BrainAnytime

Anatomy-Aware Cross-Modal Pretraining for Brain Image Analysis with Arbitrary Modality Availability

Guangqian Yang[1*], Tong Ding[1*], Wenlong Hou[1], Yue Xun[1], Ye Du[1], Qian Niu[2†], Shujun Wang[1,3†]
and for the Alzheimer's Disease Neuroimaging Initiative

[1] Biomedical Engineering, The Hong Kong Polytechnic University [2] Technology Management for Innovation, The University of Tokyo
[3] Data Science and Artificial Intelligence, The Hong Kong Polytechnic University
† Correspondence: qian.niu@weblab.t.u-tokyo.ac.jp · shu-jun.wang@polyu.edu.hk * Equal contribution

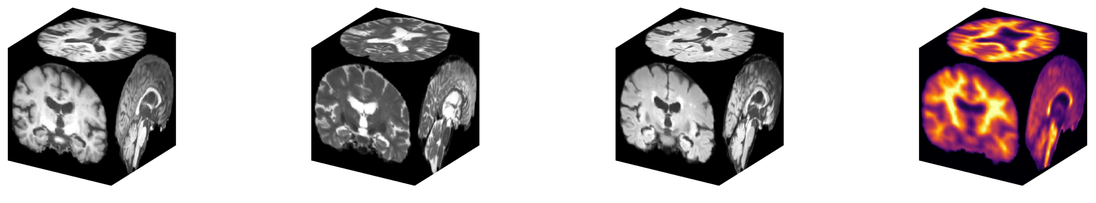
Code & resources

MICCAI 2026
Oral presentation

Inputs and clinical motivation

Research problem

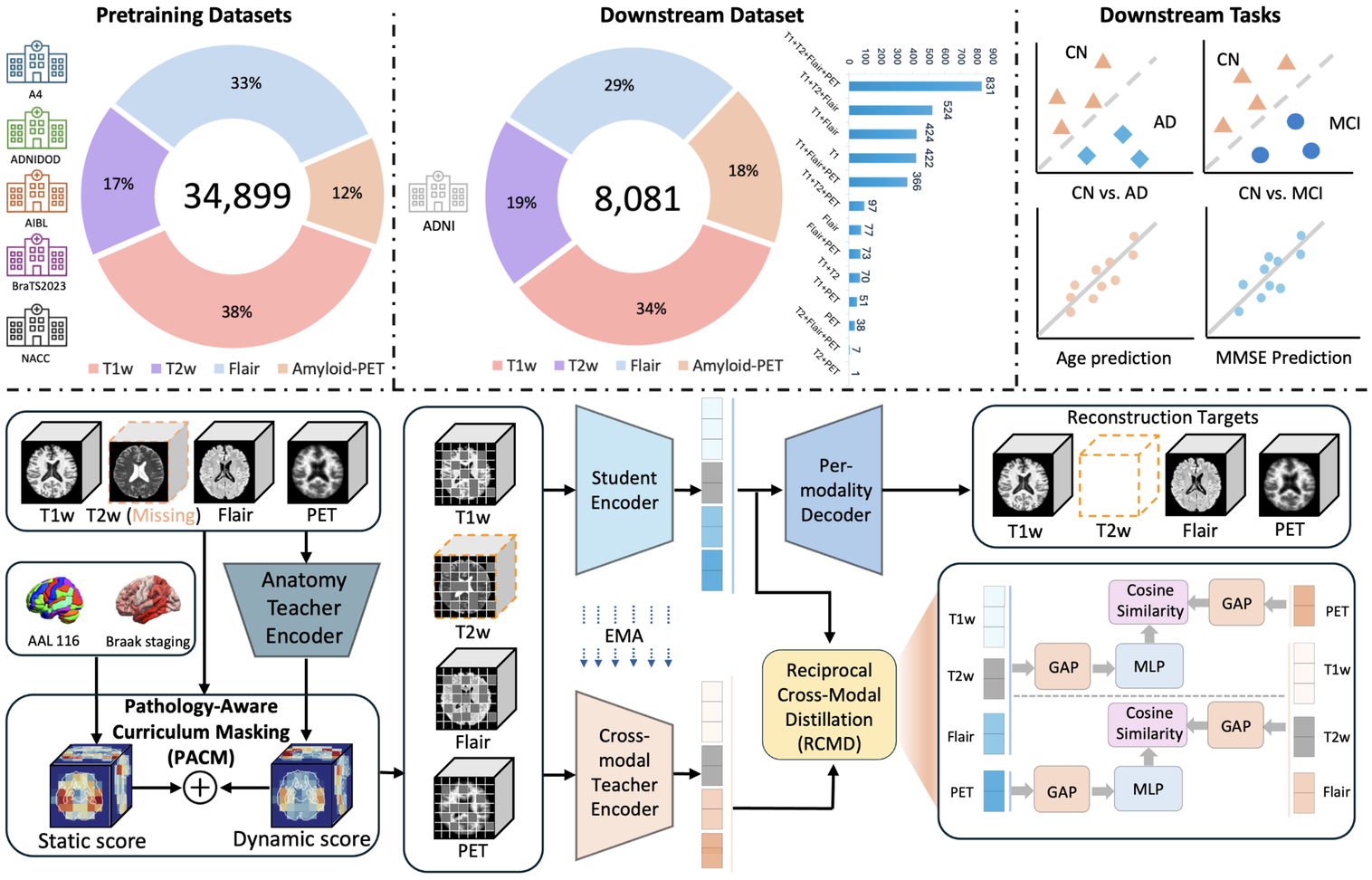
Clinical imaging availability varies: routine T1, additional T2 / FLAIR, and selective amyloid-PET.



