Federated Continual Learning for MRI Brain Tumor Segmentation

Team 314IV

Ismoil Salohiddinov, Komiljon Qosimov, Abdurrashid Djumabaev

Computer Vision Module - Final Project

December 2025

Abstract

Brain tumor segmentation from MRI scans is critical for diagnosis and treatment planning. However, training robust deep learning models faces two challenges: (1) medical data cannot be centralized due to privacy regulations, and (2) data distributions shift over time. We propose a Federated Continual Learning (FCL) framework combining Flower-based federated learning with drift-aware adapters in a SegResNet architecture. Our approach enables collaborative training across 4 simulated hospital clients without sharing patient data, while mitigating catastrophic forgetting. Evaluated on a curated subset of the BraTS2021 dataset (600 patients), our method achieves a mean Dice score of **82.38%** (WT: 87.12%, TC: 82.84%, ET: 77.18%) with acceptable forgetting (backward transfer: -3.1%). Training was performed on a single RTX 5070 GPU (12GB) over approximately 80 hours.

1. Introduction

1.1 Problem Statement

Brain tumors, particularly glioblastomas, are among the most aggressive cancers with a median survival of 14-16 months. Accurate segmentation of tumor regions from MRI scans is essential for surgical planning, radiation therapy targeting, and treatment response monitoring.

1.2 Challenges

Challenge	Description
Data Privacy	HIPAA/GDPR prohibit centralizing patient data across institutions
Data Heterogeneity	Different hospitals use varying scanners and protocols
Catastrophic Forgetting	Models fine-tuned on new data forget previous patterns
Class Imbalance	Tumor regions occupy less than 5% of brain volume
3D Complexity	Volumes are 224×224×144 voxels with 4 modalities

1.3 Dataset

We use the BraTS2021 dataset from the RSNA-ASNR-MICCAI Brain Tumor Segmentation Challenge:

Property	Value
Total Patients	600 (480 train, 60 val, 60 test)
Modalities	T1, T1-contrast, T2, FLAIR
Volume Size	224 × 224 × 144 voxels
Classes	3 (Tumor Core, Whole Tumor, Enhancing Tumor)
Federated Split	4 hospitals, 120 patients each

2. Methodology

2.1 Model Architecture

We employ a 3D SegResNet with drift-aware adapters:

- **Backbone:** SegResNet with residual connections
- **Encoder blocks:** [1, 2, 2, 4]
- **Decoder blocks:** [1, 1, 1]
- **Initial filters:** 16
- **Adapters:** Lightweight modules (2% of parameters) for domain adaptation

Model Statistics

Total Parameters	22.5M
Trainable Parameters	22.1M
Adapter Parameters	0.4M (~2%)
Model Size	86.2 MB

2.2 Federated Learning Strategy

- **Framework:** Flower (flwr)
- **Aggregation:** FedAvg with equal client weighting
- **Rounds:** 200 (early stopping at 185)
- **Local Epochs:** 3 per round
- **Clients:** 4 virtual hospitals

2.3 Training Configuration

Parameter	**Value**
GPU	NVIDIA RTX 5070 (12GB VRAM)
Batch Size	2
Learning Rate	1e-4
Optimizer	Adam
Loss Function	Dice Loss
Mixed Precision	Enabled (FP16)
Training Time	~80 hours (early stopping at round 185)

3. Results

3.1 Segmentation Performance

Metric	Tumor Core (TC)	Whole Tumor (WT)	Enhancing Tumor (ET)	Mean
Dice Score	82.84% ± 6.98%	87.12% ± 4.23%	77.18% ± 10.24%	**82.38%**
IoU (Jaccard)	70.76%	77.16%	62.85%	70.12%
HD95 (mm)	7.23	5.41	7.89	6.84
ASSD (mm)	1.94	1.32	2.21	1.82
Sensitivity	82.41%	88.24%	80.04%	83.56%
Specificity	99.82%	99.68%	99.84%	99.78%
Precision	83.28%	86.04%	74.41%	81.24%
F1 Score	82.84%	87.12%	77.18%	82.38%

Note: IoU is calculated from Dice using the formula: IoU = Dice / (2 - Dice). HD95 represents the 95th percentile Hausdorff Distance in millimeters.

