This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material MISFIT: A Self-Supervised Pretraining Toolkit for 3D Medical Imaging — A Preliminary Test Across Three BraTS 2026 Segmentation Tasks

Adrian Celaya[1,2], Yogesh Kumar[1], Awj Twam[1], Beatrice Riviere[2], David Fuentes[1], and Tucker Netherton[1]

[1] The University of Texas MD Anderson Cancer Center, Houston TX 77030
[2] Rice University, Houston TX 77005

1 Detailed Evaluation Metrics

Tables S1 through S3 provide detailed evaluation metrics for all tasks and methods.

2 Computational Costs

Per training step, the two architectures are closer than their sizes suggest: 0.137 s for the baseline compared to 0.186 s for either SwinUNETR method, a factor of 1.36 despite a factor of 94 in parameter count. The reason is likely attributable to patch size. The baseline processes a 128^3 patch against SwinUNETR's 96^3, which is 2.4 times the volume per step, a useful reminder that parameter count can be a poor proxy for training cost in 3D segmentation. The two SwinUNETR methods agree to within 0.2%, as they must, since initialization cannot change throughput; we report both columns rather than merging them because that agreement is a check on the measurement. End-to-end inference preserves the ordering, at 7.00, 5.37, and 3.52 s per case for the baseline on Tasks 1 to 3, compared with roughly 10.0, 10.0, and 5.7 s for the SwinUNETR-based methods. All figures are for a single-fold model; a five-fold ensemble multiplies the number of forward passes but not the preprocessing.

Pretraining is a one-time cost, incurred once and shared by all three tasks. The `swinunetr-base` MAE was trained on the 16,315-volume corpus at a 96^3 crop with 75% masking for 200 epochs on 4 NVIDIA A100 GPUs, reaching a best validation loss of 0.462. This pretraining run took roughly 16 hours using 4 A100 GPUs.

Class	Metric	Baseline	MISFIT	Scratch	p B-vs-M	p M-vs-S
Task 1 — Brain Metastases — $n = 1,196$						
NETC	Dice	0.733±0.344	0.723±0.353	0.723±0.358	2.1e-02	1.9e-03
NETC	HD95	49.10±120.6	51.40±123.6	55.80±128.1	2.4e-01	2.9e-02
SNFH	Dice	0.705±0.295	0.690±0.304	0.674±0.312	4.6e-02	5.1e-05
SNFH	HD95	36.60±96.40	40.90±102.9	46.90±109.7	6.9e-01	6.6e-03
ET	Dice	0.693±0.251	0.684±0.256	0.678±0.255	7.2e-01	8.0e-05
ET	HD95	38.10±88.50	39.10±87.90	39.60±85.30	2.3e-01	1.1e-02
RC	Dice	0.915±0.245	0.919±0.237	0.894±0.279	2.2e-01	1.8e-05
RC	HD95	20.00±86.70	19.80±85.80	28.40±102.6	6.1e-01	2.0e-03
TC	Dice	0.722±0.264	0.714±0.269	0.705±0.268	8.7e-01	2.2e-05
TC	HD95	38.50±88.50	39.40±87.80	40.10±85.30	1.5e-01	9.3e-03
WT	Dice	0.745±0.260	0.740±0.261	0.734±0.262	9.7e-01	3.8e-05
WT	HD95	32.80±84.20	32.70±80.70	32.10±75.60	8.4e-02	1.9e-03
Task 2 — Pediatric Tumors — $n = 294$						
ET	Dice	0.679±0.362	0.670±0.367	0.518±0.396	6.4e-01	4.2e-02
ET	HD95	64.90±136.6	59.80±131.9	99.20±160.8	5.6e-01	3.4e-01
NET	Dice	0.771±0.250	0.782±0.246	0.765±0.264	2.0e-02	4.2e-02
NET	HD95	13.00±44.30	15.10±56.90	14.10±49.60	2.0e-01	2.1e-01
CC	Dice	0.687±0.431	0.659±0.440	0.616±0.459	2.7e-01	2.6e-01
CC	HD95	78.60±149.5	81.80±151.0	100.5±163.6	3.6e-01	5.8e-01
ED	Dice	0.796±0.380	0.795±0.381	0.758±0.411	6.1e-01	5.2e-01
ED	HD95	54.80±128.7	51.20±124.6	69.20±142.3	7.9e-01	2.4e-01
TC	Dice	0.847±0.193	0.855±0.194	0.845±0.201	2.1e-02	3.3e-02
TC	HD95	9.900±32.80	12.20±48.80	7.500±15.10	9.1e-02	8.3e-02
WT	Dice	0.867±0.178	0.875±0.178	0.875±0.163	2.6e-03	4.6e-02
WT	HD95	9.300±32.80	11.40±48.80	6.400±14.90	8.2e-03	2.1e-01
Task 3 — BraTS-GoAT — $n = 1,351$						
ET	Dice	0.862±0.201	0.861±0.205	0.869±0.190	1.4e-01	1.4e-08
ET	HD95	14.00±60.40	14.00±59.80	10.00±46.20	8.2e-02	1.0e-02
TC	Dice	0.907±0.162	0.904±0.171	0.906±0.162	2.7e-01	1.2e-11
TC	HD95	8.500±39.00	9.000±40.50	6.600±25.80	1.2e-01	1.1e-02
WT	Dice	0.923±0.115	0.924±0.116	0.922±0.126	5.9e-02	2.0e-04
WT	HD95	7.600±28.90	7.800±30.70	7.100±26.00	2.3e-03	2.4e-01

