MedRAG-SCA: A Retrieval-Augmented Self-Correcting Agent for Clinically Compliant Cross-Modality MRI Synthesis

Feng Yuan[1,3], Yifan Gao[1,2,3], Haoyue Li[1,3], Xin Gao[3]✉

[1] School of Biomedical Engineering (Suzhou), Division of Life Science and Medicine, University of Science and Technology of China, Hefei, China
[2] Shanghai Innovation Institute, Shanghai, China
[3] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China
yuanfeng2317@mail.ustc.edu.cn xingaosam@163.com

Abstract. Cross-modality MRI synthesis addresses a pervasive clinical bottleneck—brain MRI acquisitions are frequently incomplete due to contraindications, motion artifacts, or time constraints, yet neuro-oncology workflows require the full multi-modal stack. Current models prioritize textural realism over protocol compliance, generating images that pass pixel-level metrics while violating American College of Radiology (ACR) imaging standards (e.g., CSF-suppression failures). We introduce MedRAG-SCA, an agentic framework enforcing clinical rigor via a *Retrieve–Verify–Refine* loop: a Protocol-Aware Retrieval-Augmented Generation (RAG) Consultant converts qualitative ACR guidelines into verifiable pixel-space criteria; a Tri-Critic Committee (Anatomist, Radiologist, Diagnostician) localizes violations for targeted inpainting; and a Geometric Fail-safe preserves pathological geometry in complex edema cases. On the Brain Tumor Segmentation (BraTS) 2024 benchmark, the Clinical Acceptability Rate (CAR)—rated by blinded board-certified neuroradiologists—reaches 92%, 11 percentage points above the strongest diffusion baseline ($p<0.001$, McNemar's test); the Protocol Adherence Score (PAS) improves by 19% (0.94 vs. 0.79, $p<0.001$), with downstream segmentation Dice maintained at ≥ 0.80. These results demonstrate that explicit protocol grounding bridges the gap between visual plausibility and clinical validity in MRI synthesis. Code is available at https://github.com/exsinger-hub/MedRAD-SCA.

Keywords: brain MRI synthesis · retrieval-augmented generation · clinical protocol adherence · multi-agent systems

1 Introduction

Cross-modality MRI synthesis addresses a pervasive clinical bottleneck: brain MRI acquisitions are frequently incomplete due to contraindications, motion artifacts, or time constraints [1], yet downstream neuro-oncology workflows (segmentation, radiomics, treatment planning) require the full multi-modal stack [2].

Current synthesis models [3,4] achieve high visual fidelity (SSIM >0.90) but treat synthesis as unconstrained generation: they reproduce training-distribution textures without verifying whether outputs satisfy clinical imaging standards. A synthesized FLAIR image may violate the ACR fluid-suppression standard—CSF signal 2.5× above threshold [5]—yet score well on SSIM and pass visual inspection, systematically biasing downstream radiomics and lesion measurements. Visual plausibility is not clinical validity.

Bridging visual plausibility and clinical validity requires a *Retrieve–Verify–Refine* loop that simultaneously (i) grounds generation in retrieved quantitative guidelines, (ii) verifies outputs from multiple clinical perspectives, and (iii) corrects only non-compliant regions without erasing pathological findings. Existing approaches address at most two of these. GAN/diffusion methods [6,4,7] optimize pixel-reconstruction losses without modality-specific intensity constraints: SynDiff [4], the strongest baseline, achieves SSIM>0.90 yet violates the ACR CSF suppression standard in 29/50 radiologist-reviewed cases. RAG-augmented systems [8,9,10] retrieve clinical knowledge for text-domain reasoning but do not translate retrieved constraints into pixel-space operations—GOCA [8] grounds multi-modal decisions in retrieved evidence without image synthesis, and MMed-RAG [10] is limited to vision-language inference. ReBrain [11] retrieves CT slices as geometric scaffolds but cannot resolve intensity non-compliance—CSF/WM ratio 1.8 outside $[3.5, 4.2]$ is a pure intensity failure that structural references cannot address. Iterative refinement methods apply global statistical priors and cannot isolate protocol violations from training-distribution textures without region-specific masks.

