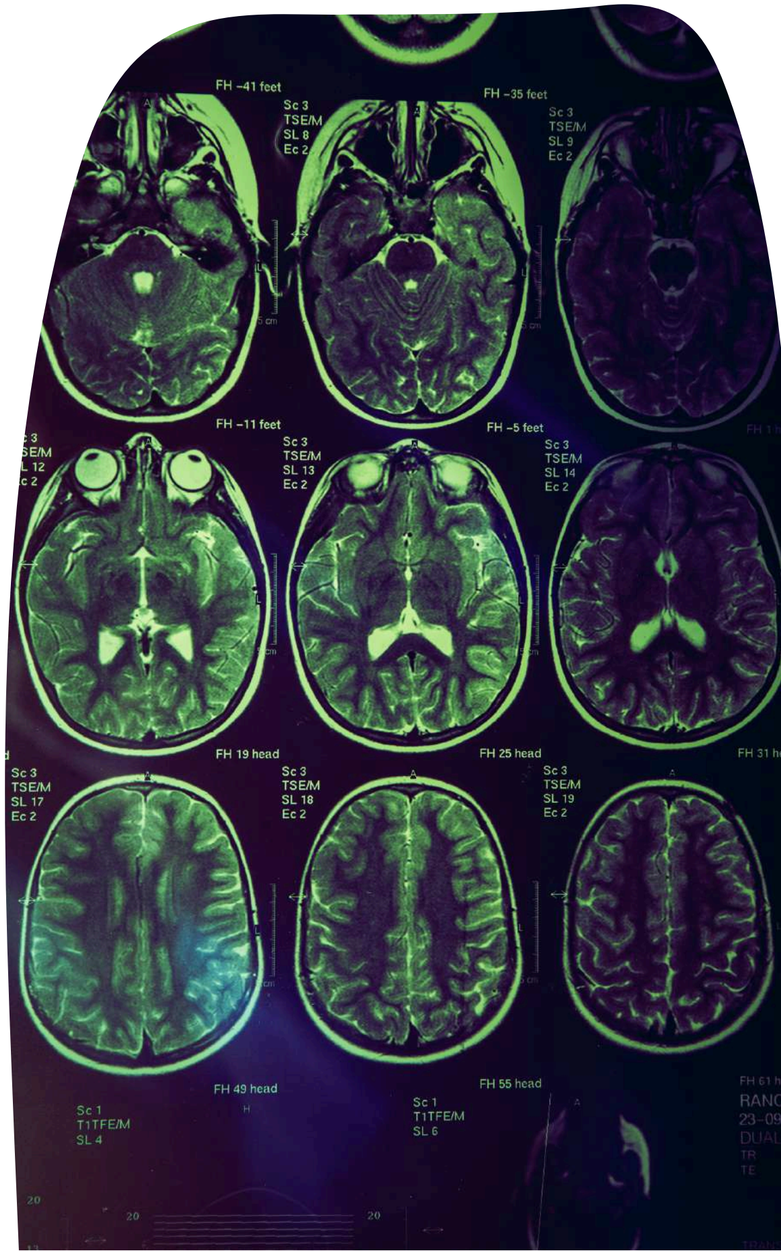
DETECTING BRAIN ANOMALIES USING CONDITIONED DIFFUSION MODELLING

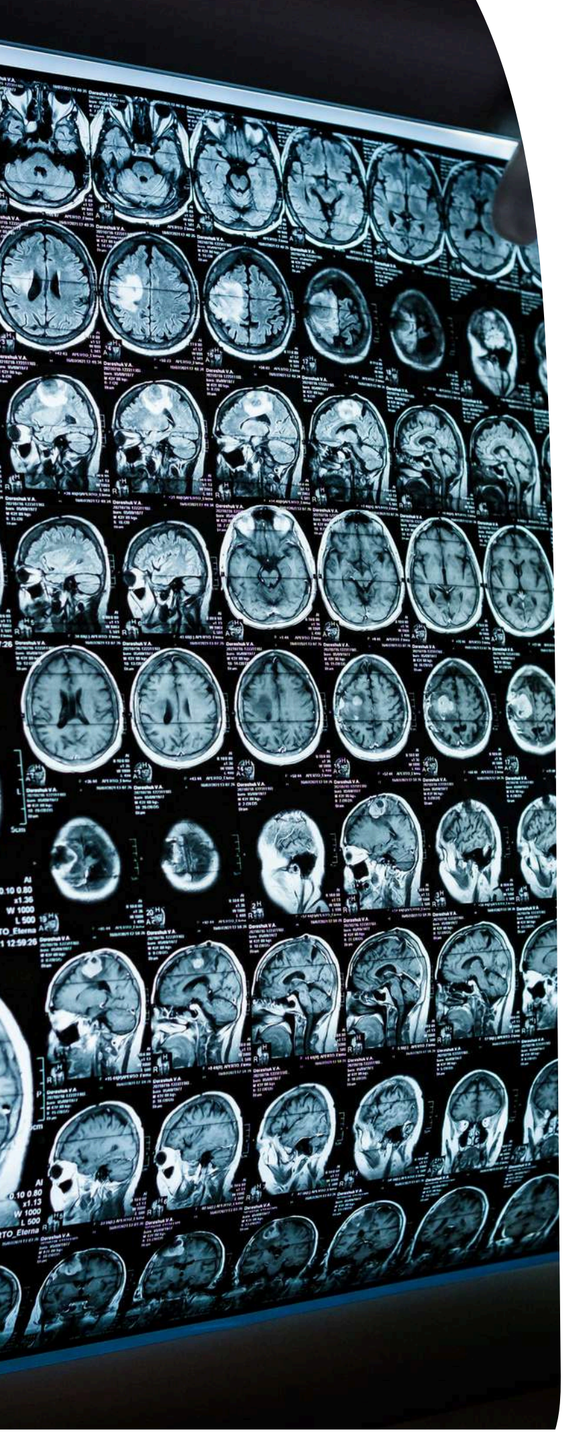
KESHAV KANT AHUJA
PRAJNA SHARMA
SUKHMANI KAUR

