A Hierarchical Multi-Task Framework for Dementia Diagnosis via Pathological Feature Learning from Multi-Organ Data

Shilun Zhao[1†], Fan Li[1,2†], Zhichao Liang[1,3†], Kaicong Sun[1], Shuwei Bai[1], Weilin Zhou[1], Xin Lin[1], Zihao Wang[4], and Dinggang Shen[1,5,6(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[2] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, Shanghai, China
[3] Faculty of Applied Sciences, Macao Polytechnic University, Macao, China
[4] School of Computer Science, The University of Sydney, Sydney, Australia
[5] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[6] Shanghai Clinical Research and Trial Center, Shanghai, China
Dinggang.Shen@gmail.com

Abstract. Current deep learning methods for dementia diagnosis are predominantly brain-centric, neglecting the disease's systemic nature. Furthermore, these models fail to decouple disease-specific alterations from normal aging, and often limit diagnosis to a single etiology, neglecting the reality of co-occurring conditions. To address these limitations, we propose MoMmNet, a hierarchical multi-task framework integrating multi-organ data (brain, heart, gut, liver, and kidney) and clinical text for pathology-aware reasoning. MoMmNet introduces three key components: 1) a Feature Alignment (FA) module that aligns peripheral and text features into brain-anchored semantic space; 2) a Pathological Feature Decoupling (PFD) module that utilizes a normative feature memory bank to reduce demographic confounders and isolate pathology-specific representations; and 3) a Progressive Multi-Etiology Inference (PMI) module that first detects cognitive impairment and then employs parallel binary classifiers to identify single or co-occurring etiologies. Evaluated on a large-scale, multi-center cohort of 15,016 subjects, MoMmNet outperforms state-of-the-art methods, achieving an AUC of 0.828 for cognitive staging. Crucially, it delivers robust multi-label predictions across 10 etiologies, demonstrating high efficacy for single etiology (e.g., AUC of 0.764 for Alzheimer's disease) and multiple co-occurring etiologies (e.g., AUC of 0.609 for co-occurring Alzheimer's disease and Lewy body dementia) among cognitively impaired subjects. These results demonstrate the value of systemic modeling, confounder decoupling, and multi-etiology reasoning for precision dementia diagnosis. Our code is publicly available at https://github.com/AIbySlz/MoMmNet.

Keywords: Dementia diagnosis · Multi-organ learning · Pathological feature decoupling · Multi-etiology inference.

† Equal contribution.

1 Introduction

Dementia is a progressive neurodegenerative syndrome causing severe cognitive and daily functional decline, presenting immense socioeconomic challenges [18]. It currently affects approximately 55 million people worldwide, and the total number of people living with dementia will be projected to 153 million by 2050 [3]. Although Alzheimer's disease (AD) is the most common etiology, other prevalent etiologies, such as vascular dementia (VD), Lewy body dementia (LBD), and frontotemporal dementia (FTD), account for a considerable proportion of cases [8, 11, 15]. Accurate differentiation among these subtypes remains challenging due to overlapping symptoms and the frequent presence of mixed pathologies.

Growing evidence suggests that dementia is not solely a brain disorder but a systemic condition involving dysfunction across multiple peripheral organs [1]. Specifically, cardiovascular dysfunction induces cerebral hypoperfusion, which in turn obstructs the systemic clearance of neurotoxic metabolites [5]. Gut microbiota dysbiosis may trigger inflammatory responses that accelerate cognitive decline [4], while liver dysfunction reduces peripheral clearance of amyloid beta [2]. Notably, these systemic alterations often emerge before detectable structural changes in the brain [10]. Despite this understanding, studies that use multi-organ data for dementia classification remain limited.

