Net improved automated configuration across segmentation tasks (9–15). Toolkits translating BraTS-derived algorithms into scientific and clinical research workflows further demonstrate the increasing maturity of AI-based brain tumor segmentation (16).

International evidence suggests that AI models can achieve promising performance for brain tumor MRI detection and segmentation, but reported accuracy varies across datasets, imaging protocols, tumor types, lesion compartments, and validation settings. A recent systematic review and meta-analysis found encouraging diagnostic and segmentation performance of AI models in brain tumor MRI, while also emphasizing concerns related to generalizability, dataset diversity, and external validation (17). Individual studies using convolutional neural networks and deep neural networks have reported strong segmentation performance when multimodal MRI sequences are available, but boundary errors remain common in infiltrative, non-enhancing, edematous, or post-treatment lesions (13,14,18). Automated MRI assessment has also shown potential for quantitative response evaluation in neuro-oncology, but expert supervision and clinical validation remain essential before routine implementation (19). Importantly, segmentation metrics such as Dice similarity coefficient do not always fully reflect expert clinical judgment, because a numerically acceptable overlap may still miss a clinically meaningful tumor extension or include treatment-related abnormality that should not be counted as tumor (20).

The key knowledge gap is therefore not whether AI can segment brain tumors under ideal benchmark conditions, but whether AI-assisted segmentation performs reliably in local clinical MRI data where

scanner protocols, image quality, referral patterns, tumor biology, and post-treatment appearances may differ from international development datasets. This gap is especially relevant in Lahore, Pakistan, where tertiary hospitals receive heterogeneous neuro-oncology referrals from urban and surrounding regional populations. For AI-assisted segmentation to be clinically useful in such settings, its performance must be evaluated against expert radiologist consensus using transparent diagnostic accuracy methodology. The clinically relevant PICO framework for this study was adult patients undergoing MRI for suspected or diagnosed brain tumor as the population, AI-assisted MRI segmentation as the index test, expert radiologist consensus segmentation as the reference standard, and sensitivity, specificity, predictive values, overall accuracy, Dice similarity coefficient, and tumor volume difference as performance outcomes.

Therefore, this prospective cross-sectional diagnostic accuracy study was conducted to evaluate the performance of AI-assisted MRI for brain tumor segmentation in adult patients in Lahore, Pakistan. The primary objective was to compare AI-generated tumor masks with expert radiologist consensus segmentation and determine the sensitivity, specificity, and segmentation overlap of the AI-assisted approach. The study was designed to answer the following research question: among adult patients undergoing MRI for suspected or diagnosed brain tumor, how accurately does AI-assisted MRI segmentation delineate tumor regions compared with expert radiologist consensus assessment?

MATERIALS AND METHODS

This prospective cross-sectional diagnostic accuracy study was designed to evaluate AI-assisted MRI segmentation of brain tumors using expert radiologist consensus segmentation as the reference standard. The study followed the methodological principles of diagnostic accuracy reporting and AI medical imaging reporting, including clear definition of the index test, reference standard, participant eligibility, image handling, blinding, and performance outcomes (21,22). The study was conducted in a tertiary care hospital in Lahore, Pakistan, with radiology and neurosurgery services. Patients were recruited from the radiology department, neurosurgery outpatient service, and admitted neurosurgical wards after institutional ethical approval. All MRI examinations were performed as part of routine clinical care, and no additional imaging sequence was acquired solely for research purposes. Clinical management decisions were made according to usual hospital practice and were not based on the AI segmentation output.