VALIDITY OF OUR PROBLEM STATEMENT

- ***Neurological disorders require accurate MRI analysis, but manual interpretation is slow, costly, and variable.***
- ***Unsupervised anomaly detection (UAD) learns normal brain patterns from healthy MRI scans to identify abnormalities without lesion labels.***
- ***We explored whether better conditioning and larger datasets improve detection performance and robustness.***

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. NeurIPS 2020.



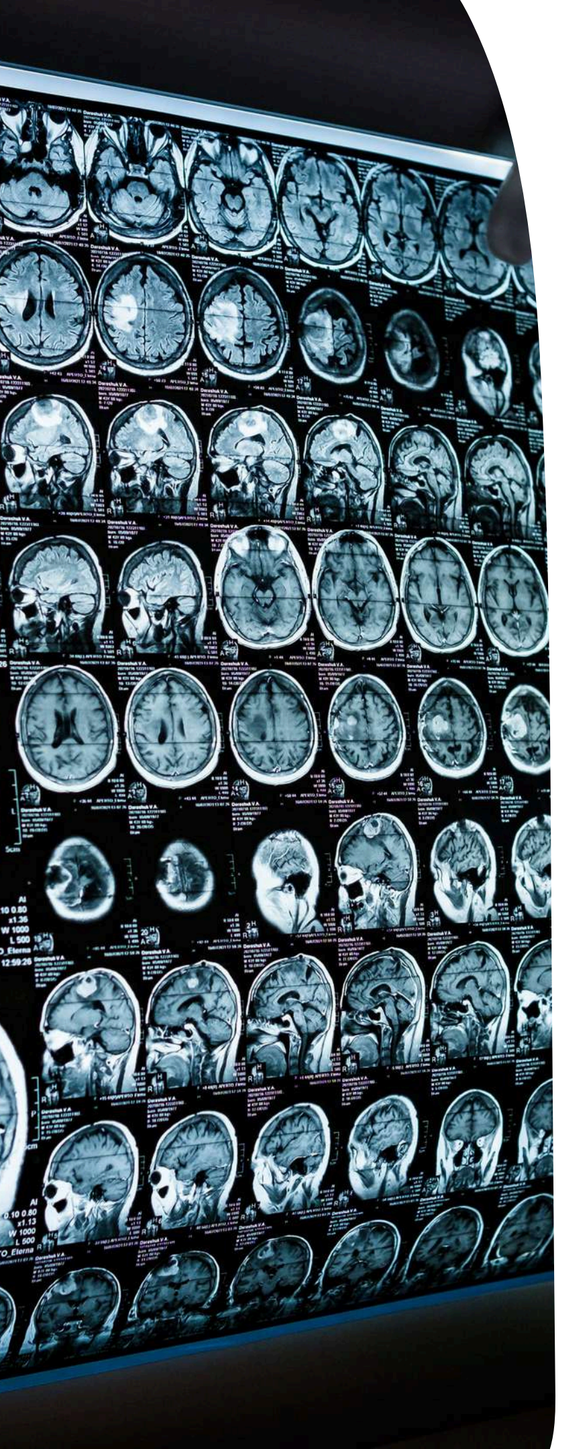


LITERATURE SURVEY

- **Why UAD & Limitations of Early Methods**

Early Generative Methods Fall Short

- VAEs produce blurry outputs due to dense bottleneck
- GANs suffer mode collapse and training instability
- Diffusion models offer sharper, more stable reconstructions — evaluated via Dice & AUPRC
- Damudi, M. Z., & Kini, A. S. (2024). Single-Step Sampling for UAD. IJBI 2024.



