Cross-Modal Alignment and Distillation for Retinal Imaging Dementia Screening

Zhuoting Xu[1,2†], Le Cao[3,2†], Tiande Zhang[2,4†], Lei Mou[2], Shijia Zhou[2], Fan Li[1], Dinggang Shen[1,5,6], Huazhu Fu[7], Bo Wu[3✉], and Yitian Zhao[1,2✉]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2] Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, China
[3] Department of Neurology, West China Hospital, Sichuan University, Chengdu 610041, China
[4] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150000, China
[5] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 201807, China
[6] Shanghai Clinical Research and Trial Center, Shanghai 201210, China
[7] Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
yitian.zhao@nimte.ac.cn, dr.bowu@hotmail.com

Abstract. Dementia screening remains a clinical challenge due to the high cost, invasiveness and sometime subjective of traditional neuroimaging or cognitive assessment questionnaires. While retinal imaging provides a non-invasive alternative, current methods struggle to capture the cerebral pathology needed for accurate diagnosis. To address this, we propose EyeDEM, a novel framework designed to distill brain structural priors into a standalone eye screening network using paired eye-brain data. To overcome cross-organ heterogeneity, we disentangle the latent space to isolate cross-modal relate representations and then apply multi-level alignment to map retinal microvasculature to brain atrophy. Furthermore, we introduce a divide-and-conquer distillation strategy that selectively transfers hierarchical diagnostic knowledge from a fused multimodal teacher to the eye student. At the inference stage, EyeDEM functions as a standalone screening method requiring solely retinal images. Validated across multi-center datasets of 2,040 participants, our standalone eye model reaches the AUC of 0.888 for dementia *vs.* normal control classification. Crucially, we validate the model's interpretability through biological correlation analysis, demonstrating that the learned retinal representations strongly correlate with critical neurodegenerative brain regions by predicting their volumes. These findings confirm that leveraging the eye-brain axis as a cross-modal prior paves the way for accessible, high-precision, and clinically transparent dementia screening.

Keywords: Retinal OCTA · Distillation · Dementia · Eye-Brain Axis

† Zhuoting Xu, Le Cao, and Tiande Zhang contributed equally to this work.

1 Introduction

Dementia, most notably Alzheimer's disease (AD) and vascular dementia (VaD), represents a rapidly escalating global public health crisis [15].Current diagnostic workups, including magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF) analysis, are unsuitable for large-scale population screening because of their high cost and limited accessibility, as well as the invasive nature of CSF sampling. Furthermore, traditional cognitive assessment questionnaires are frequently compromised by inherent subjectivity, limiting their reliability as objective early-screening tools [16].

As a developmental outgrowth of the diencephalon, the retina shares a common embryonic origin and homologous vascular architecture with the brain, offering a unique window into cerebral pathology [13]. This anatomical connection suggests that retinal microvasculature can serve as a surrogate marker for brain health. Accumulating clinical evidence substantiates this link, demonstrating that dementia is consistently associated with specific retinal alterations, including reduced vessel density in the superficial and deep capillary plexuses (SVC/DVC) and perfusion deficits in the choriocapillaris (CC) [23,21,19,10]. Crucially, these microvascular impairments also correlate with brain structural atrophy [6] and parallel cognitive decline [18]. Optical coherence tomography angiography (OCTA), which provides non-invasive, high-resolution quantification of vascular layers, has thus emerged as a promising modality for detecting these neurodegenerative footprints.

Despite this potential, existing retinal image-based brain disease screening methods remain limited by their exclusive reliance on retinal images. Current approaches typically map retinal morphology directly to disease labels, treating the eye as an isolated organ [22,20,3]. However, retinal microvascular abnormalities are merely peripheral phenotypes; they do not explicitly capture the core macro-structural neurodegeneration and volumetric loss occurring within the brain. This semantic disconnect results in models that lack interpretability and struggle to differentiate dementia-specific vascular changes from systemic vascular confounders. Without grounding retinal features in their corresponding cerebral context, the diagnostic sensitivity of eye-only models remains capped.