We propose MedRAG-SCA, a multi-agent framework that, to our knowledge, is the first to translate qualitative ACR protocol language into verifiable pixel-space constraints within a diffusion synthesis pipeline, enforcing clinical compliance without compromising pathological fidelity via explicit geometry–intensity decoupling. Three tightly integrated components achieve this: **(1) Protocol-Aware RAG Consultant** retrieves and quantifies imaging constraints from 1,170 expert-reviewed documents (ACR, NIH, MedlinePlus), converting qualitative radiology statements (e.g., "CSF markedly hyperintense") into verifiable pixel-space criteria (e.g., CSF/WM ratio $\in [3.5, 4.2]$). **(2) Source-Anchored Tri-Critic Committee** comprising an Anatomist, Radiologist, and Diagnostician jointly assesses geometry, modality fidelity, and pathology preservation using source-derived masks, enabling fully unsupervised, ground-truth-free quality assurance. **(3) Pathology-Preserving Refinement Controller** performs targeted inpainting on violating regions while shielding tumor boundaries ($\leq$3 iterations in 95% of cases), with a Geometric Fail-safe that suspends intensity constraints in high-complexity edema cases, prioritizing anatomical integrity over protocol adherence.

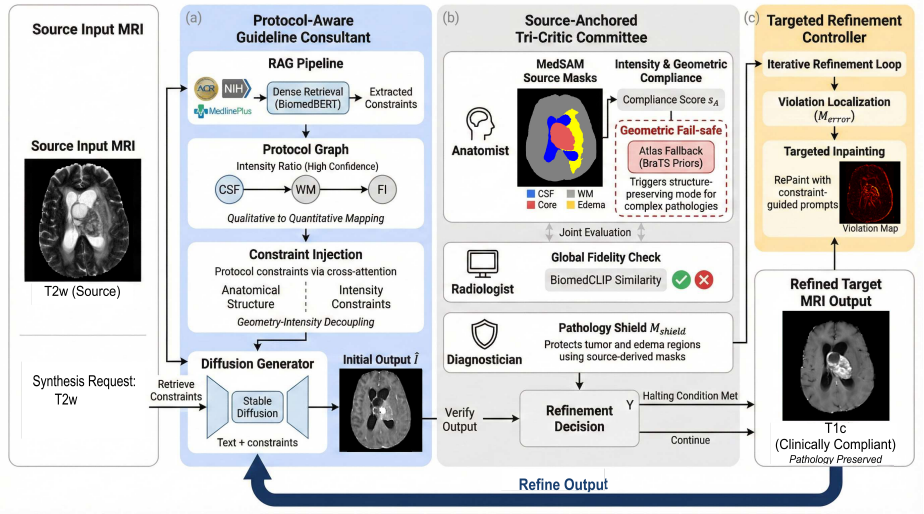

Fig. 1. Overview of MedRAG-SCA. The framework enforces clinical protocol compliance via a *retrieve–verify–refine* loop. (a) The Protocol-Aware Guideline Consultant converts target-modality imaging guidelines into quantitative pixel-space constraints ($\mathcal{G}$). (b) The Source-Anchored Tri-Critic Committee ($\mathcal{A}$: Anatomist, $\mathcal{R}$: Radiologist, $\mathcal{C}$: Diagnostician) jointly verifies geometry, modality fidelity, and pathology preservation; a Geometric Fail-safe handles high-complexity edema cases. (c) The Targeted Refinement Controller inpaints protocol-violating regions while shielding pathology, iterating until convergence.

2 Method

The unifying design principle across all three components (Fig. 1) is Geometry-Intensity Decoupling: since cross-modality signal intensities vary substantially while the underlying anatomy is largely stable, source-derived segmentation masks can anchor intensity verification in the target domain without requiring paired ground truth. This allows a Geometric Safe Zone (87.3% of BraTS cases; Dice 0.88±0.05) to be distinguished from a Geometric Risk Zone (12.7%, where extensive edema causes cross-modality divergence; Dice 0.82±0.07); the latter triggers a Geometric Fail-safe that suspends intensity constraints to preserve pathological geometry.