Evolution of Diffusion-Based UAD

From Partial Diffusion to Conditioned Reconstruction

Partial Diffusion : AnoDDPM

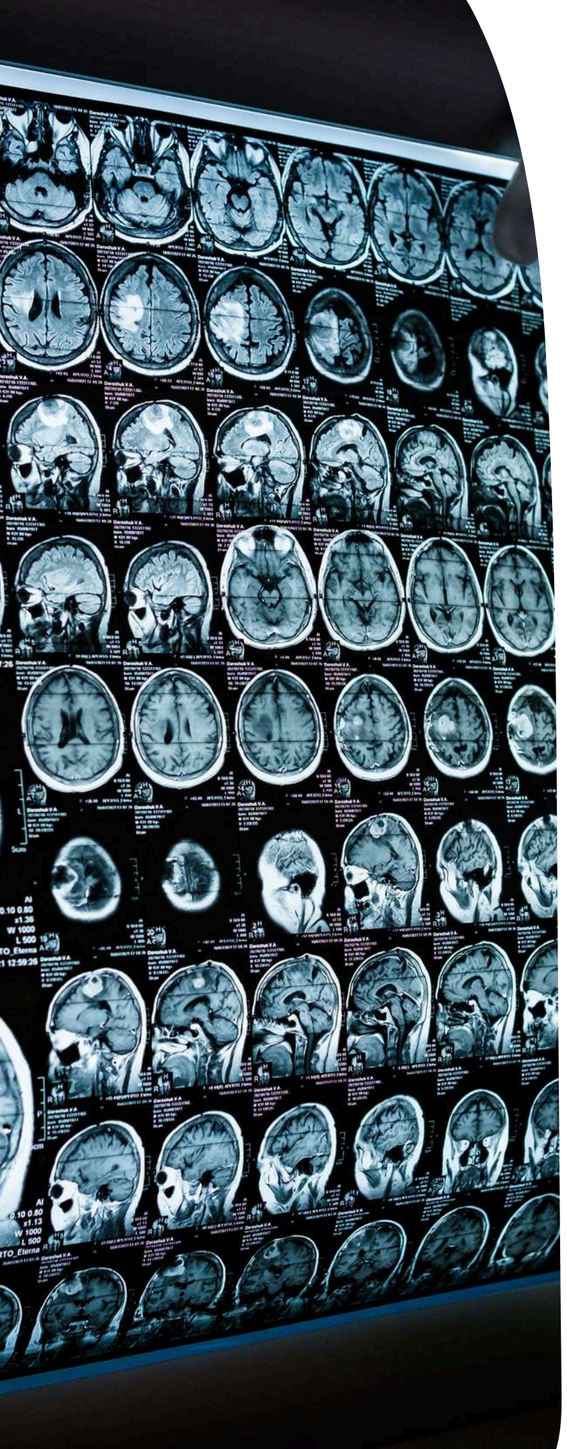
- Only partial noising needed (t_test < T); simplex noise masks large lesions better
- +25.5% Dice (segmentation coefficient), +17.6% IoU over f-AnoGAN
- Wyatt, J. et al. (2022). AnoDDPM. CVPR Workshop 2022.

Conditioning to Reduce False Positives - cDDPM

- Intensity mismatch in vanilla DDPMs causes false positives
- 128-dim conditioning vector captures global intensity without copying pathology
- Injected into UNet via FiLM; +11.9% Dice (BraTS21), +20% Dice (ATLAS) measured via AUPRC
- Behrendt, F. et al. (2025). Conditioned DDPM for UAD in Brain MRIs. CBM 2025.

Improved Anomaly Scoring

- Consistently outperforms single-σ SSIM and L1 across all pathologies on Dice & AUPRC
- Behrendt, F. et al. (2024). Diffusion Models with Ensembled SSIM Scoring. ICASSP 2024.



The Case for 3D Encoding & Our Gap

Why 3D Context Matters & What We Address

Evidence: 3D Encoding Outperforms 2D Slice-Based Methods

- Lesions span continuous slices - inter-slice context is essential for accurate detection
- 3D Residual AE (ResNet-50): Dice 0.462, AUPRC 0.844 (glioma) - beats VAE & f-AnoGAN
- Luo, G. et al. (2023). Unsupervised Anomaly Detection in Brain MRI. CBM 2023.

Enabling Infrastructure

- FOMO300K: 318,877 healthy brain scans from 920 sources maximises scanner diversity
- Cerri, S. et al. (2026). FOMO300K. arXiv 2026.
- MONAI ResNet-50 3D: full volumetric input → 128-dim feature vector via global avg pooling
- Cardoso, M. J. et al. (2022). MONAI Framework. arXiv 2022.

The Gap We Address

- cDDPM conditions on 2D slices → discards all inter-slice volumetric structure
- Authors themselves flag 3D encoding as future work (Behrendt et al., Sec. 6.4)
- We replace 2D encoder with MONAI ResNet-50 3D - same 128-dim output, same FiLM injection

Dataset and evaluation summary

Curated healthy T2 MRI cohort from OpenNeuro, HCP Wu-Minn, and WAND; external BraTS21 T2 tumour evaluation