Deep learning has demonstrated strong potential in dementia diagnosis [19, 20, 7, 22]. For instance, Xue et al. [21] achieved promising performance by developing a multi-modal network that integrates neuroimaging representations with clinical text embeddings. In the emerging area of multi-organ analysis, Li et al. [13, 12, 14] proposed a deep knowledge distillation framework to extract and fuse latent features from brain, heart, and gut data for AD identification, providing preliminary evidence that systemic representations can enhance AD diagnosis. Despite these advances, current approaches face two critical limitations. First, existing models rarely account for aging-related confounders. Since both normative aging and dementia cause structural atrophy and physiological decline [16], models often struggle to distinguish age-related changes from pathological degeneration. This confounding can lead to misinterpretation of normative aging effects as disease-specific biomarkers, thereby reducing diagnostic specificity. Second, existing frameworks typically treat dementia diagnosis as a mutually exclusive classification task [17]. Consequently, they often limit diagnosis to a single etiology, ignoring the clinical reality of multiple co-occurring etiologies.

To address these challenges, we propose MoMmNet, a multi-organ and multi-modal network for systemic dementia reasoning. MoMmNet aligns peripheral (heart, gut, liver, and kidney) and clinical text features into a unified, brain-anchored space. To mitigate biological confounding, we introduce a Pathological Feature Decoupling (PFD) module that utilizes a normative feature memory bank and Z-score normalization to isolate disease-specific patterns from physiological variations. Furthermore, we develop a Progressive Multi-Etiology Inference (PMI) module for diagnostic inference. This module first performs cognitive

staging, and then deploys parallel binary classifiers exclusively for symptomatic subjects with mild cognitive impairment (MCI) or dementia (DE) to quantify co-occurring etiologies. Experiments on a large-scale dataset of 15,016 subjects demonstrate that MoMmNet significantly outperforms state-of-the-art methods. The main contributions of this work are: 1) A novel framework integrating multi-organ data (brain, heart, gut, liver, and kidney) for comprehensive systemic characterization of dementia pathology; 2) A memory bank-based PFD module disentangling pathological features from demographic confounders using statistical priors; and 3) A hierarchical PMI module extending conventional classification paradigms to accurately identify co-occurring etiologies in mixed dementia.

2 Methods

As shown in Fig. 1, MoMmNet integrates image data from the brain and peripheral organs (heart, gut, liver, and kidney) with clinical text. The framework is composed of three key components: Feature Alignment (FA), Pathological Feature Decoupling (PFD), and Progressive Multi-Etiology Inference (PMI). More details are described below.

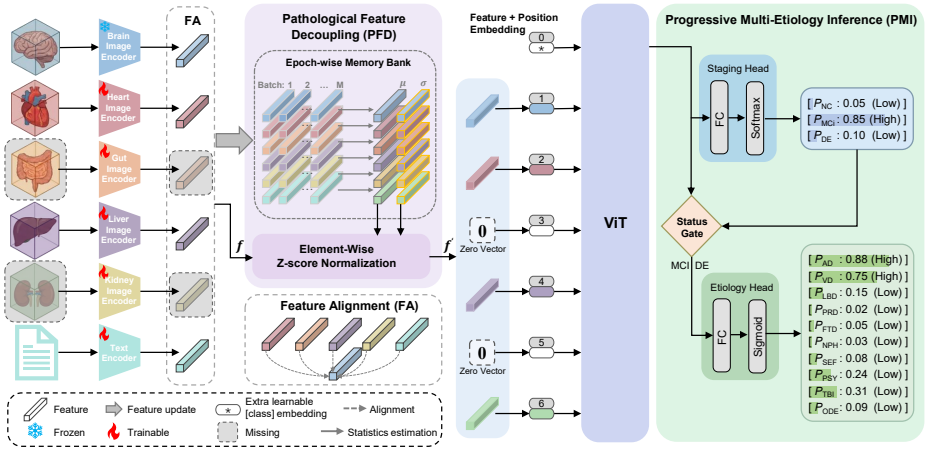

Fig. 1. Overview of the proposed MoMmNet framework. Multi-organ images and clinical text are processed by specific encoders. The FA module aligns peripheral and text features into a brain-anchored semantic space. PFD isolates pathology-specific representations, which are then fused via ViT. Finally, the fused features are processed by PMI, which performs hierarchical diagnosis of cognitive stage and etiology for subjects.

2.1 Feature Alignment (FA)

MoMmNet employs organ-specific encoders to extract latent features from the imaging data of five distinct organs, alongside a text encoder for clinical records.