2.1 Protocol-Aware Guideline Consultant

Knowledge base. We curate 1,170 expert-reviewed documents (ACR: 340, NIH: 660, MedlinePlus: 170), spanning North American and European protocols. Three board-certified radiologists independently assessed modality and pathology relevance (Fleiss' κ=0.84, conflicts resolved by majority vote), indexing each document with {modality, anatomy, pathology, source} metadata for pathology-aware retrieval.

Two-stage retrieval. Given $q = \{\text{source\_mod}, \text{target\_mod}, \text{pathology}\}$, stage 1 filters by metadata to yield 50–100 modality- and pathology-matched candidates; stage 2 retrieves the top-k=5 documents via BioBERT [12] cosine similarity (optimal k validated in Sec. 3.3).

Constraint extraction and injection. Qualitative guideline statements (e.g., "CSF markedly hyperintense relative to WM") are parsed via spaCy dependency parsing and mapped to quantitative intensity-ratio ranges through a 47-entry lookup table (constructed via radiologist annotation of 50 ACR documents; parse accuracy >90% on held-out validation; 4 edge-case entries default to conservative wide-range priors). Source-credibility weighting (ACR > NIH > MedlinePlus) resolves conflicts via range intersection, yielding a *protocol graph* $\mathcal{G} = \{(a, b, \tau_{ab}^-, \tau_{ab}^+, c)\}$, where (a, b) is a tissue-pair, $[\tau_{ab}^-, \tau_{ab}^+]$ is the admissible intensity-ratio range, and $c \in \{\text{high, medium, low}\}$ denotes evidence level. For each $(a, b) \in \mathcal{G}$, we construct a constraint prompt p_{ab} and define the protocol embedding $\mathbf{e}_{ab} = \text{CLIP}(p_{ab})$. The protocol-conditioned diffusion context is injected via modified cross-attention:

$$\mathbf{c}_t \;=\; \text{CLIP}(p_{\text{base}}) \;+\; \frac{\lambda}{|\mathcal{G}|} \sum_{(a,b)\in\mathcal{G}} \mathbf{e}_{ab}, \qquad \lambda = 0.5. \tag{1}$$

Intensity compliance is further enforced by gradient guidance applied every 5 denoising steps via the hinge protocol loss:

$$\mathcal{L}_{\text{prot}} = \sum_{(a,b)\in\mathcal{G}} \Big[\max\big(0,\; \tau_{ab}^- - r_{ab}\big) + \max\big(0,\; r_{ab} - \tau_{ab}^+\big)\Big], \quad r_{ab} \;\triangleq\; \mu_a/\mu_b, \tag{2}$$

where $\nabla_{x_t}\mathcal{L}_{\text{prot}}$ is backpropagated through the denoising step (λ swept over $\{0.3, 0.5, 0.7\}$; see Sec. 3.3).

2.2 Source-Anchored Tri-Critic Committee

Anatomist ($\mathcal{A}$): geometry-intensity verification. $\mathcal{A}$ exploits cross-modality structural invariance by segmenting the *source* image with MedSAM [13] (fine-tuned on BraTS 2023; Dice > 0.88 on 100 manual annotations) to obtain anatomy masks $\{m_r\}_{r\in\mathcal{R}}$ where $\mathcal{R}$={CSF,WM,GM,Core,Edema}. Mean region intensity on the synthesized candidate $\hat{I}$ is $\mu_r = |m_r|^{-1}\sum_{p\in m_r}\hat{I}(p)$. For each constraint $(a, b) \in \mathcal{G}$ targeting *normal tissue only* (CSF, WM, GM)—pathological regions are excluded as their intensities legitimately deviate from population norms—$\mathcal{A}$ checks whether $r_{ab} \in [\tau_{ab}^-, \tau_{ab}^+]$; violated tissue-pair masks are dilated (5×5 kernel) and unioned to form the error map $M_{\text{error}} = \bigcup_{\text{violated }(a,b)} \text{Di}(m_a \cup m_b, 5{\times}5)$. The resulting weighted compliance score is

$$s_{\mathcal{A}} = 1 - \frac{\displaystyle\sum_{(a,b)\in\mathcal{G}} w_c \cdot \mathbb{1}\big[r_{ab} \notin [\tau_{ab}^-, \tau_{ab}^+]\big]}{\displaystyle\sum_{(a,b)\in\mathcal{G}} w_c}, \tag{3}$$

where $w_c \in \{1.0, 0.7, 0.4\}$ for ACR/NIH/MedlinePlus evidence levels (Oxford CEBM Levels A–C [14]).