Pretrain → fine-tune

Transfer encoder; train task heads.
CN vs. AD · CN vs. MCI
MMSE · age regression

Pretraining architecture and downstream transfer



Learning design and data

Multi-MAE3D

Observed volumes supply patch tokens to a shared 3D encoder. Missing inputs supply no visible tokens.

RCMD

Pooled MRI-group ↔ amyloid-PET feature prediction with EMA teacher targets; paired inputs only.

PACM

Anatomical priors + teacher attention. High-importance regions are progressively masked more often.

Pretraining cohorts

Cohort	Subjects	Scans
A4	1,736	6,710
DOD-ADNI	543	783
AIBL	1,408	3,161
BraTS23	2,569	7,707
NACC	7,244	16,538
Total	13,500	34,899

Independent ADNI: 2,981 subjects. 60 / 10 / 30 split for common groups; rare groups train / val only.

Four baseline families

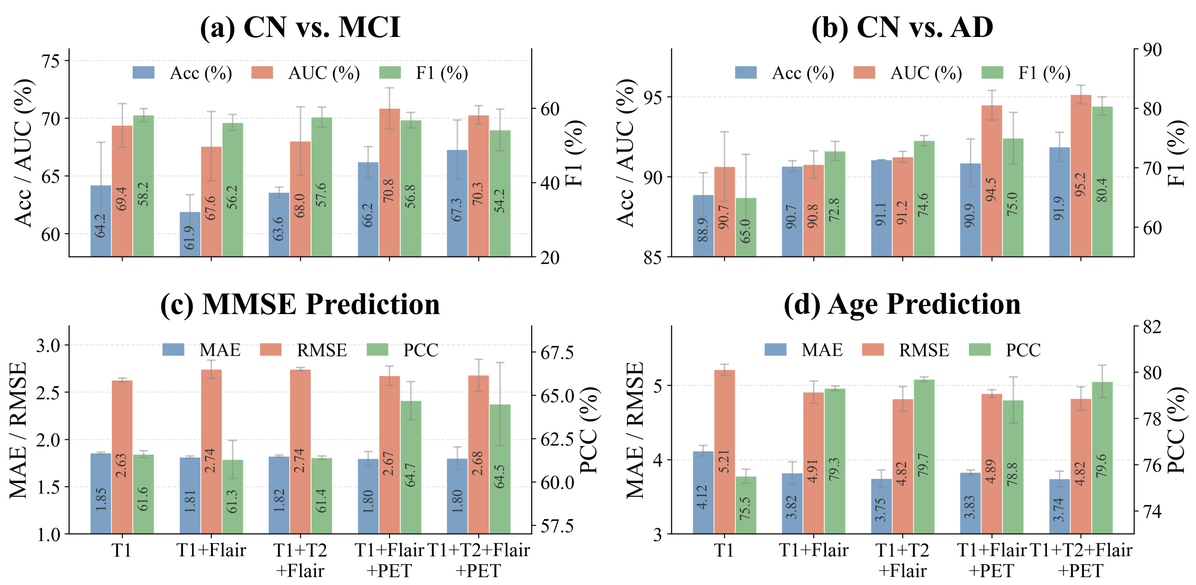
Family	Methods
Single-modal	3D ResNet50 · M3T
Dual-modal fusion	MCAD · MENet
Brain foundation	BrainIAC · SAM-Brain3D
Missing-modality	Flex-MoE · FuseMoE · MoE-retriever

