BrainMoE: Cognition Joint Embedding via Mixture-of-Expert Towards Robust Brain Foundation Model

Ziquan Wei, Tingting Dan, Tianlong Chen, Guorong Wu*

UNC Chapel Hill, USA

NeurIPS 2025

Motivation: Why we need a BrainMoE

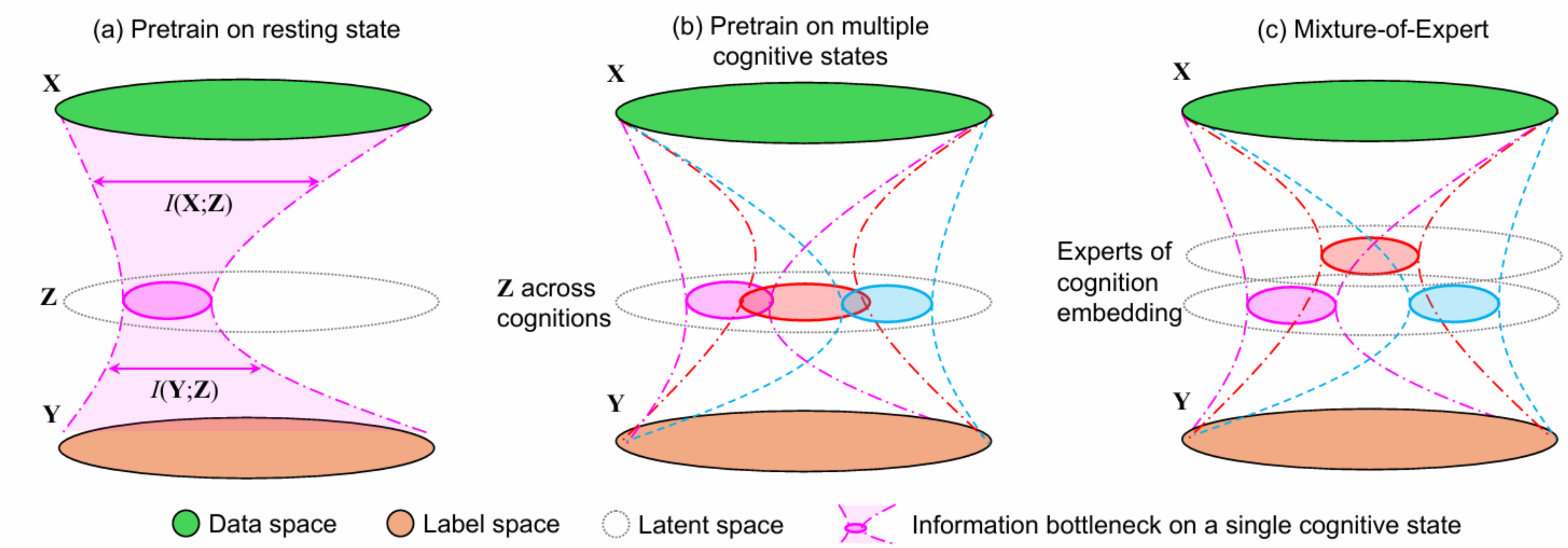


Figure 1: Motivation of BrainMoE through the lens of the information bottleneck theory. (a) Feature representation learning makes an information bottleneck between data and label space, where $\mathbf{X}$ denotes data, $\mathbf{Z}$ denotes latent feature representation, $\mathbf{Y}$ is the label of data, and $I(\cdot;\cdot)$ is the mutual information. (b) A model pre-trained on multiple cognitive states may compromise the underlying heterogeneity between different states, where $\mathbf{Z}$ cannot be optimal for all states. (c) The mixture of brain experts dedicated to diverse cognitive states leads to a joint cognition embedding so that the downstream applications can be advanced by stratified pre-training on rich behavioral tasks.