The study population consisted of adult patients aged 18 years or older who were referred for MRI brain because of a suspected or previously diagnosed intracranial tumor. Consecutive eligible patients were enrolled to reduce selection bias and better represent the clinical spectrum encountered in routine neuroimaging. Both newly diagnosed and follow-up cases were included because tumor segmentation is clinically relevant for initial assessment, treatment planning, and surveillance. Patients were eligible when MRI demonstrated an intracranial tumor or suspected tumor lesion and when the available MRI sequences were adequate for expert radiologist delineation and AI-assisted segmentation. The imaging dataset included routine brain tumor MRI sequences such as T1-weighted, T2-weighted, FLAIR, diffusion-weighted, and post-contrast T1-weighted images when contrast administration was clinically indicated and available. Enhancing tumors were assessed primarily on post-contrast T1-weighted images, whereas non-enhancing tumor components and edema were assessed using T2-weighted and FLAIR sequences. Patients were excluded if MRI images were incomplete, severely degraded by motion artefact, unsuitable for segmentation, or showed a non-tumor lesion such as abscess, infarct, demyelination, hematoma, or other mimic after expert review. Cases with missing essential clinical or imaging information were excluded from final analysis.

Written informed consent was obtained from each participant or from an authorized attendant when the patient was unable to provide consent. Clinical and demographic information was recorded on a structured proforma, including age, gender, clinical presentation where available, newly diagnosed or

follow-up status, tumor location, enhancement pattern, and relevant treatment history. Tumor location was categorized anatomically as frontal, temporal, parietal, occipital, cerebellar, brainstem, sellar or parasellar, or other intracranial region. Imaging pattern was categorized as enhancing, non-enhancing, mixed, or post-treatment lesion according to the dominant MRI appearance. Tumor volume was defined as the segmented tumor volume derived from expert radiologist consensus masks and compared with the corresponding AI-generated tumor volume.

MRI images were anonymized before analysis by removing patient identifiers and assigning each case a unique study code. The same anonymized MRI dataset was used for expert reference segmentation and AI-assisted segmentation. Image handling before AI analysis included anonymization, format conversion where required, sequence alignment, and intensity normalization. The AI-assisted segmentation system was used as the index test and generated tumor masks from the available MRI sequences. The AI output was stored separately from the radiologist segmentation files, and expert readers were blinded to the AI-generated masks during their initial manual segmentation to prevent incorporation bias. The AI-generated mask was compared with the expert consensus mask only after both segmentation processes had been completed.

Expert radiologist segmentation served as the reference standard. Two radiologists with neuroimaging experience independently reviewed the anonymized MRI scans and manually delineated visible tumor regions using radiology viewing and segmentation software. In enhancing lesions, the enhancing tumor margin was marked primarily on post-contrast T1-weighted images. In non-enhancing lesions, T2-weighted and FLAIR images were used to delineate visible tumor abnormality, while peritumoral edema was recorded separately where feasible. In post-treatment cases, residual or recurrent tumor was assessed in relation to postoperative change, gliosis, edema, necrosis, and enhancement pattern. Disagreements between the two radiologists were resolved through consensus review, and the final consensus mask was used as the reference standard for diagnostic accuracy and overlap analysis. This approach was used to reduce individual reader bias and provide a clinically defensible comparator for AI segmentation.

The primary outcomes were sensitivity, specificity, and segmentation agreement of AI-assisted MRI compared with expert radiologist consensus segmentation. Sensitivity represented the ability of the AI system to correctly identify tumor-labelled regions defined by the expert consensus mask, whereas specificity represented the ability of the AI system to correctly exclude non-tumor regions. Positive predictive value, negative predictive value, and overall accuracy were calculated from the comparison between AI-generated and expert-defined segmentation labels. Dice similarity coefficient was used as the principal overlap metric, with higher values indicating greater spatial agreement between the AI and expert masks. Mean absolute volume difference was calculated to quantify the difference between AI-derived and expert-derived tumor volumes. Subgroup performance was planned according to imaging pattern, including enhancing tumors, non-enhancing tumors, mixed-pattern tumors, and post-treatment lesions, because these categories differ in boundary clarity and segmentation difficulty.

Bias was addressed through consecutive recruitment, predefined eligibility criteria, anonymization of imaging data, independent expert segmentation, blinding of radiologists to AI output during initial review, and separate storage of AI and expert masks before comparison. Selection bias was minimized by enrolling eligible patients from routine clinical pathways rather than selecting only technically ideal cases. Measurement bias was reduced by using the same imaging dataset for both index and reference assessments. Incorporation bias was avoided by ensuring that AI output did not influence the initial expert segmentation. Spectrum effects were considered by including both newly diagnosed and follow-up cases and by reporting performance across different MRI enhancement patterns.