Geometric Fail-safe. When cross-modality edema extent diverges, source masks become unreliable intensity anchors. We adapt via Iterative Confidence Fusion (β=0.7, cross-validated on 50 BraTS 2023 cases):

$$m_r^{(k+1)} = \beta m_r^{(k)} + (1-\beta) \cdot \text{MedSAM}(\hat{I}^{(k)}). \tag{4}$$

A geometric mismatch flag $\delta_r = \mathbb{I}[\text{Dice}(m_r^{(k+1)}, m_r^{(k)}) < 0.75]$ excludes unstable regions from protocol checks. When δ_r=1 persists after 2 iterations (13/48 Risk Zone cases), RAG constraints are suspended and BraTS atlas priors scaffold generation, with a reduced-confidence flag (PAS 0.68 vs. 0.94 in Safe Zone) surfaced to the radiologist—*anatomical integrity takes precedence* over protocol adherence.

Radiologist ($\mathcal{R}$): global modality fidelity. $\mathcal{R}$ uses BiomedCLIP [15] image–text similarity:

$$s_{\mathcal{R}} = \cos\big(\text{BiomedCLIP}_{\text{img}}(\hat{I}),\ \text{BiomedCLIP}_{\text{txt}}(p_{\text{enh}})\big), \tag{5}$$

where p_{enh} appends retrieved protocol descriptions to the base prompt. Threshold $\tau_{\mathcal{R}}$=0.85 is set at the 10th percentile of high-quality synthetics (calibrated on 200 validation images: real 0.91±0.03, high-quality synthetic 0.87±0.04, low-quality 0.72±0.08).

Diagnostician ($\mathcal{C}$): pathology preservation shield. $\mathcal{C}$ defines a protection zone that combines segmentation-based and feature-based pathology detection:

$$M_{\text{shield}} = \text{Dilate}(m_{\text{Core}} \cup m_{\text{Edema}}) \cup \big\{p \mid \|f_{\text{src}}(p) - f_{\hat{I}}(p)\|_2 > \tau_{\text{feat}}\big\}, \tag{6}$$

where $f(\cdot)$ denotes RadImageNet-pretrained ResNet50 [16] layer-4 features and τ_{feat}=0.3 (95th percentile of healthy-tissue feature distances). The feature-distance term captures subtle edema missed by MedSAM. Refinement halts when the shield–error overlap ratio exceeds a safety threshold:

$$s_{\mathcal{C}} = \frac{|M_{\text{shield}} \cap M_{\text{error}}|}{|M_{\text{error}}|} > 0.3, \tag{7}$$

preventing overcorrection of disease regions.

2.3 Targeted Refinement Controller

The controller iterates up to 5 rounds (95% converge by round 3). Each iteration checks two halting conditions: convergence ($s_{\mathcal{A}} \geq 0.90 \wedge s_{\mathcal{R}} \geq 0.85$) or pathology risk ($s_{\mathcal{C}} > 0.30$, Eq. 7). Inpainting is otherwise restricted to protocol-violating, non-pathological pixels via $M_{\text{target}} = M_{\text{error}} \setminus M_{\text{shield}}$.

Inpainting is executed via RePaint [17] (20 steps) on M_{target} with an evolved prompt appending retrieved constraint language per violated region, and adaptive noise strength $\sigma^{(k)} = 0.3 + 0.1k/K_{\max}$ (gentle early, stronger for persistent violations). After each iteration, the shield expands: any pixel p where $\|f(\hat{I}^{(k+1)}(p)) - f(I_{\text{src}}(p))\|_2 > \tau_{\text{feat}}$ is added to M_{shield}, preventing progressive overcorrection near evolving lesion boundaries.