To bridge this gap, we introduce EyeDEM, a novel framework that leverages a unique cohort of paired retinal (OCTA) and brain (MRI) imaging scans to model the biological coupling between the eye and the brain, as illustrated in Fig. 1. By explicitly aligning eye representations with brain structural priors during training, our framework enables the EyePipe to internalize cross-organ pathological correlation. Consequently, during the inference stage, the model can utilize the learned associations and the transferred brain structural priors to guide a more accurate standalone eye diagnosis. The main contributions of this work are summarized as follows:

1) We propose a pioneering framework that establishes latent associations between retinal phenotypes and cerebral pathology. This enables a standalone retinal model to leverage high-level neurodegenerative priors, bridging the gap between peripheral and central biomarkers.

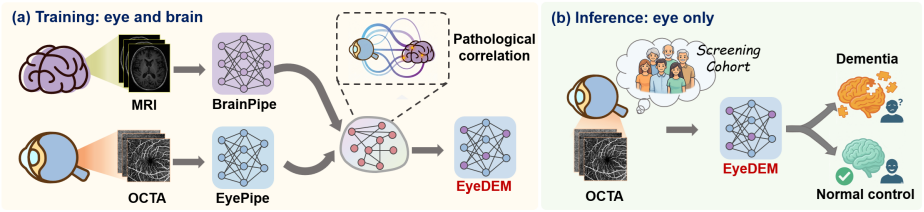

Fig. 1. Overview of our study. (a) Training: Brain MRI and retinal OCTA train modality specific networks to capture eye-brain pathological correlations. (b) Inference: The trained EyeDEM performs dementia screening using only retinal OCTA inputs.

2) We introduce an eye-brain disentanglement module to decouple shared neurodegenerative semantics from modality specific characteristics and a multi-level alignment module to align them. This ensures that the learned eye-brain alignment is grounded in genuine pathological markers rather than imaging artifacts or non-diagnostic features.

3) Our model achieves high precision dementia screening using solely retinal images. We further validate its biological interpretability by predicting corresponding dementia-related structural brain atrophy directly from eye input.

2 Method

We propose EyeDEM, a novel cross-modal knowledge alignment and distillation framework for dementia screening that explicitly transfers structural brain priors into a standalone eye network (Fig .2.). Specifically, EyeDEM comprises three core components: a coupled feature eye-brain disentanglement and multi-level alignment (MLA) mechanism, followed by a cross-modal distillation module, named divided-and-conquer distillation (DCKD). Initially, an MRI-based BrainPipe and an OCTA-based EyePipe are independently pretrained on unpaired datasets. During the training phase, BrainPipe serves as a fixed teacher, while Eyepipe is optimized through alignment and distillation to capture biologically relevant eye-brain associations related to dementia.

2.1 Eye-Brain Disentanglement and Multi-Level Alignment

Directly aligning raw eye and brain features is inherently ill-posed due to pronounced cross-organ heterogeneity; therefore, we introduce a disentanglement strategy and the MLA module to align eye representations with brain structural priors in a shared latent space, enabling EyePipe to extract highly eye-brain related features.

Eye-Brain Latent Disentanglement: We first decompose the latent space to yield a purified subspace optimized for cross-modal alignment. Specifically, we

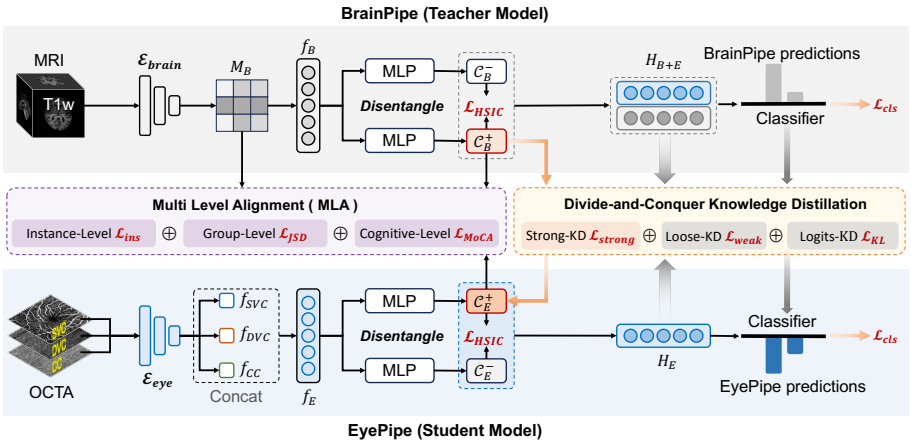

Fig. 2. Details of the proposed EyeDEM framework.