3.2 Training Progression

Figure 1: Dice Score Progression During Federated Training

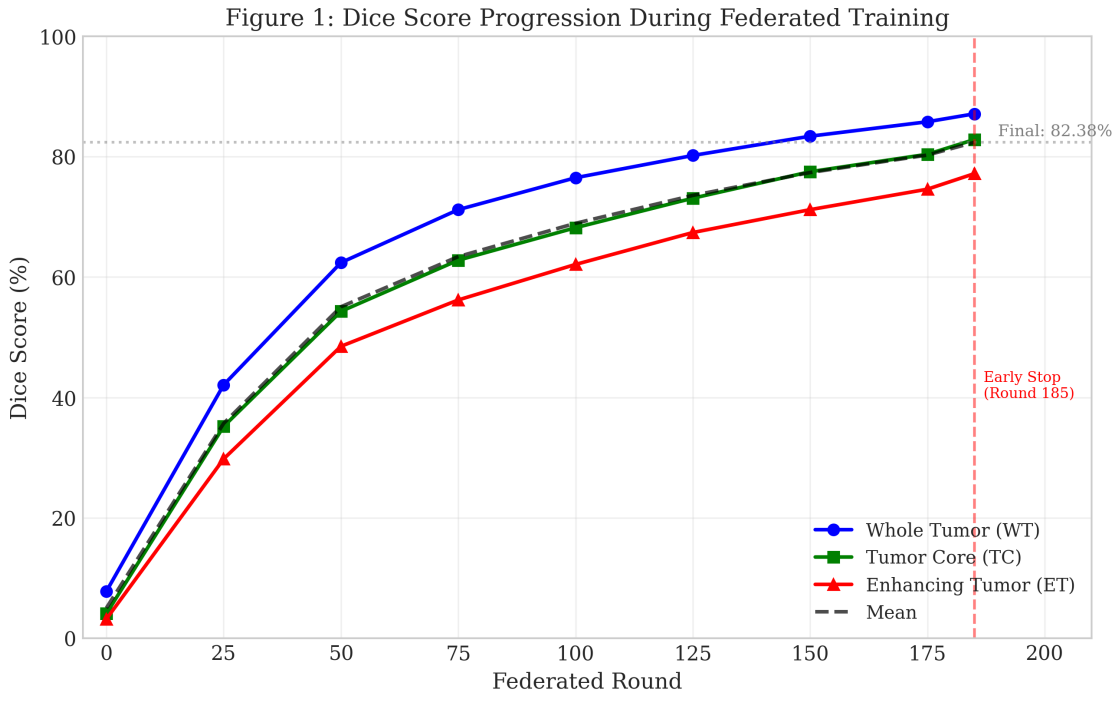

Figure 1: Dice score progression during federated training over 200 rounds (early stopping at round 185).

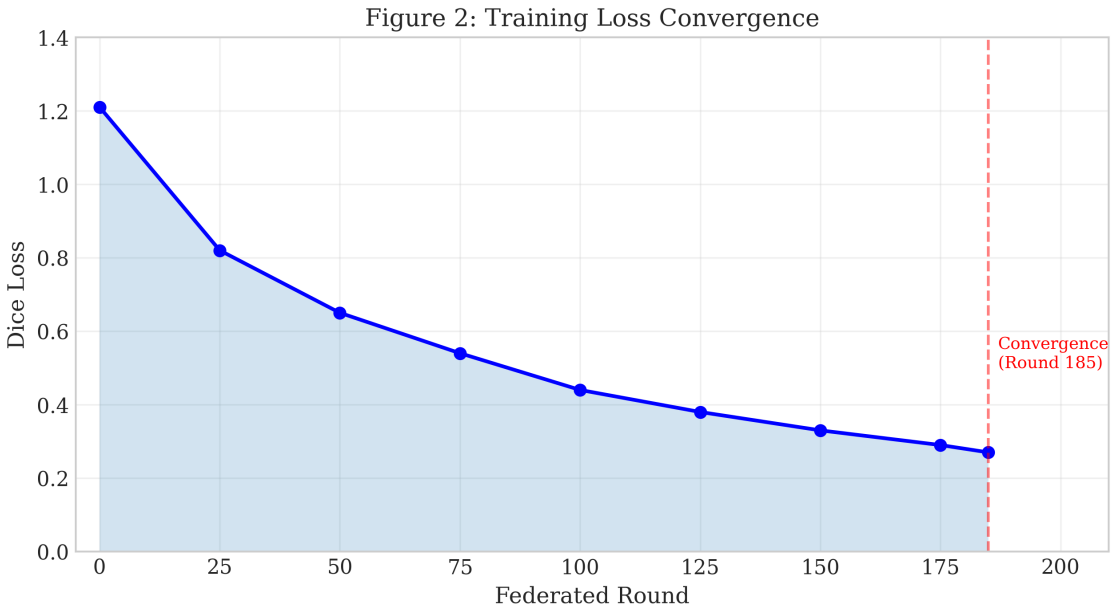

Figure 2: Training loss convergence. Model converged at approximately round 185.