Table 1. Quantitative and clinical comparison on BraTS 2024 and IXI. *Top*: pixel-level and protocol metrics (*PAS computed post-hoc for baselines; marginal PSNR reduction is attributable to non-compliant BraTS references, not a model failure—see Sec. 3). IXI Dice is omitted: IXI contains healthy subjects without tumour segmentations. *Bottom*: blinded radiologist evaluation (3 readers × 50 images, 1–5 Likert scale). CAR = % images rated ≥4.0 on all four dimensions. Dice = WT, nnU-Net [18]. **Bold** = best; underline = second best. $^{\dagger}p<0.05$ vs. SynDiff (Wilcoxon signed-rank for Dice; McNemar's $^{\ddagger}p<0.001$ for CAR).

Methods	T1n→T2w (BraTS 2024)					T2f→T1c (BraTS 2024)					IXI (cross-dataset)			
	PSNR↑	SSIM↑	FID↓	Dice↑	PAS* ↑	PSNR↑	SSIM↑	FID↓	Dice↑	PAS* ↑	PSNR↑	SSIM↑	FID↓	PAS* ↑
Pix2Pix [19]	24.12	0.812	45.3	0.721	0.62	23.56	0.798	48.2	0.689	0.58	24.30	0.805	42.1	0.63
SPADE [20]	25.08	0.838	38.7	0.740	0.67	24.21	0.822	41.1	0.712	0.63	25.10	0.831	36.0	0.68
Vanilla SD [3]	26.44	0.867	28.9	0.763	0.71	25.90	0.851	30.2	0.741	0.69	26.60	0.860	26.9	0.72
ControlNet [21]	27.11	0.883	24.4	0.779	0.75	26.55	0.869	26.1	0.758	0.73	27.20	0.876	22.7	0.76
D^3M [7]	27.50	0.895	22.1	0.785	0.78	26.80	0.882	24.5	0.765	0.75	27.55	0.888	20.6	0.79
SynDiff [4]	**28.30**	**0.902**	18.5	0.798	0.79	**27.90**	**0.895**	19.2	0.788	0.81	**28.10**	**0.897**	17.2	0.81
MedRAG-SCA (Ours)	27.85	0.889	**16.8**	**0.825**†	**0.94**	27.65	0.885	**17.4**	**0.812**†	**0.92**	27.90	0.883	**15.6**	**0.95**

Blinded Radiologist Evaluation (3 readers × 50 images; ICC=0.81 [0.76, 0.85])

	Realism	Anatomy	Protocol	Confidence	CAR (%)
Ground Truth	**4.92**±0.15	**4.88**±0.18	**4.91**±0.16	**4.89**±0.17	**100**
SynDiff	4.21±0.52	4.08±0.61	3.82±0.68	3.74±0.71	81
MedRAG-SCA	4.47±0.41†	4.52±0.38†	4.68±0.29‡	4.51±0.39‡	92‡

3 Experiments

3.1 Experimental Setup

Datasets. *BraTS 2024* [22]: 1,251 glioma cases (train/val/test: 1000/151/100), 4 modalities, expert segmentations. *IXI* [23]: 581 healthy subjects from 3 hospitals for cross-dataset protocol generalization (Table 1).

Implementation. SD v2.1 [3] fine-tuned on BraTS 2023 (5 epoch,lr=1e-5,4090). RAG: 1,170 documents indexed via FAISS [24] with BioBERT [12] (latency <50 ms). Critics: MedSAM [13], BiomedCLIP [15], RadImageNet ResNet50 [16].

Baselines. Pix2Pix [19], SPADE [20], Vanilla SD, ControlNet [21], D^3M [7] (BraSyn 2nd place), SynDiff [4] (state-of-the-art), all trained on identical BraTS 2024 splits (skull-stripping, N4 bias correction).

