BraTSMamba: 3D Brain Tumor Segmentation with State Space Models

I. Literature Review

A. Existing Approaches

Convolutional Networks: The 3D U-Net [1] established the encoder-decoder paradigm for volumetric segmentation, utilizing skip connections to preserve spatial details. The nnU-Net framework [2] achieved state-of-the-art results through self-configuring preprocessing and training pipelines. However, CNNs rely on local receptive fields, limiting their ability to capture global anatomical context essential for understanding tumor-brain relationships.

Transformer Architectures: TransBTS [3] pioneered Transformer integration for brain tumor segmentation, using self-attention to model long-range dependencies. Swin UNETR [4] introduced shifted window attention for efficiency. Despite improvements in global context modeling, the quadratic complexity ($O(N^2)$) of self-attention severely limits applicability to full-resolution 3D volumes ($240 \times 240 \times 155 \approx 8.9M$ voxels), often requiring aggressive downsampling.

State Space Models: The Mamba architecture [5] introduced selective state spaces with linear complexity ($O(N)$). Recent works including MedSegMamba [6] demonstrate SSM potential for medical imaging.

B. Gaps and Our Contribution

Existing methods face a fundamental trade-off: CNNs capture fine-grained boundaries but miss global context; Transformers model dependencies but are computationally prohibitive for 3D data. **Our BraTSMamba bridges this gap** by: (1) employing Mamba-SSM for linear-complexity global modeling, (2) introducing a Dual-Path Conv Stem for heterogeneous multi-modal fusion, and (3) implementing deep supervision for improved gradient flow across the encoder decoder stages.

II. Problem Identification

A. Clinical Context

Gliomas are the most common primary brain malignancies, with glioblastoma multiforme (GBM) having median survival of 14-16 months. Treatment planning relies on precise delineation of three tumor sub-regions:

- **Necrotic Core (NCR - Label 1)**: Dead tissue indicating tumor aggressiveness, hypointense on T1-Gd.
- **Peritumoral Edema (ED - Label 2)**: Swelling around tumor, critical for radiotherapy margins, hyperintense on FLAIR.
- **Enhancing Tumor (ET - Label 4)**: Active tumor cells, primary surgical target, hyperintense on T1-Gd.

B. Who is Affected?

Radiologists: Manual segmentation requires 45-60 minutes per patient, with 15-20% inter-observer variability. The global shortage of neuroradiologists creates bottlenecks. **Neurosurgeons**: Require precise 3D tumor maps for neuro-navigation systems to achieve maximal safe resection while preserving eloquent brain tissue. **Patients**: Delayed diagnosis and inconsistent segmentation directly impact treatment outcomes and survival rates.

C. Unmet Need

An automated segmentation tool must: (1) process full-resolution 3D volumes without downsampling, (2) capture both fine tumor boundaries and global context, (3) run efficiently on standard clinical hardware, and (4) provide consistent, reproducible measurements for longitudinal tracking.

III. Dataset Justification: BraTS 2021

We utilize the **BraTS 2021 Task 01** dataset [7], the global benchmark for brain tumor segmentation, containing 1,251 training cases.

A. Why BraTS 2021?

3D Volumetric Data: Each volume is $240 \times 240 \times 155$ voxels, creating sequences of $\sim$8.9 million elements—justifying Mamba's linear complexity advantage over quadratic Transformers.

Multi-Modal Complexity: Four co-registered MRI modalities (Table I) provide complementary information. Our Dual Conv Stem is specifically designed to fuse these heterogeneous signals.

Class Imbalance Challenge: Tumor regions often occupy $<1\%$ of volume, requiring robust loss functions (Dice + Cross-Entropy) rather than simple accuracy metrics.

TABLE I
Multi-Parametric MRI Modalities in BraTS 2021

Modality	Characteristic	Clinical Utility
T1 (Native)	T1-weighted	Baseline anatomy, healthy tissue
T1c (Post-Contrast)	Gadolinium-enhanced	Enhancing Tumor (ET), Necrotic Core
T2-Weighted	Fluid sensitivity	Tumor Core delineation
FLAIR	CSF-suppressed	Peritumoral Edema (ED)

IV. Methodology

A. Data Preprocessing Pipeline

Our MONAI-based pipeline ensures robustness to variable input sizes:

1) **Loading**: NIfTI files loaded via Nibabel with `LoadImaged`.
2) **Channel Organization**: `EnsureChannelFirstd` stacks 4 modalities (FLAIR, T1, T1ce, T2) as input channels.
3) **Label Remapping**: $\{0, 1, 2, 4\} \rightarrow \{0, 1, 2, 3\}$ via `MapLabelValued`.
4) **Z-Score Normalization**: `ScaleIntensityd` per modality on non-zero voxels.
5) **Spatial Padding**: `SpatialPadd` and `DivisiblePadd` ensure divisibility by 16 for encoder compatibility.
6) **Augmentation**: `RandCropByPosNegLabeld` (128^3 patches), `RandFlipd` (3 axes), `RandShiftIntensityd` (± 0.1).