Since the brain exhibits the most direct pathological alterations in dementia, brain representations serve as central anchors to establish a unified semantic space. By aligning peripheral imaging and clinical text features into the brain-anchored space prior to fusion, MoMmNet enriches neural representations with systemic pathological information unobservable in isolated brain scans.

In our study, since only FDG and Dixon imaging provide whole-body coverage, our organ-level alignment strictly focuses on these two shared modalities. We define the set of peripheral organs as $\mathcal{O}_{\text{peri}} = \{\text{heart, gut, liver, kidney}\}$ and the set of shared whole-body modalities as $\mathcal{M} = \{\text{FDG, Dixon}\}$. For the i-th subject, let $\mathbf{f}_i^{o,m}$ denote features of a peripheral organ $o \in \mathcal{O}_{\text{peri}}$ under modality $m \in \mathcal{M}$, and let $\mathbf{f}_i^{\text{brain},m}$ denote the corresponding brain features of the same modality. For a mini-batch of size N, the alignment loss between a peripheral organ o and the brain under a specific modality m is formulated as:

$$\mathcal{L}_{o,m} = -\frac{1}{2N}\sum_{i=1}^{N}\left[\log\frac{\exp\left(\text{sim}(\mathbf{f}_i^{\text{brain},m}, \mathbf{f}_i^{o,m})/\tau\right)}{\sum_{j=1}^{N}\exp\left(\text{sim}(\mathbf{f}_i^{\text{brain},m}, \mathbf{f}_j^{o,m})/\tau\right)} + \log\frac{\exp\left(\text{sim}(\mathbf{f}_i^{o,m}, \mathbf{f}_i^{\text{brain},m})/\tau\right)}{\sum_{j=1}^{N}\exp\left(\text{sim}(\mathbf{f}_i^{o,m}, \mathbf{f}_j^{\text{brain},m})/\tau\right)}\right] \quad (1)$$

where $\text{sim}(\cdot,\cdot)$ denotes cosine similarity, and the temperature parameter τ is set to 0.01 to control feature distribution sharpness.

Similarly, let $\mathbf{f}_i^{\text{text}}$ and $\mathbf{f}_i^{\text{brain}}$ denote the clinical text and comprehensive brain features, respectively. The text-to-brain alignment loss $\mathcal{L}_{\text{text}}$ follows the same symmetric InfoNCE formulation as Eq. (1). The overall alignment objective aggregates contributions from all peripheral organs and clinical text:

$$\mathcal{L}_{\text{align}} = \sum_{o\in\mathcal{O}_{\text{peri}}}\sum_{m\in\mathcal{M}}\lambda_{o,m}\mathcal{L}_{o,m} + \lambda_{\text{text}}\mathcal{L}_{\text{text}} \quad (2)$$

where $\lambda_{o,m}$ and λ_{text} are balancing weights controlling the contribution of each respective alignment task.

2.2 Pathological Feature Decoupling (PFD)

The latent features extracted by the encoders inherently entangle pathological indicators with demographic-driven physiological variations (e.g., age and sex). To mitigate these confounders and emphasize pathology-relevant representations, the PFD module standardizes features against a normative baseline.

Let $\mathcal{V}$ denote the set of input views, including all organ modalities and clinical text. For subject i with demographic profile c_i (e.g., age and sex), let $\mathbf{f}_i^v \in \mathbb{R}^D$ denote the feature vector extracted from view $v \in \mathcal{V}$, where $\mathbb{R}^D$ represents the D-dimensional feature space. To mitigate demographic confounding, we maintain an epoch-updated normative memory bank that stores the features of Normal Cognition (NC) subjects. For each subject i, a normative reference set is retrieved from this bank by matching the demographic profile c_i. To ensure statistical stability in sparse demographic subgroups, we employ a dynamic sliding-window strategy that progressively expands the age range until at least $N = 30$ matched NC samples are obtained. Based on this reference set, we compute the conditional mean vector $\boldsymbol{\mu}^{v|c_i}$ and standard deviation vector $\boldsymbol{\sigma}^{v|c_i}$. The feature vector $\mathbf{f}_i^v$ is

then transformed into a profile-conditioned deviation representation through Z-score normalization. Specifically, for the d-th feature dimension, the decoupled feature $\tilde{f}_i^v(d)$ is defined as:

$$\tilde{f}_i^v(d) = \frac{f_i^v(d) - \mu^{v|c_i}(d)}{\sigma^{v|c_i}(d)} \tag{3}$$

By quantifying dimension-wise deviations from a demographically matched baseline, $\tilde{\mathbf{f}}_i^v$ strictly isolates pathological representations.