The sample size was planned for a diagnostic accuracy design using anticipated sensitivity from previous MRI brain tumor segmentation literature and an acceptable precision margin, with additional recruitment to compensate for incomplete scans and poor-quality images. A total of 132 patients were initially assessed, and only complete eligible cases with adequate MRI quality and interpretable expert

segmentation were included in the final analysis. Statistical analysis was performed using statistical software. Continuous variables were summarized as mean and standard deviation when approximately normally distributed and as median with interquartile range when skewed. Categorical variables were summarized as frequencies and percentages. Diagnostic performance metrics, including sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy, were calculated by comparing AI segmentation with expert consensus segmentation. Dice similarity coefficient and absolute volume difference were calculated for overall segmentation performance and by imaging-pattern subgroup. A p-value below 0.05 was considered statistically significant where inferential comparisons were performed. Missing or incomplete imaging data were excluded from final analysis rather than imputed, because segmentation performance required complete image-mask comparability. All data were checked for completeness, internal consistency, and correct linkage between clinical records, MRI study codes, AI masks, and expert consensus masks before analysis.

The study was conducted in accordance with ethical principles for human participant research. Patient confidentiality was protected through anonymization of imaging data and use of coded study files. The AI segmentation result was used only for research comparison and did not replace expert radiologist interpretation or alter patient management. Access to study data was restricted to the research team, and segmentation files were stored separately from identifiable clinical information to maintain privacy and data integrity.

RESULTS

A total of 132 patients were assessed for eligibility during the study period. Eight patients were excluded because the MRI dataset was incomplete, image quality was inadequate for segmentation, or lesion margins were not sufficiently interpretable for valid comparison. The final analysis included 124 adult patients with MRI evidence of suspected or diagnosed intracranial tumor and complete segmentation data.

Table 1. Patient Flow and Final Analytical Sample

Study Flow Component	n	%
Patients assessed for eligibility	132	100.0
Patients excluded	8	6.1
Patients included in final analysis	124	93.9

The final analytical cohort represented 93.9% of all initially assessed patients. Exclusion was limited to 8 cases, reflecting a small proportion of screened patients; however, these exclusions were necessary because diagnostic segmentation accuracy requires complete MRI sequences and interpretable lesion boundaries for valid comparison between AI-generated and expert reference masks.

Table 2. Baseline Demographic and MRI Characteristics of Included Patients

Variable	n / Mean	% / SD
Total patients	124	100.0
Age, years	46.8	14.2
Male	70	56.5
Female	54	43.5
Frontal lobe	38	30.6
Temporal lobe	27	21.8
Parietal lobe	18	14.5
Cerebellar	13	10.5
Sellar / parasellar	10	8.1
Other sites	18	14.5
Enhancing pattern	57	46.0
Non-enhancing pattern	25	20.2
Mixed pattern	32	25.8
Post-treatment lesion	10	8.0
Expert-segmented tumor volume, cm^3	38.6	21.4

The included patients had a mean age of 46.8 ± 14.2 years, and males formed a slightly larger proportion of the sample than females. Frontal lobe tumors were the most frequent anatomical category, accounting for 38 of 124 cases, followed by temporal lobe lesions in 27 patients and parietal lobe lesions in 18 patients. The most common imaging pattern was contrast enhancement, observed in 57 patients, while non-enhancing lesions were present in 25 patients and post-treatment lesions in 10 patients. The mean tumor volume derived from expert segmentation was 38.6 ± 21.4 cm^3, indicating a clinically heterogeneous sample with substantial variation in lesion size.