employ encoders $\mathcal{E}_{eye}$ and $\mathcal{E}_{brain}$ to extract global features $\mathbf{f}_E, \mathbf{f}_B \in \mathbb{R}^d$ alongside a spatial volume $\mathbf{M}_B$. To isolate shared pathological signatures from modality specific variations, each global feature $\mathbf{f}_m$ ($m \in \{E, B\}$) is projected into two orthogonal subspaces via MLP heads. This yields a cross-modality related representation, $\mathcal{C}_m^+$, which captures shared pathology, and an inner-modality specific representation, $\mathcal{C}_m^-$, which retains unique, organ-independent characteristics. To enforce strict statistical independence between these representations, we minimize the Hilbert-Schmidt Independence Criterion (HSIC):

$$\mathcal{L}_{HSIC} = \sum_{m\in\{E,B\}} \text{HSIC}(\mathcal{C}_m^+, \mathcal{C}_m^-) = \sum_{m\in\{E,B\}} \frac{1}{(B-1)^2} \text{Tr}(\mathbf{K}_{\mathcal{C}_m^+}\mathbf{H}\mathbf{K}_{\mathcal{C}_m^-}\mathbf{H}) \quad (1)$$

where $\mathbf{H} = \mathbf{I} - \frac{1}{B}\mathbf{1}\mathbf{1}^T$ serves as the centering matrix, B denotes the batch size, and $\mathbf{K}$ represents the respective Gram matrices.

Eye-Brain Multi-Level Alignment: Once the purified related representations ($\mathcal{C}_m^+$) are isolated, we first align them at the instance level. To maintain semantic consistency, we employ a cross-attention mechanism where the shared eye representation $\mathcal{C}_E^+$ acts as the query, and the brain spatial feature map $\mathbf{M}_B$ serves as both keys and values. This mechanism enables $\mathcal{C}_E^+$ to retrieve semantically relevant cerebral patterns directly from $\mathbf{M}_B$. To ensure that $\mathcal{C}_E^+$ learns to index brain-related signatures without inappropriately updating the teacher representations, the instance level loss $\mathcal{L}_{ins}$ is minimized using a stop-gradient operation $\text{sg}(\cdot)$ applied to the BrainPipe:

$$\hat{\mathcal{C}}_B = \text{Attn}(\mathcal{C}_E^+, \text{sg}(\mathbf{M}_B), \text{sg}(\mathbf{M}_B)), \mathcal{L}_{ins} = \|\hat{\mathcal{C}}_B - \text{sg}(\mathcal{C}_B^+)\|_2^2 \quad (2)$$

where $\text{Attn}(\cdot)$ denotes the scaled dot-product attention.

While instance-level alignment enforces sample-wise semantic consistency, it does not guarantee broader distributional consistency across the dataset. To address this, we elevate the alignment to the group level by maximizing the mutual information between $\mathcal{C}_E^+$ and $\mathcal{C}_B^+$. Following the Deep InfoMax strategy [5], we maximize a lower bound of the MI by minimizing the Jensen-Shannon Divergence (JSD) between the joint and the product of marginal distributions, which encourage the shared subspaces to exhibit strong statistical dependency across sample batches. A discriminator $\mathcal{D}$ distinguishes subject-matched pairs $(\mathcal{C}_E^+, \mathcal{C}_B^+)$ from cross-subject mismatched pairs $(\mathcal{C}_E^+, \tilde{\mathcal{C}}_B^+)$, where $\tilde{\mathcal{C}}_B^+$ is a within-batch permutation of brain features. This encourages group-level statistical dependence between shared eye and brain representations:

$$\mathcal{L}_{JSD} = \mathbb{E}_{\text{joint}}[-\log \mathcal{D}(\mathcal{C}_E^+, \mathcal{C}_B^+)] + \mathbb{E}_{\text{marginal}}[-\log(1 - \mathcal{D}(\mathcal{C}_E^+, \tilde{\mathcal{C}}_B^+))] \tag{3}$$