2.3 Progressive Multi-Etiology Inference (PMI)

To mirror clinical workflows, the PMI module adopts a hierarchical strategy. After fusing the decoupled features $\tilde{\mathbf{f}}_i^v$ from all views $v \in \mathcal{V}$ via ViT, the resulting global representation is directly fed into the Staging Head to categorize the cognitive stage (e.g., NC, MCI, or DE). For subjects identified with MCI or DE, the Etiology Head subsequently uses this global feature to infer the underlying etiology. To accommodate coexisting pathologies, the Etiology Head is formulated as a multi-label classifier consisting of K independent binary classifiers, where K represents the total number of etiology types ($K = 10$ in this work).

For a mini-batch of size N, the overall inference objective is defined as:

$$\mathcal{L}_{\text{PMI}} = \frac{1}{N}\sum_{i=1}^{N}\left[\ell_{\text{CE}}(p_i^{\text{stage}}, y_i^{\text{stage}}) + \lambda_{\text{eti}} \cdot \mathbb{I}(y_i^{\text{stage}} \in \{\text{MCI, DE}\})\frac{1}{K}\sum_{k=1}^{K}\ell_{\text{BCE}}(p_i^{\text{eti},(k)}, y_i^{\text{eti},(k)})\right], \tag{4}$$

where $\ell_{\text{CE}}(\cdot)$ and $\ell_{\text{BCE}}(\cdot)$ denote the multi-class cross-entropy loss and the binary cross-entropy loss, respectively. For subject i, p_i and y_i are the predicted probabilities and ground-truth labels. The indicator function $\mathbb{I}(\cdot)$ restricts etiology supervision to symptomatic subjects (MCI or DE), and the balancing hyperparameter λ_{eti} is set to 0.1.

2.4 Overall Loss Function

The total loss function is formulated as the weighted sum of the multi-modal alignment loss and the hierarchical inference loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{align}} + \lambda_{\text{PMI}}\mathcal{L}_{\text{PMI}}, \tag{5}$$

where λ_{PMI} balances the two objectives. To bypass the non-differentiable routing gate, $\mathcal{L}_{\text{total}}$ is minimized via a hierarchical training strategy (detailed in Sec. 3.2).

3 Experiments and Results

3.1 Datasets

To evaluate the proposed MoMmNet, we curated a large-scale, multi-center dataset comprising 15,016 subjects (6,834 males and 8,182 females) from four

public databases (UK Biobank, NACC, ADNI, and OASIS) and a local clinical cohort (Zhongshan Hospital) (Fig. 2a). These subjects are annotated with 21,231 hierarchical labels that capture coexisting pathologies (Fig. 2b). These labels encompass three baseline cognitive stages: NC ($N = 6,818$), MCI ($N = 3,893$), and DE ($N = 4,305$), alongside 10 distinct underlying etiologies (e.g., AD and VD) as detailed in [21]. Overall, the dataset aggregates 62,566 multi-modal samples (Fig. 2c). The imaging modalities include structural brain MRI (T1 and FLAIR), brain PET (FDG, AV45, and AV1451), whole-body Dixon MRI, and whole-body FDG-PET. Additionally, the clinical text data include fluid biomarkers, demographics, clinical scales, and electronic health records. The cohort is predominantly elderly, with 65.0% of participants aged 60 to 80 years (Fig. 2d).