Table 3. Overall Diagnostic Accuracy and Segmentation Performance of AI-Assisted MRI

Performance Measure	Value
Sensitivity, %	91.8
Specificity, %	88.6
Positive predictive value, %	90.4
Negative predictive value, %	89.9
Overall accuracy, %	90.2
Dice similarity coefficient, mean ± SD	0.84 ± 0.09
Absolute volume difference, cm^3, mean ± SD	5.9 ± 4.7

AI-assisted MRI segmentation demonstrated high overall performance when compared with expert radiologist consensus segmentation. Sensitivity was 91.8%, indicating that most expert-defined tumor regions were captured by the AI-generated masks, while specificity was 88.6%, showing acceptable exclusion of non-tumor regions. The overall accuracy was 90.2%, with a positive predictive value of 90.4% and a negative predictive value of 89.9%. Spatial agreement between AI and expert segmentation was good, with a mean Dice similarity coefficient of 0.84 ± 0.09. The mean absolute tumor volume difference between AI-generated and expert-derived segmentations was 5.9 ± 4.7 cm^3, indicating measurable but clinically relevant variation in segmentation volume.

Table 4. Dice Similarity Coefficient According to MRI Tumor Pattern

MRI Tumor Pattern	n	Dice Similarity Coefficient, Mean ± SD
Enhancing tumors	57	0.89 ± 0.06
Non-enhancing tumors	25	0.78 ± 0.10
Mixed pattern tumors	32	0.84 ± 0.08
Post-treatment lesions	10	0.73 ± 0.11

Segmentation overlap varied according to tumor imaging pattern. The highest agreement was observed in enhancing tumors, where the mean Dice similarity coefficient was 0.89 ± 0.06. Mixed-pattern tumors showed intermediate agreement, with a mean Dice coefficient of 0.84 ± 0.08. Non-enhancing tumors showed lower agreement, with a mean Dice coefficient of 0.78 ± 0.10, reflecting the greater difficulty of differentiating infiltrative tumor from edema or adjacent brain tissue on T2-weighted and FLAIR imaging. Post-treatment lesions showed the lowest agreement, with a mean Dice coefficient of 0.73 ± 0.11, consistent with the additional segmentation difficulty created by postoperative change, gliosis, treatment effect, necrosis, and irregular enhancement patterns.

Table 5. Qualitative Error Pattern Observed During AI-Assisted Segmentation Review

Error Pattern	Main Imaging Context
Marginal over-segmentation	Peritumoral edema
Marginal under-segmentation	Non-enhancing tumor component
False-positive marking	Edema, postoperative gliosis, contrast leakage
Reduced overlap	Post-treatment lesion
Higher overlap	Enhancing tumor

Visual review showed that most segmentation differences occurred at lesion margins rather than within clearly abnormal tumor regions. AI-assisted segmentation tended to perform best when tumor borders were sharply visible on contrast-enhanced imaging. Reduced overlap was more evident in non-enhancing and post-treatment lesions, where the distinction between active tumor, edema, gliosis, necrosis, and treatment-related change was less clear. These findings indicate that AI-generated masks may provide a

useful initial segmentation outline, but expert radiologist review remains necessary, particularly in complex lesions with indistinct margins or prior treatment-related imaging changes.