The instance- and group-level alignments synchronize the cross-modal feature spaces; however, they do not inherently ensure that these synchronized spaces encode clinically actionable neurodegenerative markers. This is because the alignment objectives mainly constrain cross-modal consistency, rather than explicitly supervising disease severity or clinical relevance. To clinically ground the aligned representations and promote cognitive-level alignment consistent with the previously observed continuity, we incorporate the Montreal Cognitive Assessment (MoCA) score, a standardized measure of global cognitive function, as a continuous proxy of disease severity. Two independent regression heads, $\mathcal{R}_E$ and $\mathcal{R}_B$, are attached to the shared representations $\mathcal{C}_E^+$ and $\mathcal{C}_B^+$, respectively. To encourage the learned features to continuously reflect disease progression, we optimize a Pearson Correlation Loss:

$$\mathcal{L}_{MoCA} = \sum_{m \in \{E,B\}} [1 - \rho(\mathcal{R}_m(\mathcal{C}_m^+), \mathbf{y})] \tag{4}$$

where $\rho(\cdot)$ denotes the Pearson correlation coefficient and y represents the ground-truth MoCA score. Unlike absolute error minimization, maximizing correlation emphasizes monotonic consistency between predicted and true cognitive severity. This captures the continuous trajectory of neurodegeneration rather than merely fitting exact numerical values, ensuring that the final representations encode highly meaningful, disease-related variations.

The training objective for MLA module is formulated as a weighted sum, with the empirical parameters tuned as $\lambda_{ins} = 0.5$, $\lambda_{jsd} = 0.6$, and $\lambda_{MoCA} = 0.5$:

$$\mathcal{L}_{MLA} = \lambda_{ins}\mathcal{L}_{ins} + \lambda_{jsd}\mathcal{L}_{JSD} + \lambda_{MoCA}\mathcal{L}_{MoCA}. \tag{5}$$

2.2 Divide-and-Conquer Knowledge Distillation

Inspired by Li et al. [9,8], who distilled brain-heart-gut data to optimize standalone brain inference, we further improve upon their work by proposing the DCKD module to resolve the negative transfer caused by indiscriminate feature alignment. We first construct a multimodal teacher using fused features $\mathbf{f}_{fus} = [\mathbf{f}_E, \mathbf{f}_B]$. Given the high cross-modal consistency of the disentangled $\mathcal{C}_E^+$

and $\mathcal{C}_B^+$, we apply a strong Mean Squared Error (MSE) constraint to rigorously lock the student's latent eye representation to the teacher's reliable brain prior:

$$\mathcal{L}_{strong} = \|\mathcal{C}_E^+ - \text{sg}(\mathcal{C}_B^+)\|_2^2.$$

Let H_E, H_{B+E} denote the pre-logit embeddings from the eye-only network f_E and the fused eye-brain network f_{fus}. To avoid distortion from strict feature alignment, we adopt a weak distillation strategy that preserves relational structure. Specifically, a contrastive loss $\mathcal{L}_{weak}$ is used to align sample-wise similarities, encouraging structural consistency between student and teacher representations. In addition, a KL divergence loss $\mathcal{L}_{KL}$ is applied on output logits to transfer softened class distributions and refine classification behavior. Both $\mathcal{L}_{weak}$ and $\mathcal{L}_{KL}$ follow Li et al. [9]. With empirical weights set to $\lambda_s = 0.5$, $\lambda_w = 0.3$, and $\lambda_{kl} = 0.6$, the DCKD loss:

$$\mathcal{L}_{DCKD} = \lambda_s \mathcal{L}_{strong} + \lambda_w \mathcal{L}_{weak} + \lambda_{kl} \mathcal{L}_{KL} \tag{6}$$

The overall training objective (where $\mathcal{L}_{cls}$ denotes the standard classification loss) are formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{HSIC} + \mathcal{L}_{MLA} + \mathcal{L}_{DCKD} + \mathcal{L}_{cls}. \tag{7}$$

3 Experiments

3.1 Datasets and Implementation Details

This study includes a total of 2,040 participants sourced from multiple AD and VaD cohorts. **Internal dataset:** we curated a paired eye-brain imaging dataset comprising 137 participants (74 dementia [DEM]; 63 normal controls [NC]) with definitive CSF-derived ground truth. This cohort yielded 246 retinal OCTA volumes explicitly paired with subject-level T1-weighted brain MRIs. The OCTA images were acquired over a 6×6 mm^2 fovea-centered area with a resolution of 512 × 512. Concurrently, the T1-weighted MRI scans were preprocessed included intensity correction, skull-stripping [11], and affine alignment to the MNI152 standard space [24]. The standardized brain volumes were cropped to 160×192×160 voxels at a 1×1×1 mm^3 isotropic resolution. We supported independent modality pre-training by supplementing the core cohort with two unpaired datasets. Specifically, we included 1,046 OCTA volumes from 547 additional subjects (287 DEM; 260 NC) and 1,278 T1-weighted MRIs obtained from the ADNI database (410 DEM; 868 NC) [7,14,17]. **External dataset:** to evaluate the generalization capability and robustness of our framework, we utilized an external dataset consisting exclusively of retinal OCTA imaging from different medical center. This cohort includes a total of 78 participants (35 DEM; 43 NC).