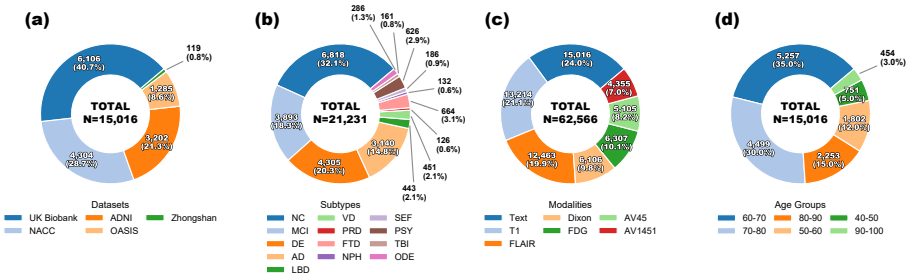

Fig. 2. Overview of our datasets. (a) Number of subjects in each dataset, (b) diagnostic label distribution, (c) number of samples in each modality, and (d) age distribution.

3.2 Implementation

We implemented MoMmNet on an NVIDIA A100 GPU. For imaging data, an organ-specific ResNet18 encoder processes all modalities associated with each organ. Meanwhile, clinical text is encoded using a multi-layer perceptron. The network was trained for 200 epochs using the AdamW optimizer with a mini-batch size of 4. The learning rate followed a warmup-cosine schedule, gradually decaying from 1×10^{-4} to 1×10^{-6}. Hyperparameters were set to $\lambda_{o,m} = \lambda_{\text{text}} = 0.1$ and $\lambda_{\text{PMI}} = 2$. During training, missing modalities were represented by zero vectors, and a class-balanced sampling strategy was adopted to mitigate dataset imbalances. Furthermore, to bypass the non-differentiable routing gate, we implemented a hierarchical training strategy for PMI: the Staging Head was first trained to convergence, followed by optimizing the Etiology Head exclusively on ground-truth symptomatic subjects (MCI and DE). To robustly assess the model, all evaluations were performed using 5-fold cross-validation, with the data split into training, validation, and testing sets at a 7:1:2 ratio. The model was evaluated using Accuracy (ACC), Precision (PRE), Sensitivity (SEN), Specificity (SPE), F1-score (F1), and Area Under the ROC Curve (AUC). For baseline comparison, we selected representative methods covering convolutional backbones,

transformer-based architectures, recent multimodal dementia diagnosis models, and a brain-only variant of our framework, including ResNet18 [9], ViT [6], ADRD [21], and our MoMmNet-B, respectively.

3.3 Results

Cognitive Staging. Table 1 presents the cognitive staging results. ResNet18, ViT, ADRD, and our MoMmNet-B use solely brain data, while the MoMmNet utilizes multi-organ data. Among the brain-only models, MoMmNet-B performs the best. By incorporating multi-organ data, MoMmNet achieves the highest overall performance (AUC: 0.828). It also consistently leads in accuracy, precision, specificity, and F1-score. Although ADRD yields slightly higher sensitivity, MoMmNet provides superior balanced classification across NC, MCI, and DE cohorts, establishing a robust foundation for downstream etiology diagnosis.

Table 1. Comparison with representative baselines for cognitive staging.

Method	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)	AUC(↑)
ResNet18 [9]	0.671 ± 0.036	0.720 ± 0.059	0.604 ± 0.077	0.753 ± 0.065	0.661 ± 0.098	0.750 ± 0.063
ViT [6]	0.648 ± 0.084	0.699 ± 0.082	0.565 ± 0.091	0.722 ± 0.053	0.625 ± 0.057	0.742 ± 0.086
ADRD [21]	0.692 ± 0.067	0.738 ± 0.043	$\mathbf{0.662 \pm 0.096}$	0.785 ± 0.099	0.694 ± 0.070	0.787 ± 0.051
MoMmNet-B	0.715 ± 0.080	0.766 ± 0.047	0.643 ± 0.061	0.792 ± 0.088	0.703 ± 0.087	0.796 ± 0.055
MoMmNet	$\mathbf{0.739 \pm 0.035}$	$\mathbf{0.798 \pm 0.073}$	0.659 ± 0.092	$\mathbf{0.818 \pm 0.056}$	$\mathbf{0.724 \pm 0.050}$	$\mathbf{0.828 \pm 0.079}$

Etiology Prediction. Table 2 presents the performance of the Etiology Head. Etiology Head accurately identifies single dementia subtype despite cascaded errors from cognitive staging. It excels in AD (ACC: 0.642, AUC: 0.764) and shows competitive outcomes for PSY and ODE. Furthermore, we extended our evaluation to subjects with co-occurring etiologies (e.g., AD+LBD and AD+FTD). These mixed pathologies exhibit lower performance compared to single etiologies (AD+LBD AUC: 0.609, AD+FTD AUC: 0.592). This decrease highlights the inherent clinical challenge of separating overlapping pathological signals.