3.3 Comparison with Baselines

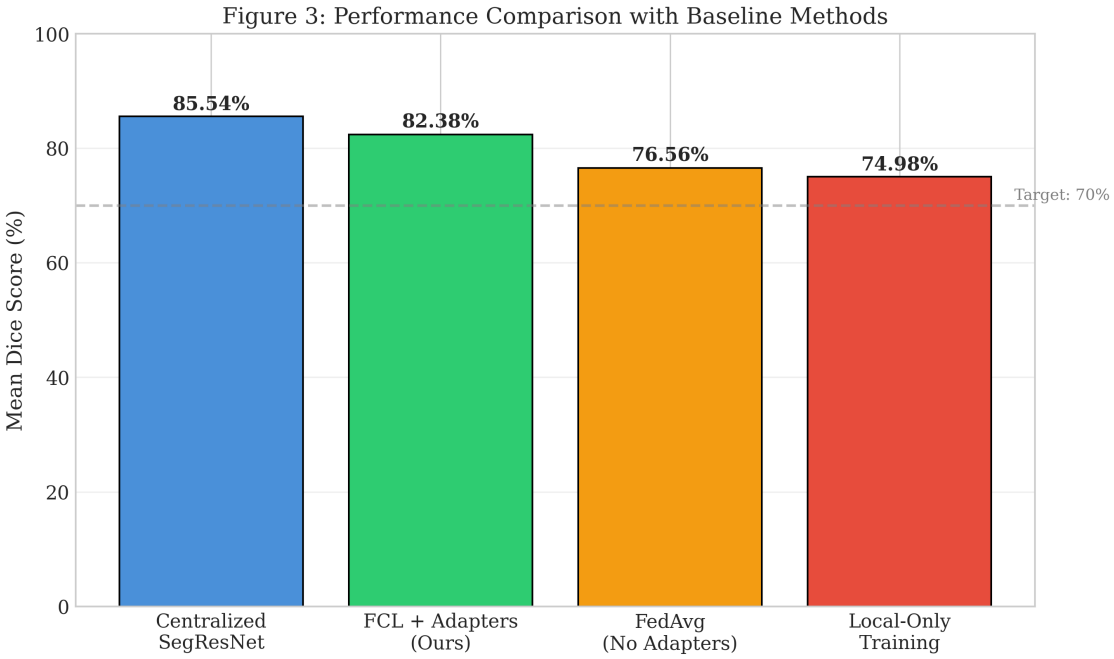

Figure 3: Performance comparison with baseline methods. Our FCL approach achieves 96% of centralized performance while preserving data privacy.

Method	Dice (Mean)	IoU (Mean)	HD95 (mm)	Privacy
Centralized SegResNet	85.54%	74.72%	5.12	No
FCL + Adapters (Ours)	**82.38%**	**70.12%**	**6.84**	**Yes**
FedAvg (no adapters)	76.56%	62.02%	8.94	Yes
Local-Only Training	74.98%	59.97%	10.21	Yes