Our framework was implemented in PyTorch and trained on four NVIDIA RTX 4090 GPUs with a batch size of 32. We optimized the network using AdamW with a warm-up mechanism and an initial learning rate of 2×10^{-4}. Performance was rigorously evaluated using five-fold cross-validation on the internal

Table 1. Quantitative dementia detection performance using retinal OCTA only.

Method	Internal			External		
	AUC	**SEN**	**F1**	**AUC**	**SEN**	**F1**
ResNet50	0.808±0.063	0.815±0.094	0.786±0.083	0.712±0.081	0.723±0.073	0.648±0.088
ConvNeXt-t	0.731±0.061	0.779±0.059	0.724±0.047	0.676±0.070	0.734±0.064	**0.706±0.083**
ViT-B	0.770±0.068	0.704±0.078	0.731±0.163	0.714±0.074	0.637±0.082	0.614±0.048
Eye-AD	0.816±0.052	0.807±0.071	0.793±0.068	0.723±0.061	0.720±0.095	0.685±0.081
MuCo	0.813±0.049	0.830±0.077	0.787±0.089	0.718±0.055	0.726±0.074	0.683±0.067
EyeDEM	**0.888±0.076**	**0.856±0.071**	**0.833±0.156**	**0.761±0.091**	**0.754±0.067**	0.705±0.074

dataset. While paired eye-brain scans guided the knowledge distillation process during training, validation and external testing relied strictly on standalone retinal OCTA inputs.

3.2 Dementia Detection Performance

To validate EyeDEM, we compared it with five representative methods on both internal and external datasets under identical training and testing protocols, including three generic methods (ResNet [4], ConvNeXt [12], and ViT [1]) and two OCTA-specific dementia methods (Eye-AD [3] and MuCo [20]). As shown in Table 1, EyeDEM achieves the best performance, particularly in AUC and SEN. On the internal dataset, it reaches an AUC of 0.888, SEN of 0.856, and F1 of 0.833, exceeding the strongest baseline by an absolute AUC gain of 0.080. When transitioning to the external dataset, all models predictably experience a performance drop due to domain shift. Despite this, EyeDEM generalize well, which attains the highest AUC (0.761) and SEN (0.754). Its F1 (0.705) remains highly competitive, though marginally outperformed by ConvNeXt-t (0.706). In contrast, EyeDEM's robust gains suggest that incorporating cerebral structural priors not only significantly improves diagnostic performance by capturing clinically relevant neurodegenerative patterns from OCTA inputs, but also ensures strong generalization capabilities across diverse datasets.

3.3 Ablation Study

We evaluated key framework components via ablation studies (Fig. 3(a)). Removing the disentanglement module caused the sharpest performance drop (AUC 0.043, SEN 0.041), highlighting its critical role in mitigating cross-organ heterogeneity to stabilize alignment. Within the MLA module, omitting the clinical constraint ($\mathcal{L}_{MoCA}$) yielded the largest decline (AUC 0.034), demonstrating that clinically guided alignment outperforms purely modality-based matching. Finally, ablating distillation reduced the AUC by 0.031, confirming its necessity for transferring discriminative brain priors to the retinal branch. Overall, these three components synergistically drive the framework's efficacy.