Healthy cohort
691 scans
OpenNeuro / HCP Wu-Minn / WAND

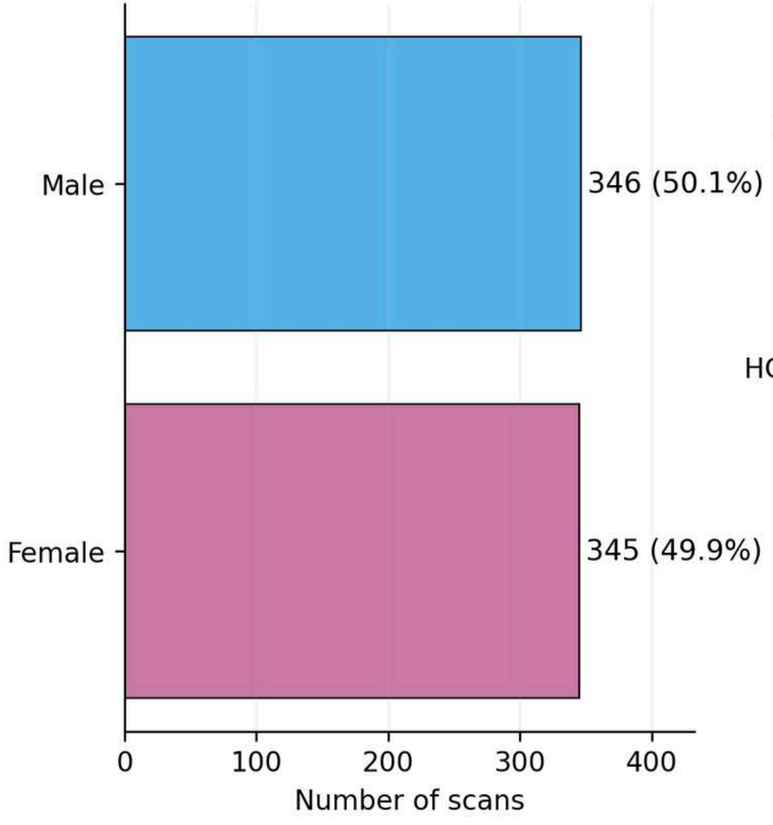
Representation
345 F / 346 M
Near-even sex composition