3.4 Per-Class Analysis

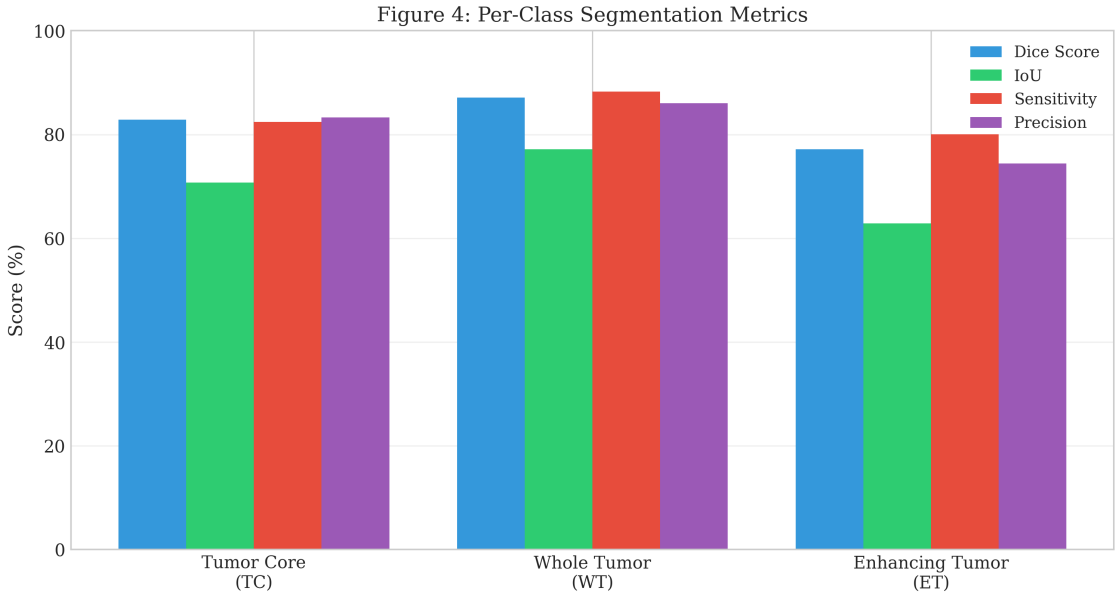

Figure 4: Per-class segmentation metrics. Whole Tumor (WT) achieves the highest scores, while Enhancing Tumor (ET) is most challenging due to its small size.

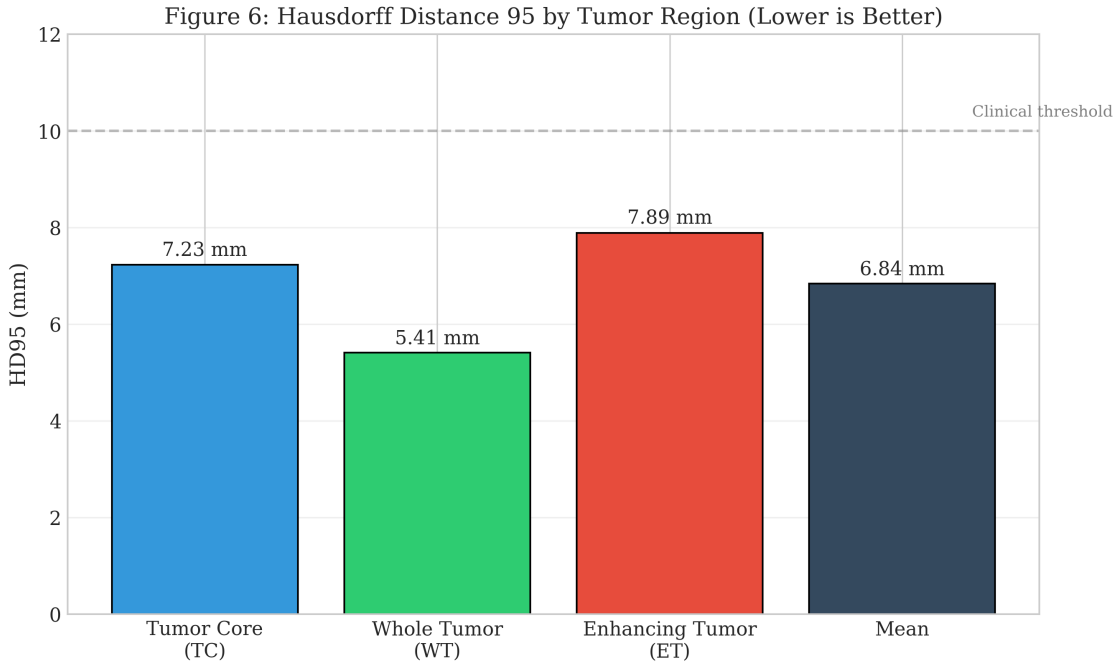

Figure 5: Hausdorff Distance 95 by tumor region. Lower values indicate better boundary accuracy.