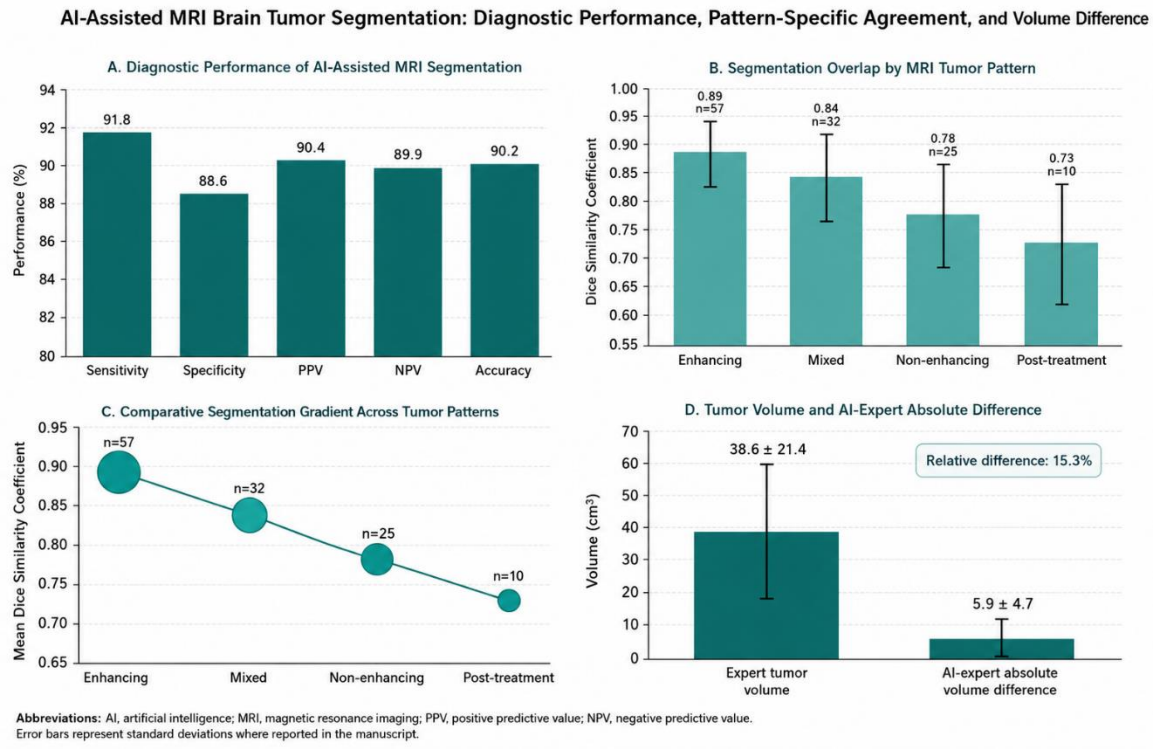

Figure 1 AI-Assisted MRI Brain Tumor Segmentation Performance Across Diagnostic, Pattern-Specific, and Volumetric Domains. The panelled figure demonstrates that AI-assisted MRI segmentation achieved consistently high diagnostic performance, with sensitivity of 91.8%, specificity of 88.6%, positive predictive value of 90.4%, negative predictive value of 89.9%, and overall accuracy of 90.2%. Pattern-specific segmentation agreement showed a clear performance gradient, with the highest Dice similarity coefficient in enhancing tumors (0.89 ± 0.06; n = 57), followed by mixed-pattern tumors (0.84 ± 0.08; n = 32), non-enhancing tumors (0.78 ± 0.10; n = 25), and post-treatment lesions (0.73 ± 0.11; n = 10). The mean expert-segmented tumor volume was 38.6 ± 21.4 cm³, while the mean AI-expert absolute volume difference was 5.9 ± 4.7 cm³, corresponding to an approximate relative difference of 15.3% against the expert-derived mean tumor volume. These findings indicate strong AI performance in clearly enhancing lesions but reduced segmentation agreement in non-enhancing and post-treatment cases, where tumor boundaries are more likely to overlap with edema, gliosis, necrosis, or treatment-related signal change.

DISCUSSION

This prospective cross-sectional diagnostic accuracy study evaluated AI-assisted MRI segmentation of brain tumors against expert radiologist consensus segmentation in an adult clinical cohort from Lahore, Pakistan. The principal finding was that AI-assisted segmentation showed high sensitivity, acceptable specificity, and good spatial agreement with expert-defined tumor masks. The AI system achieved sensitivity of 91.8%, specificity of 88.6%, positive predictive value of 90.4%, negative predictive value of 89.9%, and overall accuracy of 90.2%, while the mean Dice similarity coefficient for whole tumor segmentation was 0.84 ± 0.09. These findings suggest that AI-generated tumor masks captured most expert-defined tumor regions and provided a reproducible initial segmentation framework, although the observed false-positive and false-negative patterns confirm that expert radiologist oversight remains necessary before clinical use.