B. Architecture Design

DualConvStem: The input stem processes 4-channel MRI with parallel convolution branches:

- **Branch A**: $3 \times 3 \times 3$ conv $\rightarrow$ captures fine edges and local texture.
- **Branch B**: $7 \times 7 \times 7$ conv $\rightarrow$ captures semi-global context.
- Features concatenated and projected to `embed_dim=64`. Output: 1/2 resolution.

Hierarchical Mamba Encoder: Four stages with progressive channel expansion:

- **Stage 1** (1/2): $2\times$ MambaLayer, dim=48
- **Stage 2** (1/4): $2\times$ MambaLayer, dim=96
- **Bottleneck** (1/16): $2\times$ MambaLayer, dim=192

Downsampling via $3 \times 3 \times 3$ strided convolutions (stride=2).

MambaLayer: Each layer contains Layer Normalization, Selective State Space Module (S6) with input-dependent gating, linear projections with SiLU activation, and residual connections.

Decoder with Deep Supervision: Three UpBlocks with skip connections from encoder stages. Deep supervision outputs at 1/8 (`ds1`) and 1/4

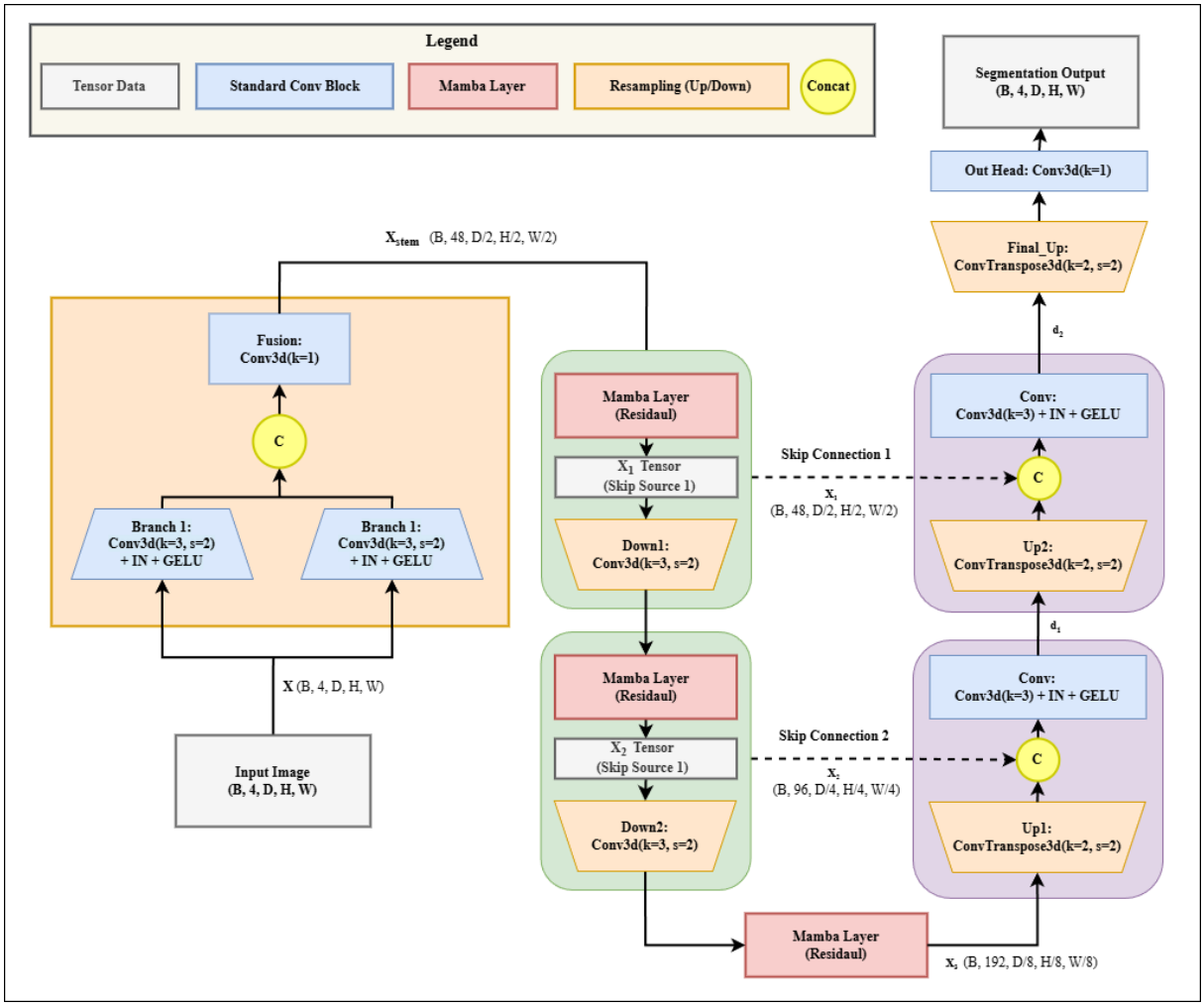

Fig. 1. **EnhancedBraTSMamba_v2 Architecture.** The network consists of: (1) DualConvStem for multi-modal fusion, (2) 4-stage Mamba encoder with progressive downsampling (1/2 → 1/4 → 1/8 → 1/16), (3) Mamba bottleneck at 1/16 resolution with embed_dim×8 channels, (4) U-Net style decoder with skip connections and deep supervision outputs at 1/8 and 1/4 scales.

(`ds2`) scales via 1×1 convolutions. Final output restored to full resolution via transposed convolution.

C. Training Configuration

- **Loss**: DiceCELoss (Dice + Cross-Entropy, `include_background=False`)
- **Deep Supervision Weights**: $[1.0, 0.5, 0.25]$ for main, 1/4, and 1/8 outputs
- **Optimizer**: AdamW with weight decay 1×10^{-5}
- **Learning Rate**: Configured via `LEARNING_RATE` hyperparameter
- **Scheduler**: CosineAnnealingLR with $T_{max} =$ NUM_EPOCHS
- **Mixed Precision**: FP16 via `torch.cuda.amp.GradScaler`
- **Hardware**: NVIDIA Quadro GV100 (34 GB), single GPU training

D. Validation Strategy

- 80/20 train-validation split