More data != more performance

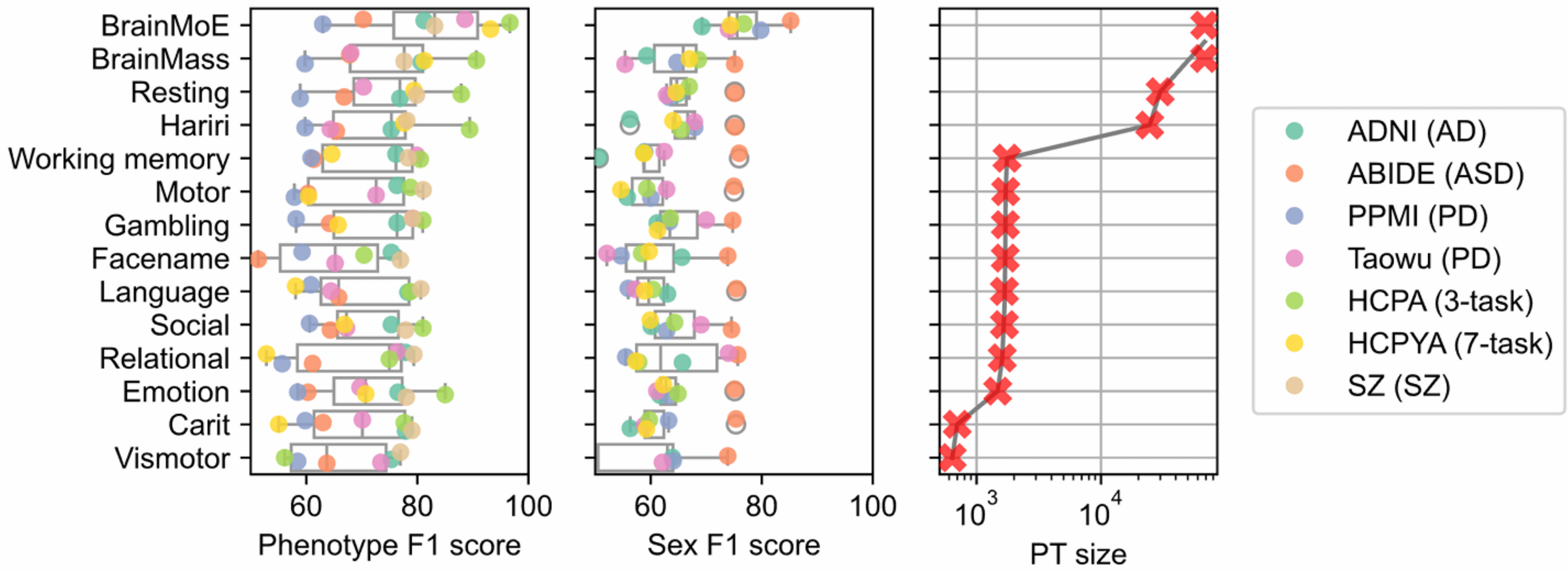