The clinical relevance of these findings is linked to the central role of MRI in neuro-oncology. In Pakistan, MRI is frequently used for brain tumor evaluation, and national epidemiological data have shown gliomas to be a major tumor group among reported brain tumor cases (1). Local and regional studies have also highlighted differences in central nervous system tumor patterns and neuro-oncology service structures when compared with international datasets, supporting the need for locally validated imaging evidence (2–4). In such settings, AI-assisted segmentation may be valuable as a supportive tool because radiology departments often manage high imaging workloads and heterogeneous referral

patterns. However, the present study validates technical segmentation performance rather than downstream clinical outcomes, and therefore the findings should be interpreted as evidence of segmentation support rather than proof of independent diagnostic or therapeutic decision-making utility.

The overall Dice similarity coefficient of 0.84 ± 0.09 is consistent with previous deep learning-based brain tumor segmentation studies, where multimodal MRI inputs have improved spatial overlap between automated and expert-derived masks (13,14,18). The observed performance also aligns with recent evidence from systematic review and meta-analysis showing that AI models for brain tumor MRI detection and segmentation can achieve promising results, while also demonstrating variability across datasets, validation methods, MRI protocols, and tumor subtypes (17). This variability is important because segmentation algorithms trained or benchmarked on curated international datasets may perform differently when applied to real-world clinical imaging from local hospitals, where acquisition quality, slice thickness, contrast timing, referral stage, and post-treatment appearances may not be uniform. The present study therefore contributes local clinical evidence from a Pakistani tertiary care setting and supports the principle that AI models should be validated in the population and imaging environment in which they are intended to be used.

Performance differed according to MRI tumor pattern. The highest overlap was observed in enhancing tumors, with a mean Dice coefficient of 0.89 ± 0.06. This finding is clinically expected because enhancing tumor margins are usually more conspicuous on post-contrast T1-weighted imaging, allowing both radiologists and AI models to distinguish abnormal tissue from surrounding brain with greater confidence. The result is consistent with the conceptual framework used in multimodal brain tumor segmentation benchmarks, where contrast-enhanced T1-weighted imaging is particularly useful for delineating enhancing tumor components, while T2-weighted and FLAIR images support assessment of edema and non-enhancing tumor regions (7,8). The strong performance in enhancing tumors indicates that AI-assisted segmentation may be most reliable when tumor boundaries are radiologically distinct and contrast between tumor and adjacent tissue is high.

By contrast, non-enhancing tumors had a lower mean Dice coefficient of 0.78 ± 0.10. This reduced agreement is clinically meaningful because non-enhancing infiltrative tumor frequently merges gradually with peritumoral edema or adjacent abnormal brain signal on T2-weighted and FLAIR sequences. The lower Dice value should therefore not be interpreted solely as a model failure; it also reflects the inherent difficulty of defining a biologically infiltrative boundary on conventional MRI. Earlier segmentation studies have similarly shown that tumor boundaries are more difficult to delineate in regions with weak contrast, edema, infiltrative growth, or heterogeneous signal intensity (13,14,18). For clinical practice, this means that AI-generated masks in non-enhancing tumors should be reviewed carefully, particularly when segmentation output is being considered for volumetric monitoring, biopsy targeting, or radiotherapy planning.

Post-treatment lesions showed the weakest agreement, with a mean Dice coefficient of 0.73 ± 0.11. This was one of the most important subgroup findings because treated brain tumors often present the greatest radiological uncertainty. Postoperative gliosis, blood products, radiation necrosis, pseudoprogression, contrast leakage, edema, and residual or recurrent tumor may overlap in appearance, making segmentation difficult even for expert readers. The RANO framework was developed partly because conventional imaging assessment of treated gliomas is complex and cannot always be reduced to simple linear or volumetric measurements (6). Automated quantitative MRI assessment has shown promise in neuro-oncology response evaluation, but such approaches still require careful validation and expert clinical interpretation, especially in post-treatment cases (19). The lower post-treatment Dice coefficient in the present study supports the conclusion that AI-assisted segmentation should not be used as a stand-alone method in treated lesions without expert verification.