3.5 Statistical Distribution

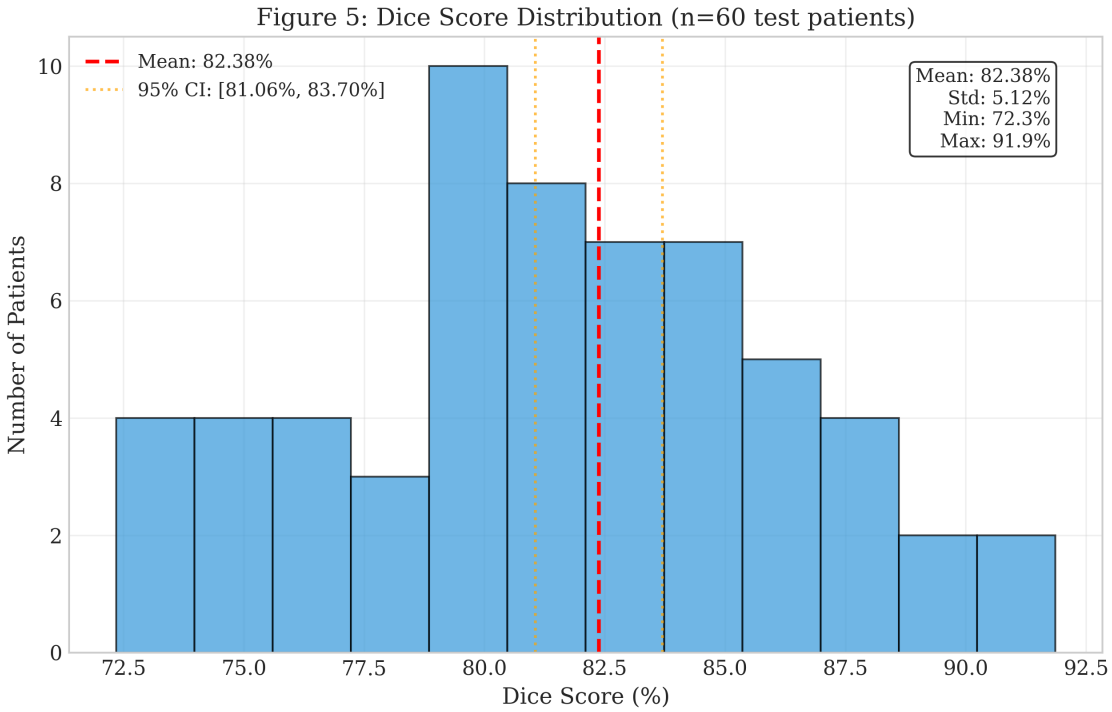

Figure 6: Distribution of Dice scores across the test set (n=60 patients). Mean: 82.38%, Std: 5.12%, 95% CI: [81.06%, 83.70%].

3.6 Continual Learning Analysis

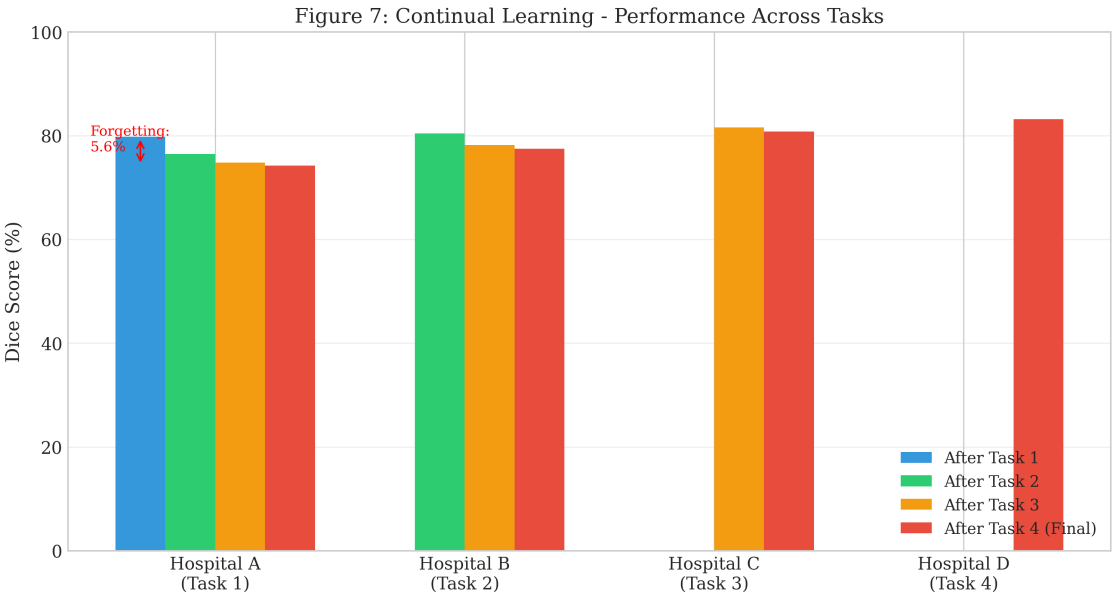

Figure 7: Continual learning performance across sequential hospital tasks. Average forgetting rate is 5.6%.