Split protocol
442 train / 54 val
Per fold; fixed healthy test = 195

External evaluation
1251 cases/fold
BraTS21 T2; failed evals = 0

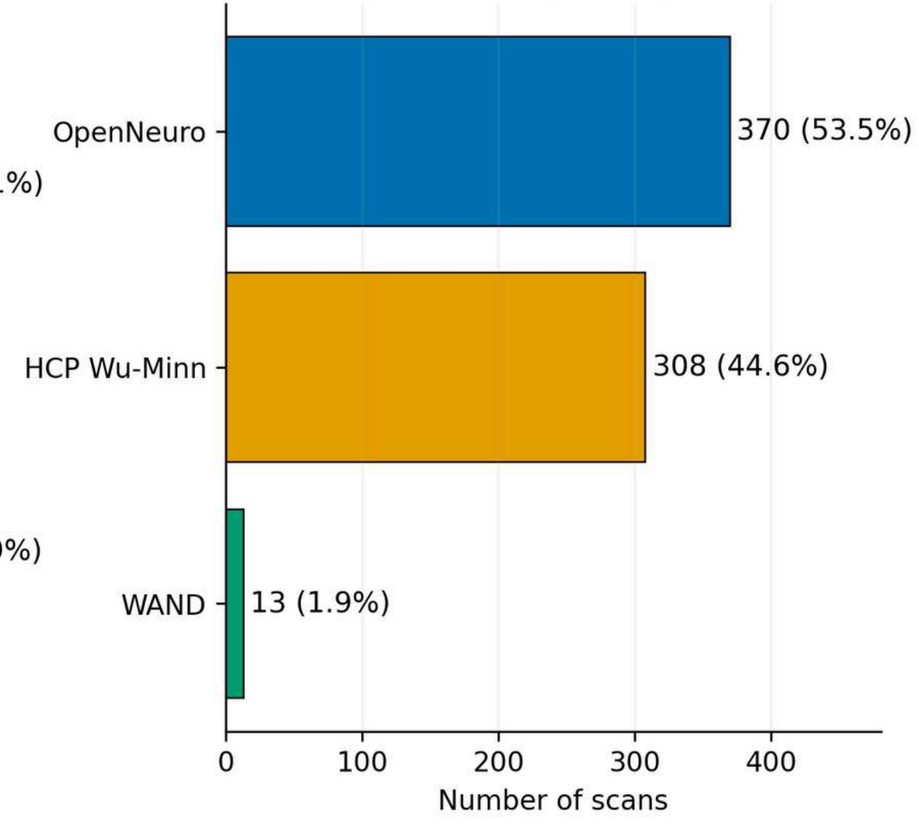
A

Healthy cohort sex composition



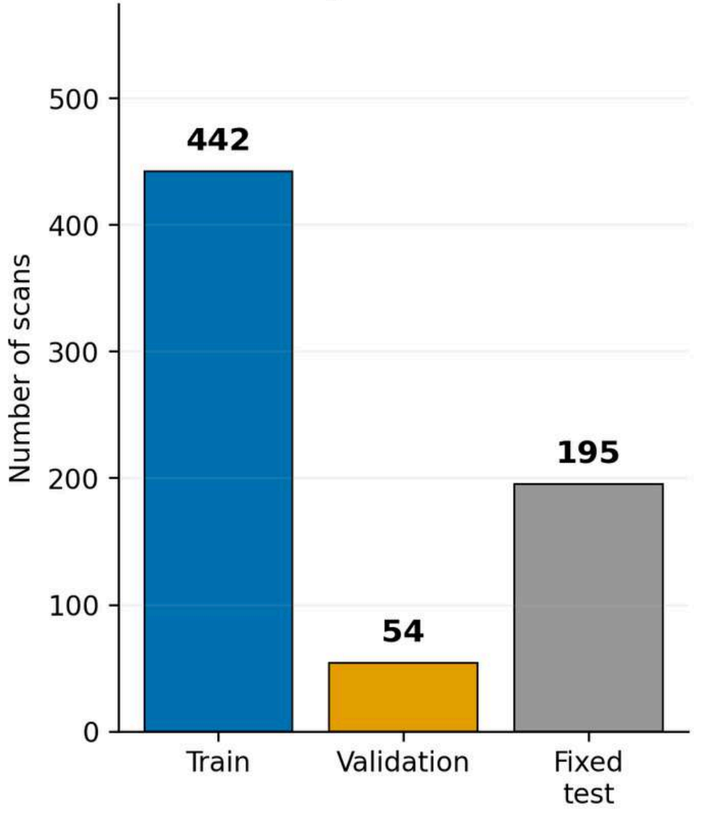
B

Source-family composition



C

Split sizes



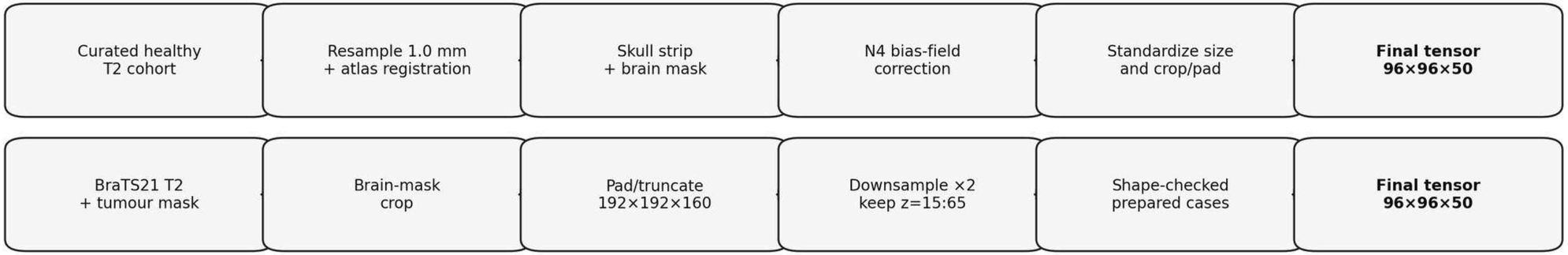
Split integrity: train/validation/test overlap = 0. Training uses healthy scans only; BraTS21 is external pathological evaluation.

Train/validation are per fold; fixed test is held out across folds.

Processing and evaluation pipeline

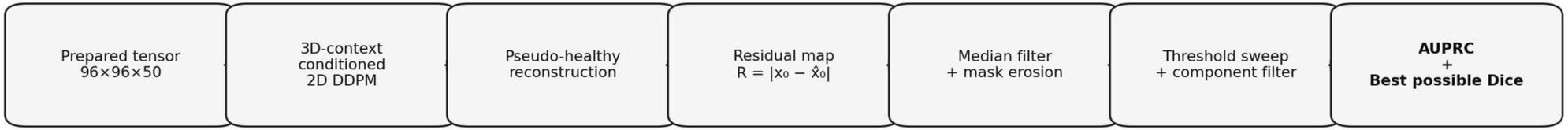
Pre-processing and residual-map post-processing adapted from Behrendt et al.; implemented for our healthy T2 and BraTS21 T2 workflow.

A Pre-processing and model-space preparation



Healthy scans undergo full preprocessing. BraTS21 cases are prepared and shape-checked into the same model-space tensor geometry.