Ablation Study. We conduct an ablation study to validate the individual contributions of the FA and PFD modules (Table 3). Adding the FA module improves ACC from 0.683 to 0.695, whereas incorporating the PFD module alone provides a more substantial gain (ACC: 0.712). The full model integrating both modules achieves the best performance (ACC: 0.739, AUC: 0.828), demonstrating their synergistic effect in enhancing the discriminative ability of the network.

Interpretability Analysis. To quantify the contributions of different modalities and organs, we employed SHAP analysis. The mean absolute SHAP values in Fig. 3a show that brain imaging features dominate model decisions, while

Table 2. Classification performance of etiology prediction.

Group	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)	AUC(↑)
			Single Etiology			
AD	0.642 ± 0.033	0.608 ± 0.041	0.663 ± 0.046	0.622 ± 0.044	0.617 ± 0.052	0.764 ± 0.045
LBD	0.575 ± 0.073	0.344 ± 0.078	0.596 ± 0.074	0.563 ± 0.070	0.429 ± 0.030	0.653 ± 0.056
VD	0.520 ± 0.036	0.288 ± 0.047	0.499 ± 0.081	0.542 ± 0.086	0.355 ± 0.053	0.616 ± 0.079
PRD	0.469 ± 0.077	0.253 ± 0.089	0.500 ± 0.057	0.438 ± 0.032	0.308 ± 0.075	0.568 ± 0.092
FTD	0.566 ± 0.063	0.376 ± 0.082	0.578 ± 0.037	0.556 ± 0.065	0.446 ± 0.031	0.643 ± 0.087
NPH	0.503 ± 0.088	0.212 ± 0.040	0.529 ± 0.084	0.471 ± 0.068	0.251 ± 0.064	0.572 ± 0.043
SEF	0.521 ± 0.034	0.243 ± 0.072	0.483 ± 0.039	0.545 ± 0.080	0.323 ± 0.076	0.615 ± 0.049
PSY	0.607 ± 0.085	0.302 ± 0.054	0.617 ± 0.083	0.567 ± 0.071	0.365 ± 0.058	0.680 ± 0.095
TBI	0.481 ± 0.059	0.267 ± 0.061	0.462 ± 0.048	0.523 ± 0.050	0.322 ± 0.055	0.562 ± 0.094
ODE	0.602 ± 0.060	0.311 ± 0.062	0.583 ± 0.066	0.613 ± 0.035	0.397 ± 0.093	0.670 ± 0.105
			Co-occurring Etiologies (Mixed Dementia)			
AD + LBD	0.528 ± 0.102	0.287 ± 0.104	0.558 ± 0.091	0.502 ± 0.097	0.374 ± 0.107	0.609 ± 0.096
AD + FTD	0.522 ± 0.100	0.338 ± 0.069	0.527 ± 0.098	0.493 ± 0.090	0.386 ± 0.099	0.592 ± 0.093

Table 3. Ablation study of the proposed FA and PFD modules.

FA	PFD	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)	AUC(↑)
×	×	0.683 ± 0.036	0.715 ± 0.059	0.614 ± 0.079	0.761 ± 0.065	0.670 ± 0.084	0.767 ± 0.063
✓	×	0.695 ± 0.043	0.741 ± 0.037	0.625 ± 0.064	0.770 ± 0.027	0.675 ± 0.051	0.773 ± 0.076
×	✓	0.712 ± 0.070	0.772 ± 0.048	0.630 ± 0.067	0.789 ± 0.061	0.701 ± 0.047	0.815 ± 0.072
✓	✓	$\mathbf{0.739 \pm 0.035}$	$\mathbf{0.798 \pm 0.073}$	$\mathbf{0.659 \pm 0.092}$	$\mathbf{0.818 \pm 0.056}$	$\mathbf{0.724 \pm 0.050}$	$\mathbf{0.828 \pm 0.078}$

peripheral-organ features, particularly heart FDG and liver Dixon, provide complementary diagnostic information. The SHAP distribution in Fig. 3b further shows that lower brain FDG values and higher brain AV45 and AV1451 values shift predictions toward disease-related stages. Peripheral-organ features exhibit weaker and more heterogeneous directional patterns, suggesting their role as auxiliary systemic cues rather than dominant decision factors. These findings demonstrate that MoMmNet integrates central neuroimaging and peripheral biomarkers into a unified diagnostic representation.