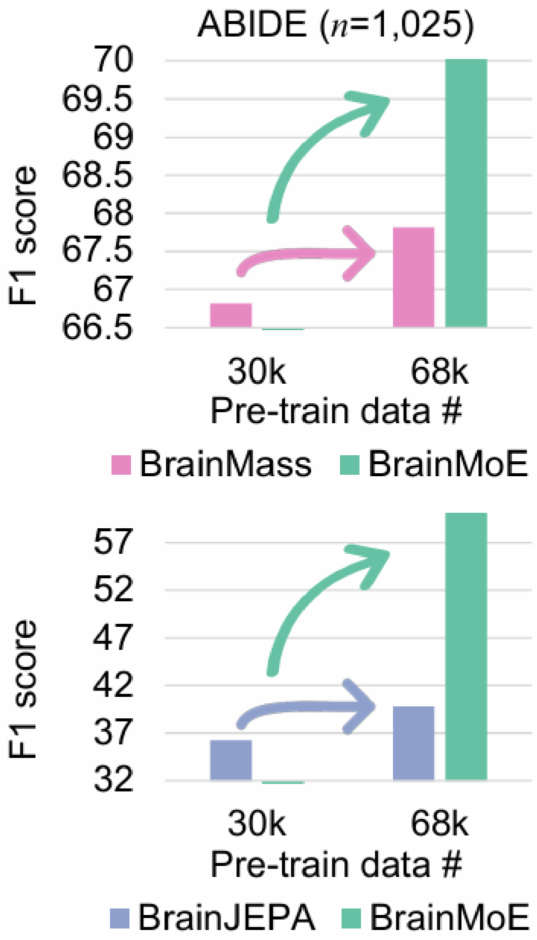
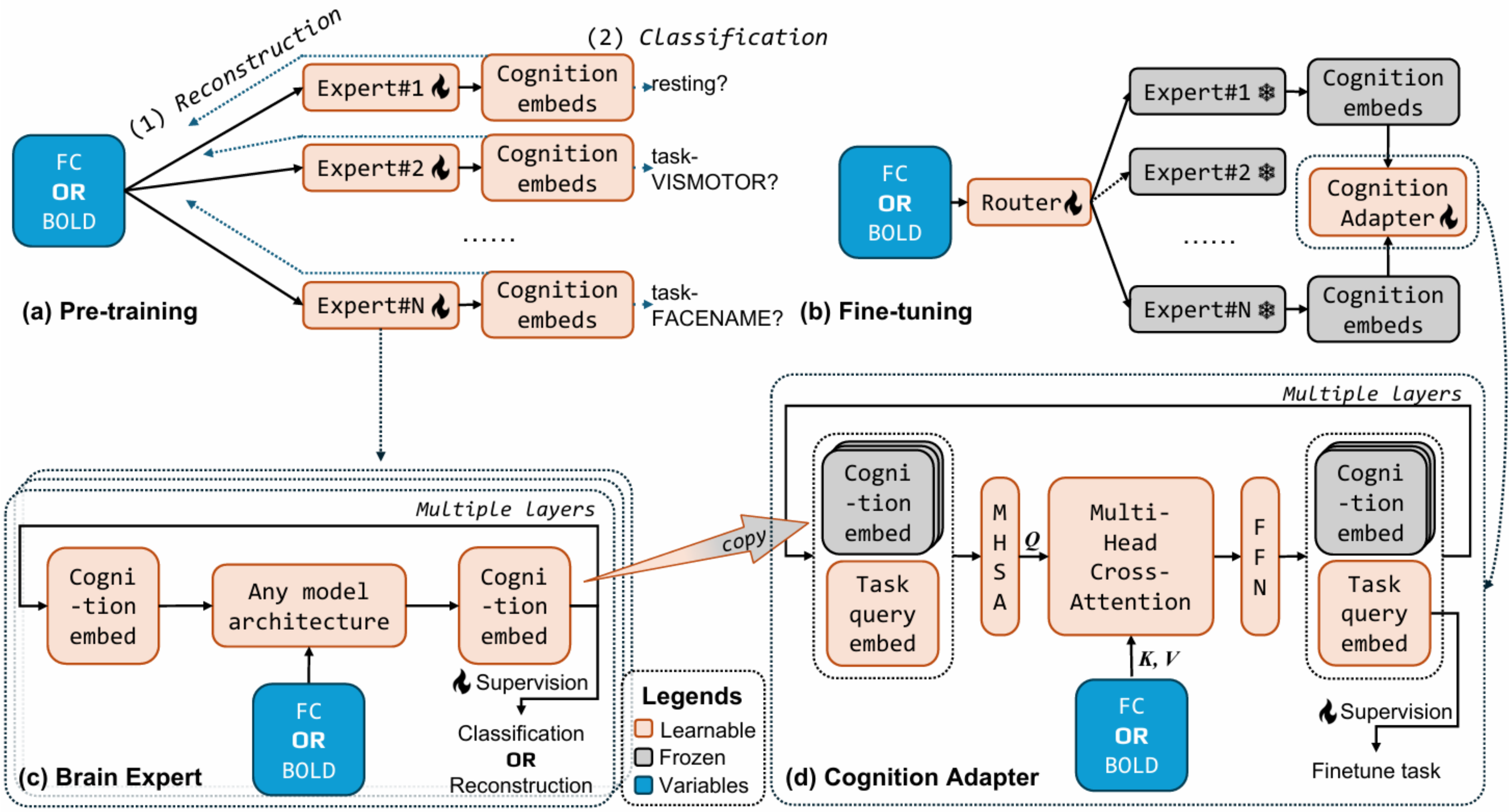