B Inference, residual map, and evaluation

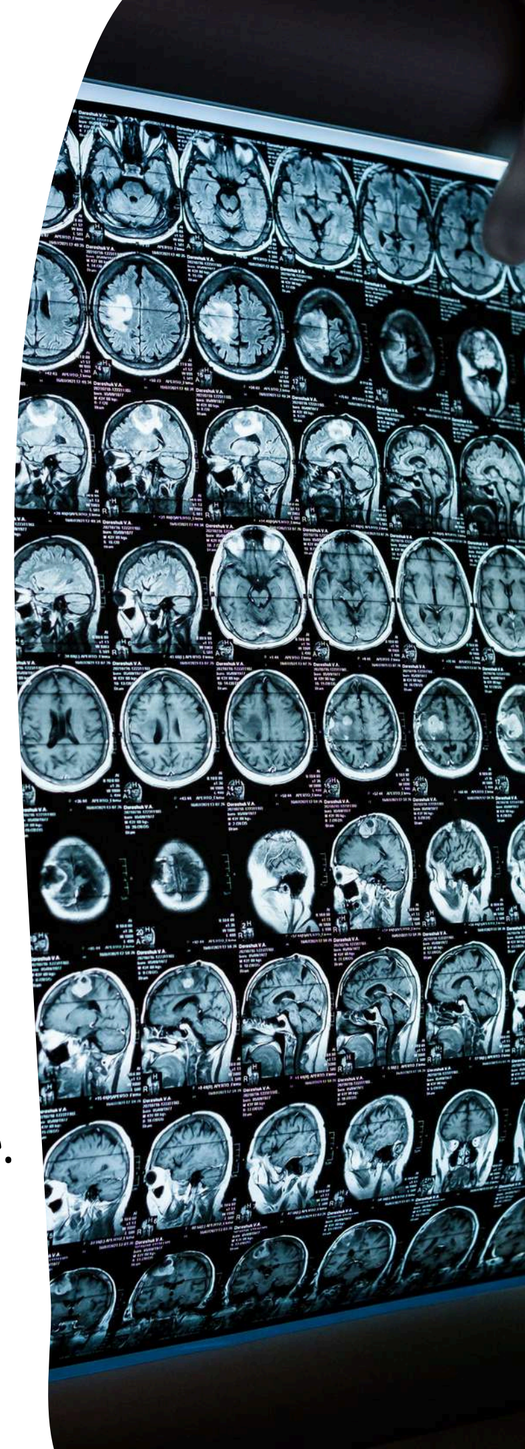


Best possible Dice = best-threshold/oracle Dice from a threshold sweep over residual maps; not fixed-threshold deployment Dice.

ML METHODOLOGY

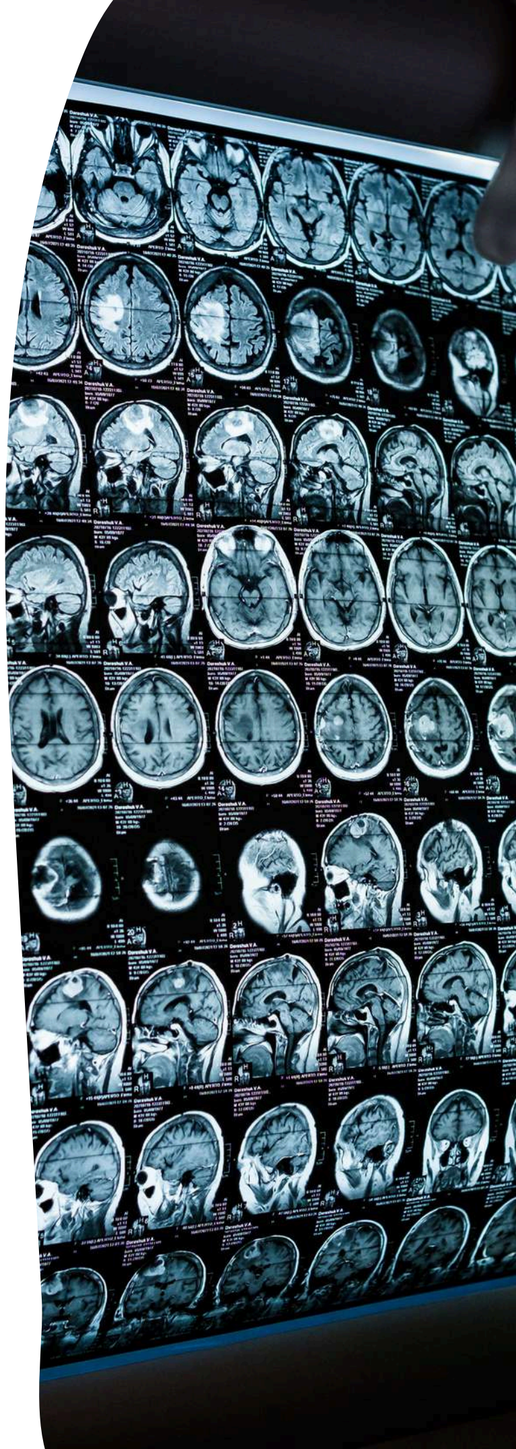
- Train only on healthy T2 MRI → learn normal brain anatomy.
- Pretrain 3D MONAI ResNet-50 encoder → capture volumetric context.
- Condition 2D DDPM U-Net → reconstruct slices as pseudo-healthy.
- Residual map |input – reconstruction| → anomaly evidence.
- Evaluate externally on BraTS21 T2 using AUPRC + best possible Dice.

Best possible Dice = threshold-sweep/oracle Dice, not fixed-threshold deployment Dice.