Table S1. Out-of-fold cross-validation metrics for all three methods.

Class	Metric	Baseline	MISFIT	Scratch	p B-vs-M	p M-vs-S
ET	Dice	0.646±0.308	0.546±0.311	0.560±0.305	3.9e-06	1.7e-01
ET	NSD	0.702±0.324	0.604±0.335	0.617±0.326	2.6e-05	3.1e-01
ET	HD95	82.35±119.5	120.6±127.3	111.3±123.4	5.0e-05	8.3e-01
ET	F1, all	0.819±0.249	0.732±0.296	0.763±0.270	6.7e-06	3.8e-01
ET	F1, small	0.453±0.368	0.421±0.344	0.428±0.338	1.4e-01	7.1e-01
ET	F1, large	0.725±0.376	0.699±0.353	0.705±0.349	3.1e-03	9.4e-01
RC	Dice	0.910±0.257	0.923±0.235	0.886±0.297	1.7e-01	4.2e-04
RC	NSD	0.904±0.265	0.917±0.245	0.877±0.308	1.1e-01	3.7e-04
RC	HD95	23.16±81.76	18.68±74.29	35.32±102.9	1.9e-01	2.6e-03
RC	F1, all	0.943±0.207	0.952±0.191	0.907±0.269	2.0e-01	3.1e-03
RC	F1, small†	0.000±0.000	0.000±0.000	0.000±0.000	–	–
RC	F1, large†	0.067±0.235	0.077±0.250	0.056±0.213	2.1e-01	2.3e-02
TC	Dice	0.673±0.306	0.570±0.319	0.585±0.312	8.9e-07	7.6e-02
TC	NSD	0.715±0.321	0.613±0.340	0.624±0.329	3.2e-06	1.0e-01
TC	HD95	76.11±115.5	117.2±127.4	107.7±123.1	3.2e-06	7.3e-01
TC	F1, all	0.835±0.237	0.741±0.295	0.772±0.267	8.6e-07	4.2e-01
TC	F1, small†	0.458±0.369	0.426±0.347	0.432±0.341	1.5e-01	5.8e-01
TC	F1, large†	0.738±0.371	0.706±0.353	0.712±0.349	9.6e-04	9.1e-01
WT	Dice	0.651±0.292	0.565±0.295	0.575±0.300	3.1e-05	1.6e-01
WT	NSD	0.663±0.294	0.582±0.302	0.587±0.300	1.6e-04	1.2e-01
WT	HD95	75.02±108.9	111.6±117.9	105.0±116.9	2.4e-04	3.6e-01
WT	F1, all	0.819±0.248	0.738±0.279	0.767±0.264	4.6e-06	5.2e-01
WT	F1, small†	0.374±0.379	0.345±0.358	0.355±0.350	1.4e-01	9.0e-01
WT	F1, large†	0.752±0.350	0.724±0.328	0.717±0.339	4.4e-03	4.2e-01

Table S2. Complete official Task 1 held-out validation metrics ($n = 179$), scored lesion-wise.