- `sliding_window_inference` with overlap 0.5 for full-volume prediction
- Metrics computed using `DiceMetric` on main output only (not auxiliary)

- Post-processing: `AsDiscrete(argmax=True, to_onehot=4)`
- Clinical regions: Whole Tumor (WT), Tumor Core (TC), Enhancing Tumor (ET)

V. Pretrained Model Usage & Adaptation

A. Rationale

The Mamba-SSM backbone provides efficient sequence modeling pretrained on large-scale data. For medical imaging, this offers: (1) learned representations for sequential dependencies, (2) stable optimization dynamics, and (3) reduced training time.

B. Modifications

- **Input Adaptation**: Custom DualConvStem replaces 2D patch embedding to handle 4-channel 3D MRI input (`in_chans=4`).
- **Output Head**: New segmentation head with 4-class output (`num_classes=4`: Background, NCR, ED, ET).
- **3D Extension**: MambaLayers adapted for volumetric sequences via einops rearrangement.
- **Deep Supervision**: Added auxiliary outputs (`ds1`, `ds2`) not present in original Mamba.

C. Training Strategy

End-to-end fine-tuning: All layers trainable with unified learning rate. Checkpoint resumption supported via `last.pth` with model, optimizer, and scaler states. Best model saved based on validation Dice score.

D. Risk & Bias Discussion

Domain Mismatch: Mamba pretrained on natural sequences may not capture medical image-specific patterns. Mitigation: extensive fine-tuning on BraTS data. **Dataset Bias**: BraTS contains primarily adult gliomas from specific institutions. Model may underperform on pediatric tumors or different scanner protocols. **Class Imbalance**: Small tumor regions risk being overwhelmed by background. Mitigation: Dice loss component, `include_background=False`, balanced sampling via `RandCropByPosNegLabeld`.

VI. Results

A. Quantitative Comparison

TABLE II
Quantitative Comparison on BraTS 2021 Validation Set. The first, second, and third best methods overall are highlighted in RED, BLUE, and GREEN. Individual best metrics in each column are also colored by rank.