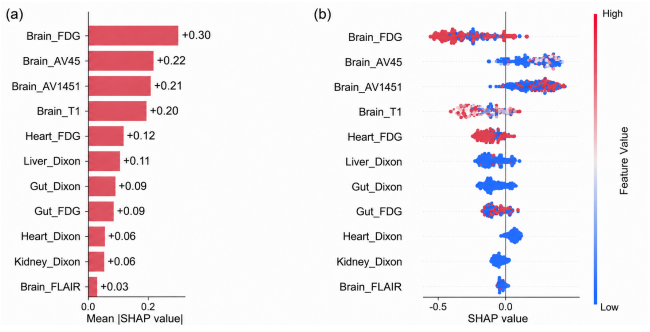

Fig. 3. Interpretability analysis. (a) mean absolute SHAP values. (b) distribution of SHAP values.

4 Conclusion

We present MoMmNet, a hierarchical multi-task framework that advances dementia diagnosis by modeling it as a systemic multi-organ disorder. By integrating multi-organ imaging and clinical text, our method aligns heterogeneous modalities into a unified brain-anchored space. To mitigate biological confounders, we introduce a memory bank-based decoupling strategy that robustly isolates pathological signals from physiological variations. Unlike conventional methods limited to single-disease prediction, MoMmNet performs multi-etiology inference to identify co-occurring pathologies, directly addressing the clinical reality of mixed dementia. Validated on a large-scale multi-center dataset of 15,016 subjects, our approach demonstrates superior diagnostic performance. Despite these promising results, several limitations remain. First, although the selected baselines cover representative diagnostic paradigms, they are not exhaustive. Second, although an adaptive age-bin expansion strategy is adopted to improve statistical stability, exact demographic matching within the PFD module may be relaxed in underrepresented age subgroups, leading to approximate rather than strictly matched normative references. Future work will incorporate larger and more diverse normative databases to further improve the reliability of profile-conditioned decoupling.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers U23A20295, 62131015, 82441023, 82394432), the China Ministry of Science and Technology (STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), the Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bettcher, B.M., Tansey, M.G., Dorothée, G., Heneka, M.T.: Peripheral and central immune system crosstalk in alzheimer disease—a research prospectus. Nature Reviews Neurology **17**(11), 689–701 (2021)
2. Bu, X., Xiang, Y., Jin, W., Wang, J., Shen, L., Huang, Z., Zhang, K., Liu, Y., Zeng, F., Liu, J., et al.: Blood-derived amyloid-β protein induces alzheimer's disease pathologies. Molecular psychiatry **23**(9), 1948–1956 (2018)
3. Collaborators, G.D.F.: Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. The Lancet. Public health **7**(2), e105–e125 (2022)
4. Cryan, J.F., O'Riordan, K.J., Cowan, C.S., Sandhu, K.V., Bastiaanssen, T.F., Boehme, M., Codagnone, M.G., Cussotto, S., Fulling, C., Golubeva, A.V., et al.: The microbiota-gut-brain axis. Physiological reviews (2019)
5. De La Torre, J.: The vascular hypothesis of alzheimer's disease: a key to preclinical prediction of dementia using neuroimaging. Journal of Alzheimer's Disease **63**(1), 35–52 (2018)