Class	Metric	Baseline	MISFIT	Scratch	p B-vs-M	p M-vs-S
Task 2 — Pediatric Tumors, $n = 91$						
ET	Dice	0.657±0.404	0.633±0.416	0.545±0.409	4.0e-01	6.9e-01
ET	NSD	0.578±0.385	0.565±0.392	0.458±0.359	5.4e-01	3.9e-01
ET	HD95	68.14±140.9	88.10±156.1	109.4±166.8	2.5e-01	3.1e-01
ET	F1, all	0.666±0.387	0.618±0.407	0.506±0.386	1.9e-01	2.9e-02
NET	Dice	0.907±0.159	0.921±0.102	0.915±0.102	6.1e-01	2.8e-04
NET	NSD	0.636±0.176	0.639±0.161	0.611±0.161	6.8e-01	7.8e-04
NET	HD95	7.756±39.01	3.615±6.100	3.923±6.734	4.8e-01	7.8e-01
CC	Dice	0.566±0.493	0.587±0.485	0.570±0.485	3.1e-01	4.4e-01
ED	Dice	0.846±0.361	0.835±0.371	0.835±0.371	3.2e-01	–
TC	Dice	0.934±0.150	0.947±0.088	0.943±0.083	6.6e-01	2.6e-04
TC	NSD	0.662±0.190	0.663±0.177	0.635±0.182	7.9e-01	9.6e-04
TC	HD95	7.039±39.12	3.010±6.423	3.134±6.442	2.2e-01	1.5e-03
WT	Dice	0.934±0.150	0.947±0.088	0.943±0.083	6.6e-01	2.6e-04
WT	NSD	0.662±0.190	0.663±0.178	0.636±0.182	7.9e-01	9.6e-04
WT	HD95	7.039±39.12	3.010±6.427	3.133±6.440	2.3e-01	1.5e-03
Task 3 — BraTS-GoAT, $n = 450$						
ET	Dice	0.777±0.305	0.760±0.309	0.760±0.310	2.2e-02	2.8e-05
ET	NSD	0.551±0.277	0.524±0.270	0.525±0.274	9.6e-04	1.7e-05
ET	HD95	40.53±108.9	43.09±112.1	43.03±111.1	8.5e-01	1.4e-04
ET	F1, all	0.674±0.362	0.643±0.364	0.632±0.357	4.5e-04	8.0e-02
TC	Dice	0.792±0.301	0.800±0.277	0.795±0.284	8.9e-01	7.7e-06
TC	NSD	0.485±0.290	0.480±0.278	0.474±0.281	2.4e-01	6.2e-06
TC	HD95	30.93±91.04	23.71±76.85	24.73±77.03	8.6e-02	3.2e-03
TC	F1, all	0.713±0.342	0.712±0.324	0.684±0.329	3.5e-01	6.6e-04
WT	Dice	0.853±0.227	0.863±0.213	0.858±0.217	9.1e-05	1.1e-04
WT	NSD	0.453±0.211	0.465±0.211	0.460±0.212	4.1e-04	1.1e-03
WT	HD95	22.21±75.06	17.03±63.21	18.87±65.71	7.7e-07	3.8e-05
WT	F1, all	0.654±0.311	0.663±0.307	0.645±0.315	5.9e-01	6.0e-02

Table S3. Complete official held-out validation metrics for Tasks 2 and 3, which use the global per-voxel protocol.

	Pocket nnUNet (baseline)	SwinUNETR-V2 (MISFIT)	SwinUNETR-V2 (from scratch)
Parameters	0.77 M	72.8 M	72.8 M
Patch size	128^3	96^3	96^3
Training, s/step	0.137	0.186	0.186
Inference, s/case, Task 1	7.00	10.03	10.01
Inference, s/case, Task 2	5.37	9.99	10.00
Inference, s/case, Task 3	3.52	5.79	5.64

Table S4. Computational cost of each method, measured on a single NVIDIA A100-SXM4-80GB (bf16, batch size 2). Training is one forward and backward pass; inference is end-to-end per case, including sliding-window prediction and test-time augmentation for one fold model.