Figure 4: The performance of BrainMoE and BrainMass pre-trained with all samples, and n=12 pre-trained individual experts on phenotypic and sex classifications among 7 datasets, where scores lower than 50% are hidden for clarity, and the pre-training (PT) size ranges from 68,251 to 637.



Framework of BrainMoE



Accuracy scores: Phenotypic prediction

Table 1: MoE improvement on phenotypic classification F1 score compared to the baseline, where 30k is pre-trained on resting-state data (n=29,951), and 68k is pre-trained on all data (n=68,251). PT stands for pre-training. Colored text indicates the performance increase/decrease from using 68k.

	BrainMass					BrainJEPA		
Predictor PT #	SVM 30k	SVM 68k	MLP 30k	MLP 68k	BrainMoE 68k	ViT 30k	ViT 68k	BrainMoE 68k
ADNI	$75.32_{\pm 7.06}$	$75.32_{\pm 7.06}$	$76.86_{\pm 7.26}$	$80.70_{\pm 7.85}$	$81.23_{\pm 11.00}$	$74.16_{\pm 8.55}$	$74.16_{\pm 8.55}$	$77.11_{\pm 6.64}$
↳AD		**0.00**		**3.84 ↑**	**4.37 ↑**		**0.00**	**2.95 ↑**
ABIDE	$62.31_{\pm 1.95}$	$64.12_{\pm 2.31}$	$66.81_{\pm 4.18}$	$67.81_{\pm 3.91}$	$70.26_{\pm 3.40}$	$36.25_{\pm 6.93}$	$39.82_{\pm 3.91}$	$54.55_{\pm 9.89}$
↳ASD		**1.81 ↑**		**1.00 ↑**	**3.45 ↑**		**3.77 ↑**	**18.30 ↑**
PPMI	$54.87_{\pm 15.76}$	$56.52_{\pm 14.86}$	$58.90_{\pm 14.29}$	$59.77_{\pm 14.22}$	$62.97_{\pm 13.94}$	$38.69_{\pm 13.91}$	$38.69_{\pm 13.91}$	$60.49_{\pm 11.59}$
↳PD (staged)		**1.65 ↑**		**0.87 ↑**	**4.07 ↑**		**0.00**	**21.80 ↑**
Taowu	$58.33_{\pm 34.78}$	$65.67_{\pm 20.55}$	$70.29_{\pm 17.97}$	$68.00_{\pm 21.46}$	$88.57_{\pm 12.51}$	$36.08_{\pm 26.38}$	$36.94_{\pm 20.98}$	$79.86_{\pm 14.46}$
↳PD (binary)		**7.34 ↑**		**2.29 ↓**	**18.28 ↑**		**0.86 ↑**	**43.78 ↑**
SZ	$76.95_{\pm 9.01}$	$76.95_{\pm 9.01}$	$79.85_{\pm 8.69}$	$77.63_{\pm 8.56}$	$83.10_{\pm 11.33}$	$76.98_{\pm 9.00}$	$78.97_{\pm 9.79}$	$82.86_{\pm 9.19}$
↳Schizophrenia		**0.00**		**2.22 ↓**	**3.25 ↑**		**1.99 ↑**	**5.88 ↑**
HCPA	$85.16_{\pm 0.41}$	$89.73_{\pm 0.58}$	$87.91_{\pm 0.48}$	$90.63_{\pm 0.74}$	$96.67_{\pm 0.77}$	$59.54_{\pm 15.47}$	$53.12_{\pm 14.19}$	$81.74_{\pm 0.51}$
↳3-task,rest		**4.57 ↑**		**2.72 ↑**	**8.76 ↑**		**6.42 ↓**	**22.20 ↑**
HCPYA	$77.51_{\pm 2.42}$	$80.87_{\pm 1.77}$	$79.40_{\pm 1.78}$	$81.27_{\pm 1.27}$	$93.19_{\pm 0.72}$	$50.68_{\pm 25.20}$	$56.10_{\pm 29.16}$	$74.59_{\pm 3.79}$
↳7-task		**3.36 ↑**		**1.87 ↑**	**13.79 ↑**		**5.42 ↑**	**23.91 ↑**

Accuracy scores: Sex classification

Table 2: MoE improvement on sex classification F1 score compared to the baseline, where the sample size of the downstream dataset is indicated. PT stands for pre-training.