6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
7. Han, X., Fei, X., Wang, J., Zhou, T., Ying, S., Shi, J., Shen, D.: Doubly supervised transfer classifier for computer-aided diagnosis with imbalanced modalities. IEEE Transactions on Medical Imaging **41**(8), 2009–2020 (2022)
8. Hayden, K.M., Zandi, P.P., Lyketsos, C.G., Khachaturian, A.S., Bastian, L.A., Charoonruk, G., Tschanz, J.T., Norton, M.C., Pieper, C.F., Munger, R.G., et al.: Vascular risk factors for incident alzheimer disease and vascular dementia: the cache county study. Alzheimer Disease & Associated Disorders **20**(2), 93–100 (2006)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Jack, C.R., Knopman, D.S., Jagust, W.J., Petersen, R.C., Weiner, M.W., Aisen, P.S., Shaw, L.M., Vemuri, P., Wiste, H.J., Weigand, S.D., et al.: Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The lancet neurology **12**(2), 207–216 (2013)
11. Kane, J.P., Surendranathan, A., Bentley, A., Barker, S.A., Taylor, J.P., Thomas, A.J., Allan, L.M., McNally, R.J., James, P.W., McKeith, I.G., et al.: Clinical prevalence of lewy body dementia. Alzheimer's research & therapy **10**(1), 19 (2018)
12. Li, F., Bai, S., Liu, Y., Chen, Z., Zhao, S., Ding, Z., Xie, F., Xu, Y., Yue, L., Zhang, H., et al.: Enhancing diagnosis of mild cognitive impairment through brain-heart-gut metabolic networks in whole-body pet imaging. Cell Reports Medicine **7**, 102629 (2026)
13. Li, F., Zhao, S., Bai, S., Liu, Y., Zhang, K., Xu, Y., Zhang, Y., Sun, K., Shen, D.: Brain-heart-gut guided multi-constraint knowledge distillation for early alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 45–54. Springer (2025)
14. Li, F., Zhao, S., Liang, Z., Xu, Z., Bai, S., Zhou, W., Zheng, Q., Yan, W., Wang, L., Sun, K., Zhang, H., Shen, D.: Brain-body interactions in systemic diseases: a survey from an imaging perspective. MedScience (2026)
15. Onyike, C.U., Diehl-Schmid, J.: The epidemiology of frontotemporal dementia. International review of psychiatry **25**(2), 130–137 (2013)
16. Poos, J.M., Grandpierre, L.D., van der Ende, E.L., Panman, J.L., Papma, J.M., Seelaar, H., van den Berg, E., van't Klooster, R., Bron, E., Steketee, R., et al.: Longitudinal brain atrophy rates in presymptomatic carriers of genetic frontotemporal dementia. Neurology **99**(24), e2661–e2671 (2022)
17. Qiu, S., Miller, M.I., Joshi, P.S., Lee, J.C., Xue, C., Ni, Y., Wang, Y., De Anda-Duran, I., Hwang, P.H., Cramer, J.A., et al.: Multimodal deep learning for alzheimer's disease dementia assessment. Nature communications **13**(1), 3404 (2022)
18. Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C.E., Cummings, J., van der Flier, W.M.: Alzheimer's disease. The Lancet **397**(10284), 1577–1590 (2021)
19. Sun, K., Zhang, Y., Liu, J., Yu, L., Zhou, Y., Xie, F., Guo, Q., Zhang, H., Wang, Q., Shen, D.: Achieving multi-modal brain disease diagnosis performance using only single-modal images through generative ai. Communications Engineering **3**(1), 96 (2024)

20. Tanveer, M., Richhariya, B., Khan, R.U., Rashid, A.H., Khanna, P., Prasad, M., Lin, C.T.: Machine learning techniques for the diagnosis of alzheimer's disease: A review. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **16**(1s), 1–35 (2020)
21. Xue, C., Kowshik, S.S., Lteif, D., Puducheri, S., Jasodanand, V.H., Zhou, O.T., Walia, A.S., Guney, O.B., Zhang, J.D., Poésy, S., et al.: Ai-based differential diagnosis of dementia etiologies on multimodal data. Nature Medicine **30**(10), 2977–2989 (2024)
22. Zhou, T., Liu, M., Fu, H., Wang, J., Shen, J., Shao, L., Shen, D.: Deep multi-modal latent representation learning for automated dementia diagnosis. In: International conference on medical image computing and computer-assisted intervention. pp. 629–638. Springer (2019)