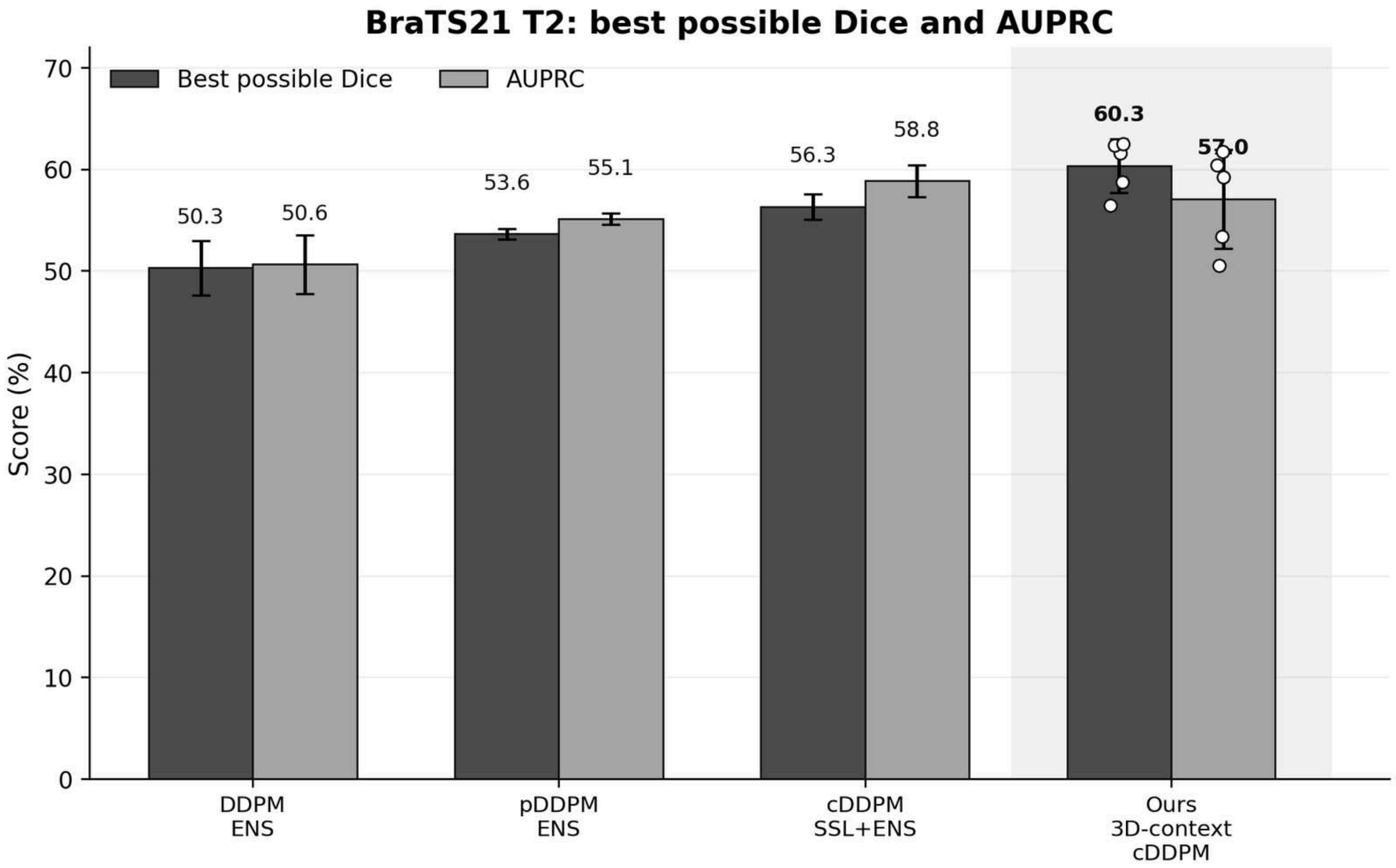
CHALLENGES AND LIMITATIONS

- Training instability in early iterations required pipeline debugging and reconfiguration
- Compute constraint: the model combines a 3D ResNet-50 context encoder with a 2D DDPM, giving ~90M trainable parameters. To keep fold-wise training feasible, we used batch size 8 and 16-bit AMP.
- Our final evaluation used t_test = 500. Behrendt et al.'s strongest cDDPM setting uses SSL + ENS, where multiple t_test values are ensembled. Therefore, our comparison is contextual, not a perfectly matched benchmark.
- Dice scores report the best threshold-swept Dice; conclusions should be based on the full fold aggregate rather than a single fold.



BraTS21 T2 segmentation performance

Selected DDPM-family rows from Behrendt et al. Table 2 vs. our 3D-context conditioned DDPM.



Interpretation

Contextual comparison, not a perfect one-to-one benchmark.

This work differs by:
- Gold_700 healthy T2 cohort
- 3D-context encoder extension
- single t_test = 500
- no t_test ensemble

Behrendt strongest cDDPM row uses SSL + ENS.

Our result: AUPRC 57.05% ± 4.85% | Best possible Dice 60.31% ± 2.64% (1251/1251 cases per fold; 0 failed evaluations)

Best possible Dice = threshold-sweep/oracle Dice over residual maps; not fixed-threshold deployment Dice. Bars show mean ± SD across folds.