	BrainMass					BrainJEPA		
Predictor	SVM	SVM	MLP	MLP	BrainMoE	ViT	ViT	BrainMoE
PT #	30k	68k	30k	68k	68k	30k	68k	68k
ADNI	$48.60_{\pm 6.55}$	$54.30_{\pm 12.48}$	$64.82_{\pm 4.30}$	$59.30_{\pm 13.05}$	$69.22_{\pm 5.26}$	$37.70_{\pm 8.73}$	$36.42_{\pm 8.22}$	$62.98_{\pm 7.76}$
↳n=138		**5.70** $\uparrow$		**5.52** $\downarrow$	**4.40** $\uparrow$		**1.28** $\downarrow$	**25.28** $\uparrow$
ABIDE	$73.84_{\pm 3.49}$	$73.84_{\pm 3.49}$	$75.12_{\pm 5.27}$	$75.12_{\pm 6.32}$	$85.21_{\pm 3.77}$	$78.08_{\pm 5.84}$	$78.08_{\pm 5.84}$	$78.08_{\pm 6.15}$
↳n=1025		**0.00**		**0.00**	**10.09** $\uparrow$		**0.00**	**0.00**
PPMI	$52.03_{\pm 14.16}$	$56.58_{\pm 12.28}$	$63.32_{\pm 14.96}$	$64.73_{\pm 14.12}$	$79.81_{\pm 8.46}$	$46.23_{\pm 9.00}$	$46.23_{\pm 9.00}$	$67.57_{\pm 7.24}$
↳n=209		**4.55** $\uparrow$		**1.41** $\uparrow$	**16.49** $\uparrow$		**0.00**	**21.34** $\uparrow$
Taowu	$46.24_{\pm 23.96}$	$46.24_{\pm 23.96}$	$62.86_{\pm 28.20}$	$55.38_{\pm 20.93}$	$74.00_{\pm 27.79}$	$46.24_{\pm 23.97}$	$51.67_{\pm 21.12}$	$90.00_{\pm 12.96}$
↳n=40		**0.00**		**7.48** $\downarrow$	**11.14** $\uparrow$		**5.43** $\uparrow$	**43.76** $\uparrow$
HCPA	$66.20_{\pm 1.58}$	$68.25_{\pm 1.70}$	$66.93_{\pm 0.63}$	$68.58_{\pm 0.99}$	$76.76_{\pm 0.93}$	$40.32_{\pm 4.07}$	$40.32_{\pm 4.07}$	$44.24_{\pm 7.01}$
↳n=4863		**2.05** $\uparrow$		**1.65** $\uparrow$	**9.83** $\uparrow$		**0.00**	**3.92** $\uparrow$
HCPYA	$63.33_{\pm 3.01}$	$65.47_{\pm 3.51}$	$64.57_{\pm 2.53}$	$66.98_{\pm 3.30}$	$74.36_{\pm 4.43}$	$40.20_{\pm 4.22}$	$40.20_{\pm 4.22}$	$43.24_{\pm 8.19}$
↳n=3293		**2.14** $\uparrow$		**2.41** $\uparrow$	**9.79** $\uparrow$		**0.00**	**3.04** $\uparrow$