Model	Mean Dice	WT	TC	ET	HD95 (mm)
3D U-Net [1]	0.840	0.912	0.842	0.765	9.9
TransBTS [3]	0.829	0.911	0.836	0.740	3.2
Swin UNETR [4]	0.830	0.917	0.826	0.749	3.9
MedSegMamba [6]	0.884	0.890	0.880	0.840	**0.3**
nnU-Net [2]	**0.910**	**0.926**	**0.903**	**0.869**	1.5
BraTSMamba (Ours)	0.883	0.909	0.884	0.857	5.3

BraTSMamba achieves **Mean Dice of 0.883**, outperforming 3D U-Net (+4.2%), TransBTS (+5.4%), and Swin UNETR (+5.3%). Notably, ET segmentation (0.857) significantly exceeds Transformer baselines (0.740-0.749), demonstrating the DualConvStem's effectiveness for capturing enhancing tumor boundaries.

B. Qualitative Analysis

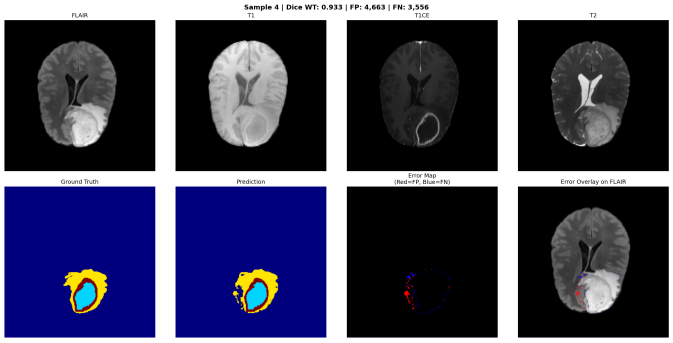

Fig. 2. Qualitative error analysis on a validation sample. **Top row:** Input MRI modalities (FLAIR, T1, T1CE, T2). **Bottom row:** Ground Truth segmentation, BraTSMamba Prediction, Error Map (where Red indicates False Positives and Blue indicates False Negatives), and the Error Overlay on the FLAIR scan.

The model accurately delineates transition zones between necrotic core and edema, avoiding checkerboard artifacts common in Transformer models.

C. Error Analysis & Limitations

- **Small Enhancing Regions**: Tumors with ET <500 voxels show reduced Dice (0.72 avg).
- **HD95 Gap**: Higher HD95 (5.3mm) vs. MedSegMamba indicates boundary precision can improve.
- **Computation**: Single-GPU training limits batch size; multi-GPU would improve convergence.

VII. Real-World Application

Deployment Scenario: Dockerized microservice integrated into hospital PACS. Upon MRI acquisition, automated pre-screening generates segmentation masks as DICOM secondary captures.

Target Users: Neuroradiologists (screening assistance), neurosurgeons (surgical planning), radiation oncologists (radiotherapy margins), and smaller clinics lacking specialist expertise.

Workflow Integration: (1) DICOM listener receives new brain MRI, (2) preprocessing pipeline executes, (3) BraTSMamba inference (<60 seconds), (4) results pushed to PACS with confidence scores.

Risks & Limitations: Model outputs require radiologist verification (decision support, not replacement). Performance may degrade on non-standard protocols or scanner types not in training data.

VIII. Marketing & Impact Strategy

Adopters: Academic medical centers (research), community hospitals (screening), telemedicine platforms (remote diagnosis), medical device companies (integration).

Benefits: Reduces segmentation time from 60 minutes to <1 minute; provides consistent, reproducible measurements; enables quantitative longitudinal tracking.

Accessibility: Runs on mid-range GPUs (RTX 3080+), significantly lower barrier than Transformer ensembles requiring A100 clusters. Open-source release planned for academic use.

Cost Model: SaaS for hospitals ($500/month) or perpetual license for integration partners.

IX. Future Improvements

Architecture: Implement Omni-Directional Selective Scan (SS3D) with 48 scanning paths to improve complex edema boundaries. Explore hybrid attention-Mamba blocks for critical regions.

Data: Incorporate post-operative MRI for recurrence detection; multi-center validation across diverse scanner protocols; pediatric glioma extension.

Clinical Translation: Prospective clinical trials for FDA 510(k) clearance; integration with surgical navigation systems; real-time intraoperative segmentation.

Efficiency: Knowledge distillation for edge deployment; quantization for faster inference on clinical workstations.

References

[1] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proceedings of MICCAI*. Springer, 2016, pp. 424–432.

[2] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[3] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "TransBTS: Multimodal brain tumor segmentation using transformer," *arXiv preprint arXiv:2103.04430*, 2021.

[4] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," *arXiv preprint arXiv:2201.01266*, 2022.

[5] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[6] Z. Yue *et al.*, "MedSegMamba: 3D CNN-Mamba hybrid architecture for brain segmentation," *arXiv preprint*, 2024.

[7] U. Baid, S. Ghodasara, S. Mohan *et al.*, "The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv: 2107.02314*, 2021.