Main results across four downstream tasks

Method	CN vs. AD ↑			CN vs. MCI ↑			MMSE			Age		
Method	ACC	AUC	F1	ACC	AUC	F1	MAE ↓	RMSE ↓	PCC ↑	MAE ↓	RMSE ↓	PCC ↑
Flex-MoE	84.2 (2.3)	91.0 (1.9)	74.8 (3.5)	58.9 (4.5)	63.2 (4.8)	40.5 (12.4)	2.452 (0.260)	3.304 (0.300)	42.4 (3.6)	5.015 (0.461)	6.297 (0.564)	70.3 (2.7)
FuseMoE	82.1 (1.6)	87.5 (0.9)	67.8 (3.8)	62.8 (2.3)	67.4 (1.7)	54.2 (3.3)	2.400 (0.160)	3.384 (0.220)	38.7 (5.6)	4.728 (0.103)	6.041 (0.103)	65.9 (1.4)
MoE-retriever	83.1 (4.7)	87.2 (6.0)	74.7 (6.3)	60.6 (2.1)	63.6 (4.1)	58.5 (5.5)	2.065 (0.053)	2.868 (0.093)	52.3 (6.8)	4.502 (0.461)	5.682 (0.513)	71.5 (6.1)
BrainAnytime	**89.4** (1.9)	**94.7** (1.2)	**82.4** (3.7)	**67.2** (2.4)	**72.5** (1.9)	**63.9** (4.6)	**1.984** (0.056)	**2.768** (0.092)	**60.2** (2.6)	**4.010** (0.133)	**5.158** (0.154)	**77.2** (1.1)

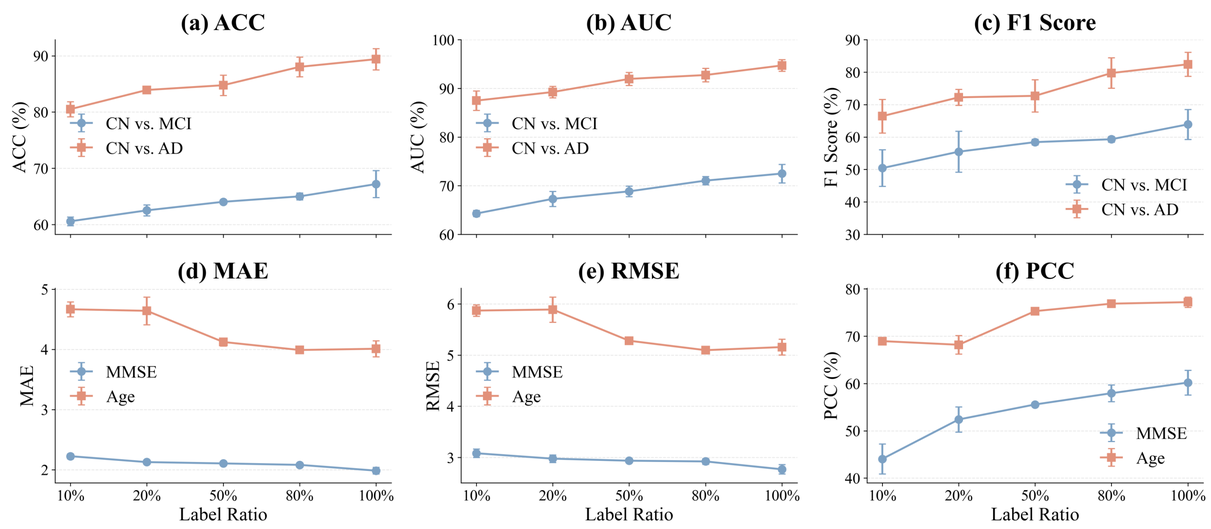
Mean (SD) across five real modality groups; classification / PCC in %. Bold: best displayed mean. Groups contain different subjects.