The qualitative error pattern further supports this interpretation. Most AI segmentation errors occurred at tumor margins, where the distinction between tumor, edema, gliosis, necrosis, and treatment effect was least clear. Some AI masks overestimated tumor extent by including surrounding edema or treatment-related signal abnormality, while others underestimated small non-enhancing components. These edge-related discrepancies are particularly important because a segmentation mask may have an apparently acceptable Dice score while still missing a clinically relevant tumor extension or including a region that should not be treated as active tumor. Expert-centered evaluations of segmentation algorithms have emphasized that conventional numerical metrics do not always capture clinical acceptability, particularly when errors occur in anatomically or therapeutically important areas (20). For this reason, Dice coefficient, sensitivity, and specificity should be interpreted together with visual review and clinical context rather than as isolated performance indicators.

The study has several strengths. It used a diagnostic accuracy design with expert radiologist consensus segmentation as the reference standard, blinded initial radiologist assessment, anonymized imaging data, and separate storage of AI and expert masks before comparison. These design features reduce incorporation bias and improve the credibility of the index test comparison. The use of sensitivity, specificity, predictive values, overall accuracy, Dice coefficient, and volume difference provides a broad assessment of both classification and spatial overlap performance. The inclusion of enhancing, non-enhancing, mixed-pattern, and post-treatment lesions also improves clinical relevance because routine neuro-oncology imaging includes a spectrum of lesion appearances rather than only ideal benchmark cases.

However, several limitations must be acknowledged. First, the study was conducted in a single tertiary care setting in Lahore, which limits generalizability to other hospitals, scanner platforms, MRI protocols, and patient populations. Second, the AI model was evaluated against expert imaging masks rather than histopathological validation of every segmented region; this is standard for segmentation studies but means that the reference standard reflects expert radiological interpretation rather than absolute biological tumor boundaries. Third, the study did not report confidence intervals or the underlying voxel-level, lesion-level, slice-level, or patient-level confusion matrix denominators; therefore, the precision of sensitivity, specificity, predictive values, and accuracy could not be independently assessed. Fourth, although subgroup Dice coefficients were clinically informative, smaller groups such as post-treatment lesions had limited sample size, and the subgroup findings should be interpreted cautiously. Fifth, workflow outcomes such as reporting time, radiologist workload reduction, inter-reader improvement after AI assistance, or treatment-planning impact were not formally measured. Therefore, the study supports AI as a segmentation-assistance tool but does not establish independent clinical implementation effectiveness.

Future studies should address these limitations through multicentre external validation across different Pakistani hospitals, scanner manufacturers, field strengths, acquisition protocols, and neuro-oncology referral patterns. They should report complete confusion matrices, confidence intervals, inter-reader agreement before consensus, and subgroup analyses by tumor type, enhancement pattern, treatment status, and lesion volume. Where possible, studies should also assess AI-assisted radiologist performance compared with unassisted radiologist segmentation, including time efficiency, correction burden, inter-reader consistency, and clinical acceptability. Integration of advanced MRI sequences, radiomics, molecular tumor information, and longitudinal follow-up may further improve segmentation reliability, particularly in non-enhancing and post-treatment lesions. Until such evidence is available, AI-assisted segmentation should be used as a supervised support tool that may improve initial tumor delineation but requires expert radiologist review before clinical interpretation.

CONCLUSION

AI-assisted MRI demonstrated promising diagnostic accuracy and good spatial agreement with expert radiologist consensus segmentation for brain tumor delineation in this Lahore-based clinical cohort. The strongest performance was observed in enhancing tumors, while non-enhancing and post-treatment lesions showed lower agreement because of indistinct margins, edema, gliosis, necrosis, and treatment-related imaging changes. These findings support the use of AI-generated masks as an initial segmentation aid that may improve consistency and assist radiology workflow, but they do not support unsupervised replacement of expert radiologist assessment. Larger multicentre Pakistani studies with complete confidence intervals, external validation, inter-reader agreement analysis, and formal workflow evaluation are needed before routine clinical integration of AI-assisted brain tumor segmentation.