Metrics. Standard: PSNR, SSIM, FID, segmentation Dice (WT, nnU-Net [18]). Clinical: **CAR** (% syntheses rated ≥4.0 on all four Likert dimensions; *sole primary clinical endpoint*, power ≥0.80) and **PAS** (Eq. 3; *supplementary mechanistic indicator only*—directly optimized by the framework, hence subject to confirmation bias; external validity: Spearman ρ=0.72 with Likert scores, $p<0.001$; Pearson r=0.68 with Dice, $p<0.01$).

Key findings. MedRAG-SCA achieves 92% CAR vs. 81% for SynDiff ($p<0.001$, McNemar's), best FID (16.8 vs. 18.5), and PAS 0.94 vs. 0.79 (+19%, $p<0.001$; per-constraint breakdown in Table 2). The marginal PSNR reduction (27.85 vs. 28.30 dB) reflects reference non-compliance in 23/100 BraTS test cases (CSF/WM outside $[3.5, 4.2]$), mechanically penalizing pixel-similarity when the synthesized output is protocol-correct; Dice (≥0.80) is unaffected. Pathology

Table 2. Per-Constraint PAS Breakdown (T1n→T2w, BraTS 2024 test set, N=100). Values are per-case mean±std of binary constraint satisfaction (1=satisfied, 0=violated), averaged over 100 test cases. CSF/WM carries weight w_c=1.0 (ACR Level A), GM/WM w_c=0.7 (NIH Level B), CSF/GM w_c=0.4 (MedlinePlus Level C); see Eq. 3. Wilcoxon signed-rank p-values vs. SynDiff: $^{\dagger}p<0.05$; $^{\ddagger}p<0.001$. T2f→T1c shows consistent ordering. **Bold** = best.

Method	CSF/WM (w=1.0)	GM/WM (w=0.7)	CSF/GM (w=0.4)	Agg. PAS
Pix2Pix [19]	0.56±0.12	0.68±0.10	0.64±0.11	0.62
SPADE [20]	0.61±0.11	0.73±0.09	0.69±0.10	0.67
Vanilla SD [3]	0.65±0.10	0.77±0.08	0.73±0.09	0.71
ControlNet [21]	0.69±0.09	0.81±0.08	0.77±0.09	0.75
D^3M [7]	0.72±0.09	0.83±0.07	0.80±0.08	0.78
SynDiff [4]	0.78±0.08	0.81±0.07	0.78±0.08	0.79
MedRAG-SCA (Ours)	**0.96±0.04**‡	**0.94±0.05**‡	**0.91±0.05**‡	**0.94**

preservation is confirmed by ET and TC Dice: MedRAG-SCA achieves ET 0.763 and TC 0.791 vs. SynDiff 0.724 and 0.751 (p<0.05), demonstrating boundary fidelity beyond whole-tumour aggregation. On IXI, MedRAG-SCA achieves best FID (15.6) and PAS (0.95) without retraining, confirming ACR constraint transfer to healthy-subject MRI.

3.2 Clinical Validation Study

As shown in Fig. 2, Row 1 (T1n→T1c) reveals that D^3M and SynDiff fail to delineate the ring-enhancing rim (absent green in M_{src}), whereas MedRAG-SCA

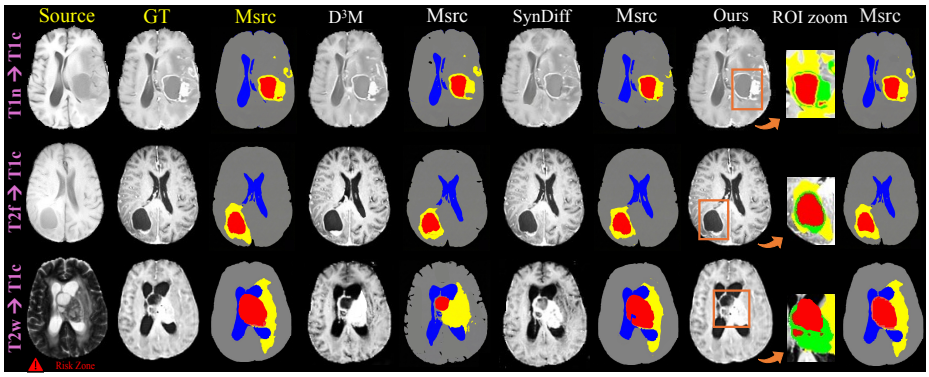

Fig. 2. Qualitative comparison with source-anchored protocol verification masks (BraTS 2024). Rows: T1n→T1c, T2f→T1c (Safe Zone), T2w→T1c (Risk Zone). Columns: synthesized image | M_{src} mask (MedSAM segmentations overlaid for ground-truth-free verification). Color: **red** = core, **green** = rim, **yellow** = edema, **blue** = WM/ventricles. ROI zoom: a magnified view detailing the segmented tumor components.