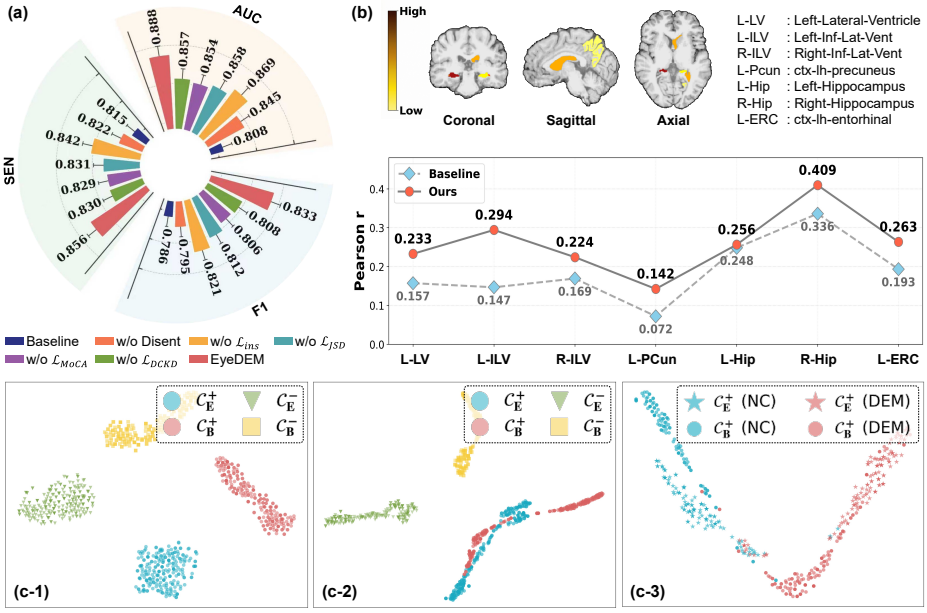

Fig. 3. Ablation study and model interpretability. (a) Performance impact of different model components. (b) ROI localization and Pearson correlation (r) comparison between ResNet50 and EyeDEM across anatomical regions. (c) Embedding evolution: Visualization of clusters $\{C_E^+, C_E^-, C_B^+, C_B^-\}$ at initialization (c-1) and post-training (c-2). c-3 displays $\{C_E^+, C_B^+\}$ distributions color-coded by clinical class (NC vs. DEM).

3.4 Interpretability and Biological Correlation Analysis

To demonstrate that our framework goes beyond superficial pattern matching and captures biologically plausible neurodegenerative markers, we investigated the correlation between the learned eye features and established brain atrophy patterns. Specifically, without further training, we assessed the biological validity of the extracted features by predicting structural volumes across seven dementia-associated regions of interest (ROIs) extracted by FastSurfer [2] (e.g., hippocampus, entorhinal cortex, and lateral ventricles). As illustrated in Fig. 3 (b), our proposed method achieves consistently higher Pearson correlations (r) than the baseline across all seven regions. Notably, the strongest correlation is observed in the right hippocampus ($r = 0.409$), a critical region fundamentally implicated in early dementia pathogenesis. These results demonstrate that EyeDEM captures eye–brain associations related to dementia, supporting their biological relevance and the framework's clinical interpretability.

To investigate the latent relationships between eye and brain modalities, we visualized the embedding space in Fig. 3(c). At initialization (c-1), the eye and brain features are largely dispersed and exhibit no discernible structure or alignment. After training (c-2), the shared components C_E^+ and C_B^+ become well

aligned across modalities, while the modality-specific components C_E^- and C_B^- remain clearly separated, indicating that modality-unique information is preserved without interference. As shown in (c-3), the aligned shared representations further exhibit improved separability between NC and DEM groups, supporting the effectiveness of cross-modal alignment and its potential relevance for clinical discrimination.

4 Conclusions

In this work, we propose EyeDEM, which transfers structural brain priors to an OCTA-only dementia screening network through cross-modal disentanglement, MLA, and DCKD. The framework achieves strong standalone screening performance while providing supportive biological evidence for eye-brain coupling. It offers a non-invasive approach to scalable dementia screening and a general paradigm for cross-organ medical imaging.

Acknowledgments. This work is supported by the National Science Foundation Program of China (62422122, 62272444, 62401551), the Key Research and Development Program of Zhejiang Province (2024C03101, 2024C03281), the Ningbo 'KeChuang Yongjiang 2035' Key R&D Program (2024Z230).