Controlled modality dropout



Same subjects in every condition. AD F1: 65.0 → 80.4.

Performance with fewer training labels



BrainAnytime at 10–100% labels. This varies supervision within the pretrained model.

Results by available inputs

Method	T1	T1+F	T1+T2+F	T1+F+P	All four
3D ResNet50	93.1	—	—	—	—
M3T	81.2	—	—	—	—
MCAD	—	93.9	—	—	—
MENet	—	95.4	—	—	—
BrainIAC	82.5	87.8	82.1	—	—
SAM-Brain3D	91.7	95.3	85.8	—	—
Flex-MoE	81.7	95.5	88.8	95.3	93.6
FuseMoE	80.2	94.3	83.0	92.0	87.9
MoE-retriever	72.0	93.9	86.9	92.5	90.6
BrainAnytime	**93.4**	**97.0**	**90.6**	**97.5**	**95.2**

CN vs. AD AUC (%) ↑; means over 3 seeds. All nine baselines, grouped as above. —: not reported. F: FLAIR; P: amyloid-PET.

Compare within columns: groups have distinct subjects. SDs: Table 2 / accompanying data.

Component ablation

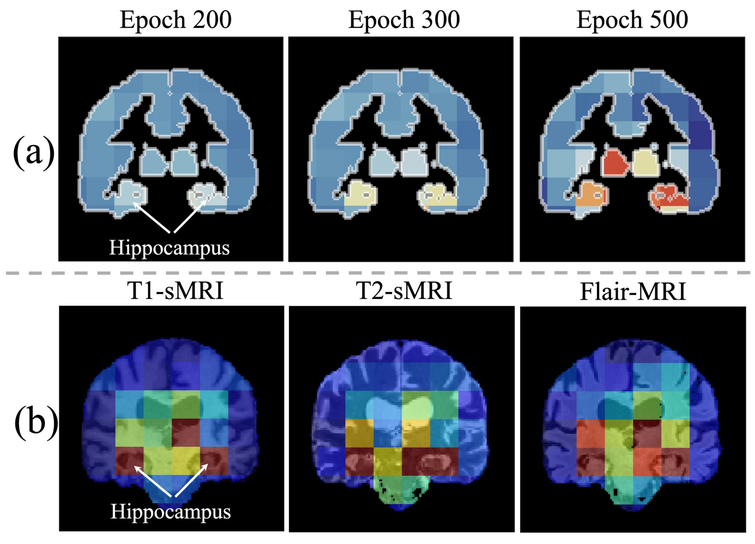
Pretraining	AD ACC	AD AUC	MCI ACC
ViT-B (scratch)	78.8 (2.2)	84.8 (1.1)	59.3 (2.1)
MAE3D	86.9 (0.4)	92.8 (1.9)	64.4 (1.8)
MAE3D+RCMD	88.8 (0.3)	94.1 (0.9)	65.7 (0.6)
MAE3D+PACM	88.1 (1.6)	94.8 (0.5)	66.7 (1.1)
BrainAnytime	89.4 (0.5)	94.7 (0.5)	67.2 (1.4)

What the components add

Full vs. MAE3D:
+2.5 / +2.8 pp AD / MCI ACC

Mean (SD), separate experiment. Full model leads in ACC; PACM alone gives the highest AD AUC (94.8).

Anatomical attention



Scope: retrospective cohort evidence; prospective clinical utility remains to be established. MRI / PET images are teaching illustrations, not a diagnostic case.