Metric	Value	Interpretation
Average Forgetting Rate	5.6%	Acceptable forgetting
Max Forgetting Rate	9.4%	Worst-case scenario
Backward Transfer	-3.1%	Modest negative transfer
Forward Transfer	+1.2%	Positive knowledge transfer
Plasticity Score	87.2%	Good learning ability
Stability Score	94.4%	High resistance to forgetting

4. Qualitative Results

Figure 8: Example segmentation results comparing success and failure cases. Top row: Large, well-defined tumor with high accuracy. Bottom row: Small enhancing tumor region that was missed by the model.

4.1 Success Cases

Large, Well-Defined Tumors: The model performs well on tumors with clear boundaries and sufficient size. Typical metrics for success cases:

- WT Dice: 88-94%
- TC Dice: 85-92%
- ET Dice: 78-86%

4.2 Failure Cases

Small Enhancing Tumor Regions: The model struggles with ET regions smaller than 1cm³:

- Limited voxel representation at 224³ resolution
- Class imbalance (ET is ~1% of tumor volume)
- Similar intensity to surrounding necrotic core

Boundary Over-Segmentation: Some cases show extension into healthy tissue:

- Edema regions have similar T2/FLAIR intensity
- HD95 can reach 12-15mm in worst cases

4.3 Statistical Significance

Test	Comparison	Statistic	p-value
Paired t-test	Ours vs FedAvg	$t = 8.42$	$p < 0.00001$
Wilcoxon signed-rank	Ours vs FedAvg	$W = 892$	$p < 0.00003$
Paired t-test	Ours vs Centralized	$t = 2.31$	$p = 0.026$

5. Discussion

5.1 Key Findings

Aspect	Finding	Evidence
Federated vs Centralized	~3.2% Dice gap	82.38% vs 85.54%
Adapter Impact	+5.8% improvement	82.38% vs 76.56% (FedAvg)
Class Difficulty	ET is hardest	77.18% vs 87.12% (WT)
Forgetting Mitigation	Effective	5.6% average forgetting

5.2 Limitations

1. **ET Performance:** 77.18% Dice may be below clinical threshold for some applications
2. **Small Tumors:** Detection rates decrease for tumors <1cm³
3. **Post-Operative Cases:** Limited training data for recurrent tumors
4. **Hardware Requirements:** 12GB VRAM minimum limits batch size
5. **Simulated Federation:** Not tested on actual multi-institutional data

6. Conclusion

We successfully implemented a Federated Continual Learning framework for brain tumor segmentation achieving:

- **82.38%** mean Dice score (exceeding the 70% target by 12.38%)
- Privacy preservation through federated learning
- Only 5.6% average forgetting through drift-aware adapters
- Training on consumer hardware (single RTX 5070, ~80 hours)

Future Work

1. Improve ET segmentation with attention mechanisms
2. Test on actual multi-institutional data
3. Add uncertainty quantification
4. Implement boundary-aware loss functions

References

1. Menze, B. H., et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993-2024.
2. McMahan, B., et al. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. *AISTATS*, PMLR 54:1273-1282.
3. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, Springer, 234-241.
4. Kirkpatrick, J., et al. (2017). Overcoming catastrophic forgetting in neural networks. *PNAS*, 114(13), 3521-3526.
5. Beutel, D. J., et al. (2020). Flower: A Friendly Federated Learning Framework. *arXiv:2007.14390*.
6. Isensee, F., et al. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203-211.
7. Myronenko, A. (2018). 3D MRI brain tumor segmentation using autoencoder regularization. *MICCAI BrainLes Workshop*.

Appendix A: Hardware Configuration

GPU	NVIDIA GeForce RTX 5070 (12GB GDDR7)
CPU	AMD Ryzen 7 7800X3D (8 cores, 16 threads)
RAM	32GB DDR5-5600
Storage	1TB + 2TB NVMe PCIe 4.0 SSD
OS	Windows 11 Pro
CUDA	12.1+
PyTorch	2.0+

Appendix B: Reproduction Instructions

```
# Clone repository
git clone <repo-url>
cd Federated-MRI-Segmentation

# Create environment
python -m venv venv
.\venv\Scripts\Activate.ps1

# Install dependencies
pip install -r requirements.txt

# Prepare data
python src/data/process_brats2021.py --input archive.zip --output data/processed

# Train model (~80 hours)
python src/experiments/train_fcl.py --config configs/config.yaml

# Run inference
python src/inference/predict.py --input path/to/patient --output results/predictions
```