Performance interpretation

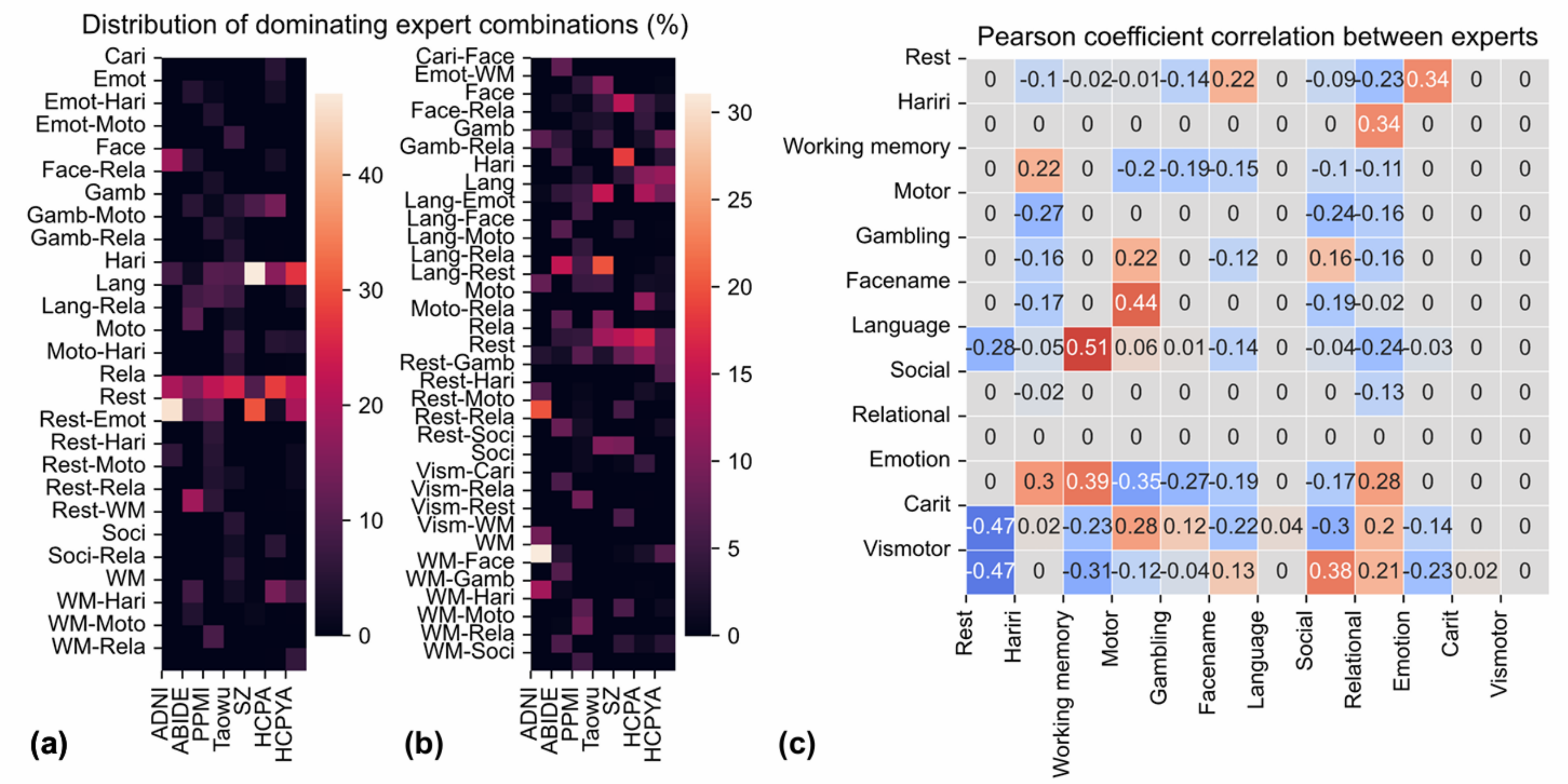


Figure 5: The distribution of dominating expert combinations of **(a)** late fusion MoE and **(b)** BrainMoE shows the router preference, where the first 4 letters of cognitive state are used as abbreviation. **(c)** The correlation between cognition embeddings of experts shows expert diversity.

More applications

Table 5: Age regression performance compared to the baseline, where the sample size of the downstream dataset is indicated, and unit is year.

MSE	ADNI	ABIDE	PPMI	Taowu	HCPA	HCPYA
BrainMass	$36.28_{\pm 19.83}$	$36.77_{\pm 17.2}$	$33.09_{\pm 21.87}$	$38.10_{\pm 28.33}$	$22.66_{\pm 7.51}$	$5.46_{\pm 2.69}$
BrainMoE	$\mathbf{36.27}_{\pm 10.23}$	$\mathbf{4.86}_{\pm 2.67}$	$\mathbf{29.89}_{\pm 8.39}$	$\mathbf{30.72}_{\pm 12.25}$	$\mathbf{10.56}_{\pm 2.86}$	$\mathbf{3.45}_{\pm 1.08}$

Table 6: BrainMoE applies on a multimodal dataset, NATVIEW [27].

NATVIEW	8-task (F1)	Sex (F1)	Age (MSE)
BrainMass (fMRI)	$67.66_{\pm 5.74}$	$63.67_{\pm 5.16}$	$8.05_{\pm 5.58}$
CBraMod (EEG)	$68.71_{\pm 1.46}$	$65.39_{\pm 2.33}$	$8.26_{\pm 5.97}$
BrainMoE (13 Ex.)	$\mathbf{68.73}_{\pm 3.72}$	$\mathbf{65.47}_{\pm 5.38}$	$\mathbf{7.99}_{\pm 5.53}$

Ablation studies