**Abstract**

We reproduce and adapt **3D-TransUNet**—a U-Net-style 3D segmentation model augmented with Vision Transformers—to segment glioma sub-regions from BraTS-2019 multi-parametric MRI. We keep the paper's *decoder-only* configuration (Transformer decoder + CNN encoder) and train with nnU-Net's engine on a single GPU in Colab. Our main methodological addition is a **Hausdorff 95th percentile distance (HD95)** metric computed online (per-epoch, in millimeters) for the BraTS composite regions (ET/TC/WT). On a fold-0 split (≈ 208 train / 53 validation) trained for 80 epochs with AdamW+cosine decay, we observe stable convergence, class-wise Dice reaching ~0.74 (ET), ~0.86 (TC) and ~0.87 (WT), and HD95 decreasing quickly in early epochs and stabilizing near 5–7 mm. This boundary-aware signal complements Dice and is particularly informative for small enhancing components. We discuss limitations (single-fold, short training horizon) and outline next steps to reach paper-level performance.

## 1. Introduction

### Goal & motivation

Accurate, reproducible delineation of brain-tumor sub-regions—**Enhancing Tumor (ET), Tumor Core (TC)** and **Whole Tumor (WT)**—from multiparametric MRI underpins response assessment and therapy planning. Classic CNN-only U-Nets capture local texture well but struggle with long-range context; Transformers contribute global context via self-attention but can lose spatial detail. **3D-TransUNet** explicitly combines both: a CNN path for high-resolution features and Transformer blocks (encoder, decoder, or both) for global reasoning. The authors show that *decoder-only* designs favor small/heterogeneous targets such as tumors, while *encoder-only* favors multi-organ tasks.

### Previous work

- **U-Net / nnU-Net.** The U-Net family dominates medical segmentation; nnU-Net contributes self-configuring pre-/post-processing and strong baselines.
- **TransUNet / 3D-TransUNet.** Introduces Transformer components inside U-Net. The 3D variant provides three options—**encoder-only**, **decoder-only**, and **encoder+decoder**—and attains state-of-the-art results on diverse datasets (Synapse, MSD Hepatic Vessel, BraTS-2021, BraTS-MET-2023).

**Our focus.** We re-implement the **decoder-only** 3D-TransUNet on **BraTS-2019** and add a **Hausdorff-95 (HD95)** evaluation not reported for brain tumors in the original paper, logging ET/TC/WT HD95 (mm) every epoch.

## 2. Method

### 2.1 Architecture: 3D-TransUNet (decoder-only)

We follow the paper's decoder-only variant: a CNN encoder supplies multi-scale features to a **Transformer decoder** that reframes segmentation as **mask classification with learnable queries**. Queries are iteratively refined by **masked cross-attention** against CNN features (coarse-to-fine); the final masks are produced by dot products with the last CNN feature map followed by per-class classification of masks.

**Our configuration:** ViT depth = 12, hidden = 768, MLP dim = 3072, 12 heads; batch = 4; **80** epochs; **AdamW** (3e-4) with **cosine annealing**; mixed precision; deep supervision enabled.

### 2.2 Loss & optimization

We use nnU-Net's **Dice + cross-entropy** hybrid applied to deep-supervision outputs. The implementation minimizes **–Dice + α·CE**, so the *total* training loss can become **negative** once Dice dominates—matching our plots and slide explanation.

### 2.3 Data & pre-processing

**Dataset.** BraTS-2019 (Kaggle mirror), 4 MRI channels per case (T1, T1ce, T2, FLAIR) with labels {0,1,2,4}. Composite regions: **WT = {1,2,4}**, **TC = {1,4}**, **ET = {4}**.
**nnU-Net integration.** We generate dataset.json, rename to _0000…_0003.nii.gz, and run nnUNet_plan_and_preprocess to normalize spacing, intensities, and discover patch size.