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

References

1. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
2. Estrada, S., Kügler, D., Bahrami, E., Xu, P., Mousa, D., Breteler, M.M., et al.: Fastsurfer-hypvinn: Automated sub-segmentation of the hypothalamus and adjacent structures on high-resolutional brain mri. Imaging Neuroscience **1**, imag–1 (2023)
3. Hao, J., Kwapong, W.R., Shen, T., Fu, H., Xu, Y., Lu, Q., et al.: Early detection of dementia through retinal imaging and trustworthy ai. npj Digital Medicine **7**(1), 294 (2024)
4. 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)
5. Hjelm, R.D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., et al.: Learning deep representations by mutual information estimation and maximization. arXiv preprint arXiv:1808.06670 (2018)
6. Hu, Z., Wang, L., Zhu, D., Qin, R., Sheng, X., Ke, Z., Shao, P., Zhao, H., Xu, Y., Bai, F.: Retinal alterations as potential biomarkers of structural brain changes in alzheimer's disease spectrum patients. Brain Sciences **13**(3), 460 (2023)
7. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)

8. Li, F., Bai, S., Liu, Y., Chen, Z., Zhao, S., Ding, Z., et al.: Enhancing diagnosis of mild cognitive impairment through brain-heart-gut metabolic networks in whole-body pet imaging. Cell Reports Medicine **7**(2), 102629 (2026)
9. Li, F., Zhao, S., Bai, S., Liu, Y., Zhang, K., Xu, Y., et al.: Brain-heart-gut guided multi-constraint knowledge distillation for early alzheimer's disease diagnosis. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Lecture Notes in Computer Science, vol. 15974, pp. 45–54. Springer Nature Switzerland, Cham (2025)
10. Li, F., Zhao, S., Liang, Z., Xu, Z., Bai, S., Zhou, W., Zheng, Q., Yan, W., Wang, L., Sun, K., et al.: Brain–body interactions in systemic diseases: a survey from an imaging perspective. MedScience pp. 1–33 (2026)
11. Liu, J., Liu, F., Sun, K., Jiang, C., Wang, Y., Sun, T., et al.: Brainparc: Unified lifespan brain parcellation with anatomy-guided progressive transmission. In: Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM) (2025)
12. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022)
13. London, A., Benhar, I., Schwartz, M.: The retina as a window to the brain—from eye research to cns disorders. Nature Reviews Neurology **9**(1), 44–53 (2013)
14. Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., et al.: Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). Alzheimer's & Dementia **1**(1), 55–66 (2005)
15. Nichols, E., Steinmetz, J.D., Vollset, S.E., Fukutaki, K., Chalek, J., Abd-Allah, F., et al.: 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)
16. Palmer, N.P., Ortega, B.T., Joshi, P.: Cognitive impairment in older adults: epidemiology, diagnosis, and treatment. Psychiatric Clinics **45**(4), 639–661 (2022)
17. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., et al.: Alzheimer's disease neuroimaging initiative (adni) clinical characterization. Neurology **74**(3), 201–209 (2010)
18. Wang, M., Chen, L., Gong, Y., Pazo, E.E., Gao, F., Hu, L., et al.: Cognitive function correlates with retinal structural and vascular imaging parameters in chinese older adults. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring **17**(3), e70147 (2025)
19. Wang, X., Wang, Y., Liu, H., Zhu, X., Hao, X., Zhu, Y., et al.: Macular microvascular density as a diagnostic biomarker for alzheimer's disease. Journal of Alzheimer's Disease **90**(1), 139–149 (2022)
20. Wang, X., Li, H., Xiao, Z., Fu, H., Zhao, Y., Jin, R., et al.: Screening of dementia on octa images via multi-projection consistency and complementarity. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 688–698. Springer (2022)
21. Wu, J., Zhang, X., Azhati, G., Li, T., Xu, G., Liu, F.: Retinal microvascular attenuation in mental cognitive impairment and alzheimer's disease by optical coherence tomography angiography. Acta ophthalmologica **98**(6), e781–e787 (2020)
22. Xie, J., Yi, Q., Wu, Y., Zheng, Y., Liu, Y., Macerollo, A., et al.: Deep segmentation of octa for evaluation and association of changes of retinal microvasculature with alzheimer's disease and mild cognitive impairment. British Journal of Ophthalmology **108**(3), 432–439 (2024)

23. Yoon, S.P., Grewal, D.S., Thompson, A.C., Polascik, B.W., Dunn, C., Burke, J.R., et al.: Retinal microvascular and neurodegenerative changes in alzheimer's disease and mild cognitive impairment compared with control participants. Ophthalmology Retina **3**(6), 489–499 (2019)
24. Zhang, Y., Sun, K., Liu, Y., Xie, F., Guo, Q., Shen, D.: A modality-flexible framework for alzheimer's disease diagnosis following clinical routine. IEEE Journal of Biomedical and Health Informatics **29**(1), 535–546 (2024)