REFERENCES

1. Pakistan Brain Tumour Consortium, Enam SA, Shah MM, Bajwa MH, Khalid MU, Bakhshi SK, et al. The Pakistan Brain Tumour Epidemiology Study. J Pak Med Assoc. 2022;72(11 Suppl 4). doi:10.47391/JPMA.11-S4-AKUB01.

2. Danish F, Salam H, Qureshi MA, Nouman M. Comparative clinical and epidemiological study of central nervous system tumours in Pakistan and global database. Interdiscip Neurosurg. 2021;25:101239. doi:10.1016/j.inat.2021.101239.

3. Bajwa MH, Khalid MU, Shah MM, Shamim MS, Laghari AA, Akhunzada NZ, et al. Treatment patterns of glioma in Pakistan: an epidemiological perspective. J Pak Med Assoc. 2022;72(11 Suppl 4). doi:10.47391/JPMA.11-S4-AKUB05.

4. Abdullah UE, Laghari AA, Khalid MU, Rashid HB, Jabbar AA, Mubarak F, et al. Current management of glioma in Pakistan. Glioma. 2019;2(4):139-144. doi:10.4103/glioma.glioma_15_19.

5. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021;23(8):1231-1251. doi:10.1093/neuonc/noab106.

6. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, et al. Updated response assessment criteria for high-grade gliomas: Response Assessment in Neuro-Oncology Working Group. J Clin Oncol. 2010;28(11):1963-1972. doi:10.1200/JCO.2009.26.3541.

7. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2015;34(10):1993-2024. doi:10.1109/TMI.2014.2377694.

8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017;4:170117. doi:10.1038/sdata.2017.117.

9. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Cham: Springer; 2015. p. 234-241. doi:10.1007/978-3-319-24574-4_28.

10. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, editors. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2016. Cham: Springer; 2016. p. 424-432. doi:10.1007/978-3-319-46723-8_49.

11. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision. Stanford: IEEE; 2016. p. 565-571. doi:10.1109/3DV.2016.79.

12. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017;36:61-78. doi:10.1016/j.media.2016.10.004.

13. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging. 2016;35(5):1240-1251. doi:10.1109/TMI.2016.2538465.

14. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with deep neural networks. Med Image Anal. 2017;35:18-31. doi:10.1016/j.media.2016.05.004.

15. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z.

16. Kofler F, Berger C, Waldmannstetter D, Lipkova J, Ezhov I, Tetteh G, et al. BraTS Toolkit: translating BraTS brain tumor segmentation algorithms into clinical and scientific practice. Front Neurosci. 2020;14:125. doi:10.3389/fnins.2020.00125.

17. Wang TW, Shiao YC, Hong JS, Lee WK, Hsu MS, Cheng HM, et al. Artificial Intelligence Detection and Segmentation Models: a systematic review and meta-analysis of brain tumors in magnetic resonance imaging. Mayo Clin Proc Digit Health. 2024;2(1):75-91. doi:10.1016/j.mcpdig.2024.01.002.

18. Di Ieva A, Russo C, Liu S, Jian A, Bai MY, Qian Y, et al. Application of deep learning for automatic segmentation of brain tumors on magnetic resonance imaging: a heuristic approach in the clinical scenario. Neuroradiology. 2021;63(8):1253-1262. doi:10.1007/s00234-021-02649-3.

19. Kickingereder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. Lancet Oncol. 2019;20(5):728-740. doi:10.1016/S1470-2045(19)30098-1.

20. Hoebel KV, Patel JB, Beers AL, Chang K, Singh P, Brown JM, et al. Expert-centered evaluation of deep learning algorithms for brain tumor segmentation. Radiol Artif Intell. 2024;6(1). doi:10.1148/ryai.220231.

21. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Moher D, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi:10.1136/bmj.h5527.

22. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell. 2020;2(2):e200029. doi:10.1148/ryai.2020200029.