### 2.4 New contribution: on-the-fly HD95 (mm)

We extend the trainer to compute **per-epoch HD95** for **ET/TC/WT** directly during validation:

- **Labels used:** ET = 4; TC = {1,4}; WT = {1,2,4}.
- **Metric:** MedPy's hd95 on hard labels (argmax), using voxel spacing from the nnU-Net plans to report **millimeters**.
- **Edge cases:** if both prediction and reference are empty for a region → **NaN**; if one is empty → **∞**; we aggregate with **nanmean** across the epoch.
- **Why HD95?** It is **boundary-sensitive** and interpretable in mm; it highlights small boundary errors or spurious islands that can be invisible to Dice, especially for ET.

### 2.5 Engineering & training loop

- Single-GPU Colab; copy preprocessed/ to local SSD for speed.
- Custom train.py fuses the paper's config with nnU-Net's get_default_configuration, auto-registers nnUNetTrainerV2_HD95, and supports --resume/--validation_only.
- We disable deep-supervision at inference (matching nnU-Net's practice)

## 3. Experiments and Results

### 3.1 Experimental protocol

- **Data split:** nnU-Net 5-fold generator; we report **fold 0** only (≈ 208 training / 53 validation).
- **Training:** 80 epochs, **AdamW** (β=0.9/0.999, weight-decay 1e-2), **cosine** LR from 3e-4 → 1e-6, mixed precision, batch 4.
- **Augmentations:** 3D rotations (±30°), scaling (0.7–1.4), flips, intensity transforms—as in nnU-Net defaults.

### 3.2 Training dynamics

- **Loss curves.** Rapid drop in the first ~10 epochs, then smooth convergence with **train/val curves closely tracking** and no late-epoch divergence (Figure "Loss curves"). The negative values are expected because the objective includes –Dice.
- **Dice curves. ET** rises from <0.2 to ~0.70–0.75; **TC** and **WT** plateau around **0.84–0.88** by epoch 80.
- **HD95 curves (mm).** All regions show a steep reduction (≈ 20–45 mm → ≈ 5–8 mm) within ~10 epochs; **WT** tends to be lowest (~5–6 mm), **ET/TC** ~6–7 mm with occasional spikes, then flatten.

### 3.3 Quantitative (fold 0, ~epoch 80)

| Region | Dice (≈) | HD95 (mm, ≈) |
|---|---|---|
| ET | 0.72–0.75 | 6–7 |
| TC | 0.85–0.86 | 6–8 |
| WT | 0.86–0.88 | 5–6 |

*(Noted the **best val** loss around epoch 65.)*

### 3.4 Context vs. prior work

The 3D-TransUNet paper reports BraTS-2021 5-fold performance of **ET 88.85, TC 92.48, WT 93.90** (Dice), surpassing nnU-Net-Large (avg 91.47 vs 91.74).
Our pilot is **not directly comparable** (BraTS-2019 vs. -2021, single fold, 80 vs. 1000 epochs, no extensive post-processing). Nevertheless, the **shape** of our Dice and the **early HD95 collapse** are consistent with the decoder-only design benefiting tumor/lesion targets seen in the paper.

## 4. Discussion

**Why the HD95 addition matters.** Dice is insensitive to thin boundary errors; HD95 in **millimeters** highlights clinically relevant contour deviations (e.g., "ragged" edges or small spurious islands). In our run, HD95 provided fast-moving validation feedback in early epochs and a complementary view to Dice for **ET**, the hardest region.

**Strengths of this pipeline.**

- Fully **reproducible** end-to-end nnU-Net → 3D-TransUNet training in Colab (automatic dataset JSON, planning, caching to SSD).
- **Decoder-only** configuration matches the tumor-centric setting and shows stable convergence on a **single GPU**.
- **New HD95 logging** requires no extra I/O (computed from in-memory predictions each epoch).

**Limitations.**

- Single fold, **80** epochs — considerably shorter than the paper's regimen; no test-time ensembling or extensive post-processing.
- We did not include **empty-case handling** in loss weighting; ET imbalance still affects stability (occasional HD95 spikes).
- Our baseline comparison is to the paper's reported numbers (different year of BraTS).