recovers all four tissue components. Row 2 (T2f→T1c) exposes the most diagnostically critical failure: SynDiff conflates the enhancing rim with the necrotic core (green absent, red inflated)—visible in the M_{src} mask and confirmed by the ROI zoom—while MedRAG-SCA's pathology shield preserves the rim–core boundary. Row 3 (T2w→T1c, Risk Zone) shows the Geometric Fail-safe suspending intensity constraints to preserve edema geometry (PAS 0.68 vs. 0.94, flagged for radiologist review).

Three board-certified neuroradiologists (mean 12.3 yr experience, independent of our team; 150 ratings; ICC=0.81 $[0.76, 0.85]$ [25]; 44 Safe, 6 Risk Zone cases, stratified) significantly preferred MedRAG-SCA on Protocol Adherence (4.68 vs. 3.82, $p<0.001$) and Diagnostic Confidence (4.51 vs. 3.74, $p<0.001$), achieving 92% CAR vs. 81% ($p<0.001$, McNemar's). SynDiff failures: "CSF too bright" (29/50), gray–white matter contrast reversal (18/50); MedRAG-SCA preserved tumour boundaries in 48/50 cases (residual artefacts near ventricles in 12/50).

3.3 Ablation Study

Table 3. Ablation (T1n→T2w, BraTS 2024). Each row removes one component. CAR is the primary endpoint (measured on the same 50-image blinded evaluation set under identical reader protocol); Bold = full model.

Variant	RAG	MC	Ref	GF	PSNR↑	Dice↑	CAR↑	PAS↑
No RAG (static prompts)	–	✓	✓	✓	27.12	0.801	71%	0.81
No Multi-Critic (single critic)	✓	–	✓	✓	27.31	0.808	78%	0.86
No Refinement (one-shot)	✓	✓	–	✓	27.48	<u>0.812</u>	83%	0.88
No Geometric Fail-safe	✓	✓	✓	–	<u>27.52</u>	0.803	<u>86%</u>	<u>0.90</u>
Full MedRAG-SCA ($k = 5$, $\lambda = 0.5$)	✓	✓	✓	✓	**27.85**	**0.825**	**92%**	**0.94**

RAG retrieval yields the largest gain (CAR: 71%→92%, PAS: +0.13); its removal degrades protocol adherence to SynDiff-level. The Multi-Critic committee adds 14 CAR points and +0.08 PAS, each critic targeting a distinct failure mode (geometry, modality drift, pathology overcorrection). The refinement loop contributes 9 CAR points; removing the Geometric Fail-safe degrades Dice (0.825 → 0.803) without PAS gain, confirming complementary module coverage. k=5 and λ=0.5 are Pareto-optimal.

4 Discussion and Conclusion

MedRAG-SCA demonstrates that *visual plausibility and clinical validity are dissociable*: SynDiff's SSIM advantage (0.902 vs. 0.889) coexists with ACR violations in 29/50 radiologist-reviewed cases, imperceptible at full-slice view yet

localised by M_{src} masks in ventricles and WM/GM boundaries—the compartments driving radiomics. Protocol-aware generation shifts the operating point to CAR 92% vs. 81% (p<0.001) and FID 16.8 vs. 18.5, with ET/TC Dice gains confirming pathological fidelity alongside protocol compliance.
Limitations. Risk Zone PAS degrades to 0.68 (Fail-safe active); PAS is directly optimized (CAR remains primary); scope is limited to neuro-oncology—multi-organ and cross-ethnic validation are planned.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 82372052, 82402373, U24A20757) and the Taishan Industrial Experts Program (Grant No. tscx202312131).