PERFORMANCE METRICS

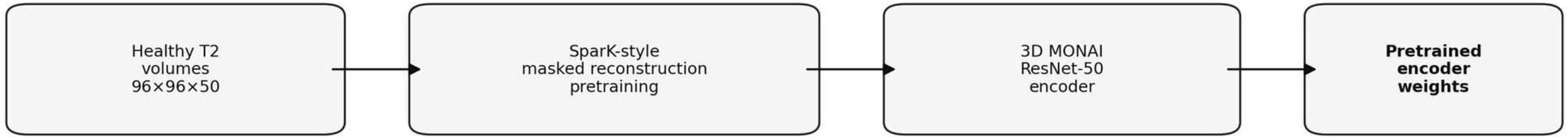
Interpretations: The performance comparison is contextual, not a perfectly matched benchmark. Our model uses a 3D-context encoder extension and single t_test = 500; Behrendt et al.'s strongest cDDPM row uses SSL + ENS. Best possible Dice is threshold-sweep/oracle Dice, not fixed-threshold deployment Dice.

MODEL ARCHITECTURE

3D-context conditioned diffusion model

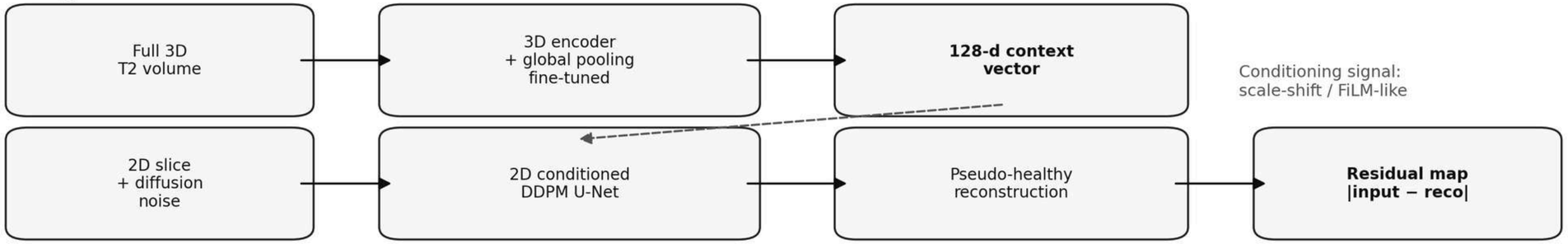
A 3D encoder summarizes healthy-volume context, which conditions 2D DDPM slice reconstruction.

A Stage 1: self-supervised 3D context encoder pretraining



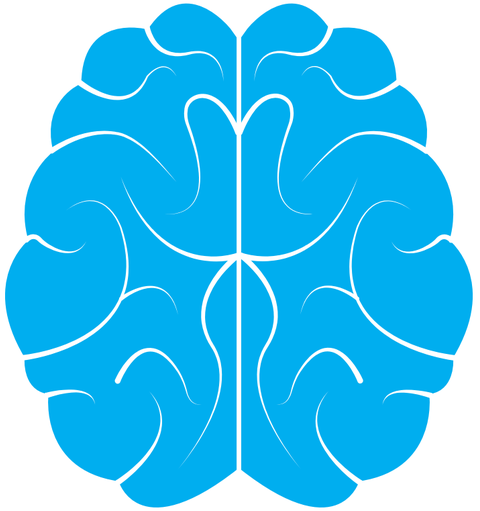
Goal: learn healthy 3D anatomical context before conditioning the diffusion model.

B Stage 2: 3D-context conditioned 2D DDPM reconstruction



At inference, tumour-containing slices are reconstructed as healthy; residuals form the anomaly map.

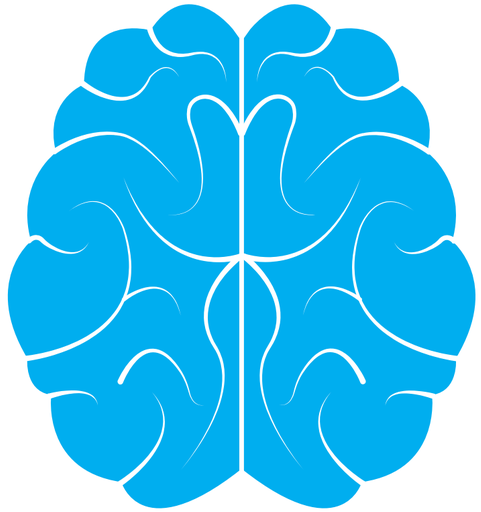
DEPLOYABILITY



What works

- Healthy-only training setup
- External pathological evaluation on BraTS21 T2
- Voxel-wise anomaly maps
- Public repository with code, configs, figures, tables, and audit reports
- Clear reporting of limitations

What is still needed



- Prospective multi-site validation
- Scanner/protocol robustness checks
- Runtime optimization for diffusion inference
- Comparison against clinical baselines
- Radiologist reader study
- Regulatory, privacy, and safety review

This model is best framed as a research prototype for unsupervised tumour anomaly localization, not a deployable clinical diagnostic system.