## 5. Conclusion & Future Work

We implemented 3D-TransUNet (decoder-only) for glioma sub-region segmentation on BraTS-2019 and contributed an **HD95-aware trainer** that logs ET/TC/WT **in mm** each epoch. The model converges stably; boundary quality (HD95) improves sharply early on and plateaus near 5–7 mm. To move toward paper-level results:

- **Scale up training** to ≥ 1000 epochs and run **all 5 folds**; select *ValBest* checkpoints per fold before ensembling.
- Add **morphological post-processing** (island removal for ET) and **test-time augmentation**.
- Explore **encoder+decoder** variants and **pretrained ViT** backbones.
- Extend to BraTS-2021/2023 splits for direct comparison to the paper.

## References


- Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M., Xing, L., Lu, L., Yuille, A., and Zhou, Y., "**3D TransUNet: Advancing Medical Image**

**Segmentation through Vision Transformer**s," *arXiv preprint* arXiv:2310.07781, 2023. arXiv+1

- Ronneberger, O., Fischer, P., and Brox, T., "**U-Net: Convolutional Networks for Biomedical Image Segmentation**," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), Lecture Notes in Computer Science, vol. 9351, Cham: Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28. SpringerLinkarXiv

- Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H., "**nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation**," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021. doi:10.1038/s41592-020-01008-z. (Epub Dec. 7, 2020.) NaturePubMed



**Ethics Statement: Project Stakeholders and Implications Assessment**

**1) Introduction**

**Student names:** Ari Aharon Shemesh; Itay Asael
**Project title:** *BT-Seg: Brain-Tumor Segmentation with 3D-TransUNet and HD95*
**Project description.** We train a 3D-TransUNet to segment tumor sub-regions (ET, TC, WT) from multi-sequence brain MRI and add HD95 to evaluate boundary quality in millimeters. The goal is a reproducible research pipeline that can inform future clinical tooling and experimentation, not a clinical product.

**2) LLM-generated answers about stakeholders, explanations, and responsibility**

**Prompt used:**

*"Given a student project that trains 3D-TransUNet on BraTS MRI to segment tumor sub-regions and evaluates HD95, identify three stakeholder groups affected by such technology; draft a concise explanation appropriate for each group; and state who should be responsible for delivering those explanations."*

**a. Three stakeholder types.**

1. **Patients and caregivers** (whose scans could one day be analyzed by similar models).

2. **Clinicians and radiology teams** (who may review, correct, or rely on model contours).
3. **Hospital/health-system data stewards & IRBs** (guardians of data governance, privacy, and validation standards).

**b. What an explanation to each might look like (≤ 1 paragraph each).**

- *Patients/caregivers*. "This research prototype outlines tumor regions on MRI using patterns learned from past anonymized scans. It can miss or over-segment areas and is not a diagnostic tool. Experts review and correct its suggestions. Your privacy is protected through de-identification, and no decisions are made solely by the model."
- *Clinicians*. "The model outputs ET, TC, and WT masks and reports HD95 (mm) to indicate boundary accuracy. It was trained on BraTS-2019 with nnU-Net-style pre-processing; performance varies by cohort and scanner. Use contours as an assistive prior and verify against raw images; do not rely on them without validation on your local data."
- *Data stewards/IRBs*. "The pipeline consumes de-identified MRIs under data-use agreements, logs no PHI, and supports audit of training/evaluation. Before deployment, prospective validation, bias assessment (scanner/site, sex, age), and monitoring are mandatory. Model artifacts must be versioned; access is controlled."

**c. Who is responsible for giving each explanation (≤ 1 paragraph).**

- *Patients/caregivers:* the **treating clinical team** (attending radiologist/neuro-oncologist) during consent or results discussion; communications crafted with input from **ethics/communications** offices.
- *Clinicians:* the **ML/clinical research team** publishing the tool, with **department leadership** setting usage policy and mandatory training.
- *Data stewards/IRBs:* the **principal investigator** and **institutional privacy office**, ensuring adherence to IRB protocols, DUA terms, and security reviews.

**3) Reflection on the AI output (manual)**

To strengthen the above explanations ethically, we would add: (1) **Explicit uncertainty metrics** (using conformal prediction especially when ET is tiny). (2) **Equity & generalization risks** (pediatric vs. adult cases, under-representation minorities).