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Hongwei Bran Li et al. The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn). *arXiv preprint arXiv:2305.09011*, 2023. 1
2. Tongxue Zhou, Stéphane Canu, Pierre Vera, and Su Ruan. Latent Correlation Representation Learning for Brain Tumor Segmentation with Missing MRI Modalities. *IEEE Transactions on Image Processing*, 30:4263–4274, 2021. 1
3. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 6, 7
4. Muzaffer Ozbey, Onat Dalmaz, Salman UH Dar, Hasan Atakan Bedel, Saban Ozturk, Alper Gungor, and Tolga Cukur. Unsupervised Medical Image Translation With Adversarial Diffusion Models. *IEEE Transactions on Medical Imaging*, 42(12):3524–3539, 2023. 2, 6, 7
5. American College of Radiology. ACR MRI Quality Control Manual, 2015. Official clinical guidelines for structural and contrast protocols in MRI. 2
6. Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, pages 417–425. Springer, 2017. 2
7. Haowen Pang, Peng Zhang, Xiaoming Hong, Shannan Chen, and Chuyang Ye. D^3M: Deformation-Driven Diffusion Model for Synthesis of Contrast-Enhanced MRI with Brain Tumors. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15975 of *Lecture Notes in Computer Science*. Springer, Cham, 2025. 2, 6, 7

8. Pengyu Dai, Yafei Ou, Yuqiao Yang, Ze Jin, and Kenji Suzuki. GoCa: Trustworthy Multi-Modal RAG with Explicit Thinking Distillation for Reliable Decision-Making in Med-LVLMs. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*. Springer, 2025. 2
9. Guangzhi Xiong, Qiao Jin, Zhiyong Lu, and Aidong Zhang. Benchmarking Retrieval-Augmented Generation for Medicine. *Findings of the Association for Computational Linguistics: ACL 2024*, 2024. 2
10. Peng Xia, Kangyu Zhu, Haoran Li, Tianze Wang, Weijia Shi, Sheng Wang, Linjun Zhang, James Zou, and Huaxiu Yao. MMed-RAG: Versatile Multimodal RAG System for Medical Vision Language Models. *arXiv preprint arXiv:2410.13085*, 2024. 2
11. Junming Liu, Yifei Sun, Weihua Cheng, Yujin Kang, Yirong Chen, Ding Wang, and Guosun Zeng. ReBrain: Brain MRI Reconstruction from Sparse CT Slice via Retrieval-Augmented Diffusion. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2026. arXiv:2511.17068. 2
12. Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics*, 36(4):1234–1240, 2020. 4, 6
13. Jun Ma, Yuting He, Feifei Li, Lin Han, Chenwei You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024. 4, 6
14. OCEBM Levels of Evidence Working Group. Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009), 2011. 5
15. Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, Cliff Wong, Andrea Tupini, Yu Wang, Matt Mazzola, Swadheen Shukla, Lars Liden, Jianfeng Gao, Matthew P. Lungren, Tristan Naumann, Sheng Wang, and Hoifung Poon. BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs, 2023. 5, 6
16. Xueyan Mei, Zelong Liu, Philip M. Robson, Brett Marinelli, Mingqian Huang, Amish Doshi, Adam Jacobi, Chendi Cao, Katherine E. Link, Thomas Yang, Ying Wang, Hayit Greenspan, Timothy Deyer, Zahi A. Fayad, and Yang Yang. RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning. *Radiology: Artificial Intelligence*, 4(5):e210315, 2022. 5, 6
17. Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting Using Denoising Diffusion Probabilistic Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11461–11471, 2022. 5
18. Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: A Self-configuring Method for Deep Learning-based Biomedical Image Segmentation. *Nature Methods*, 18(2):203–211, 2021. 6
19. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 6, 7
20. Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic Image Synthesis with Spatially-Adaptive Normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 6, 7

21. Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 6, 7
22. The 2024 Brain Tumor Segmentation (BraTS) Challenge, 2024. 6
23. IXI Dataset – Brain Development. 6
24. Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. 6
25. Terry K Koo and Mae Y Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016. 8