REVEAL++: Differentiable Phenotypic Grouping for Vision–Language Retinal Modeling of Alzheimer's Disease Risk

Ethan Meidinger[1], Seowung Leem[2], Zeyun Zhao[2], and Ruogu Fang[2⋆]

[1] University of Virginia, Charlottesville, VA, USA
[2] J. Crayton Pruitt Family Department of Biomedical Engineering, Herbert Wertheim College of Engineering, University of Florida, Gainesville, FL, USA

Abstract. The retina offers a noninvasive window into neurodegenerative disease, capturing subtle structural patterns associated with a risk of future cognitive decline. Vision–language alignment frameworks such as REVEAL have shown that pairing retinal fundus images with structured clinical risk narratives improves early prediction of Alzheimer's disease (AD). A key design choice in these approaches is the use of phenotypic grouping, where individuals with similar risk profiles are treated as multi-positive pairs during contrastive learning. However, existing methods operationalize phenotypic similarity as a discrete construct, relying on hard group assignments that impose rigid supervision and decouple group formation from representation learning. We propose a continuous formulation of phenotypic structure within contrastive learning. Rather than assigning samples to fixed clusters, we model inter-subject similarity as a differentiable weighting function derived from intra-modality embedding similarities in both retinal images and risk profiles. These weights define soft multi-positive relationships through a continuous aggregation operator, enabling graded supervision that reflects the spectrum nature of disease risk. We further introduce a soft-target contrastive objective that jointly learns cross-modal alignment and phenotypic structure in an end-to-end manner. Evaluated on UK Biobank retinal imaging data for incident AD prediction, the proposed framework consistently outperforms discrete group-based contrastive learning and standard vision–language baselines. By treating phenotypic similarity as a learnable, continuous signal rather than a fixed grouping rule, our approach provides a principled and robust foundation for population-scale neurodegenerative risk modeling from multi-modal retinal and clinical data. Code is available at: https://github.com/lab-smile/REVEALpp.

Keywords: Neurodegenerative Disease · Multi-Modal Alignment · Preclinical Prediction

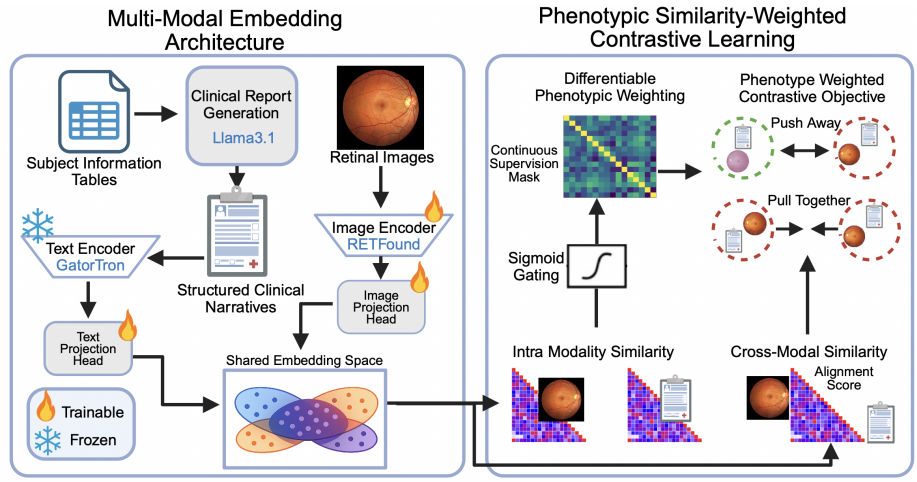

Fig. 1. Architecture of the proposed differentiable phenotypic weighting framework for group-aware contrastive learning. Image and text embeddings are aligned via a similarity-weighted multi-positive contrastive loss, where continuous phenotypic weights replace hard grouping to model the heterogeneous spectrum of Alzheimer's disease risk.

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by a long preclinical phase during which pathological changes accumulate before clinical symptoms [6]. Advances in brain imaging and plasma-based biomarkers have substantially improved the ability to detect disease-related pathology. However, these approaches may remain costly, invasive, or impractical for large-scale population screening. Complementary modalities that are scalable and noninvasive therefore play an important role in early risk stratification. The retina has emerged as one such modality, as its structure and microvasculature share developmental and physiological links with the central nervous system and are associated with neurodegenerative and vascular processes relevant to AD [3]. In parallel, systemic and lifestyle-related risk factors, including cardiometabolic health and sleep patterns, capture longitudinal exposures that contribute to dementia risk decades before diagnosis [4,2]. Rather than serving as standalone diagnostic markers, these signals offer complementary population-level information that may help characterize early disease susceptibility.

Recent advances in vision–language models (VLMs) have enabled joint representation learning across heterogeneous data modalities through contrastive alignment. Inspired by CLIP-style architectures, medical VLMs have increasingly been adapted to retinal imaging, leveraging large-scale pretraining to learn clinically meaningful visual representations [7,17,21]. Building on this paradigm, the

* Corresponding author: ruogu.fang@ufl.edu

REVEAL framework aligned retinal fundus images with individualized clinical risk narratives derived from structured health data, enabling multi-modal modeling of neurodegenerative risk [13]. A central innovation of REVEAL was group-aware contrastive learning (GACL), which encouraged subjects with similar phenotypic profiles to act as multi-positive pairs during training. This strategy improved robustness to individual-level noise and promoted learning of shared disease-relevant structure, leading to improved downstream AD risk prediction compared with uni-modal and standard pairwise contrastive approaches.

Despite these advantages, phenotypic grouping in existing GACL formulations is constructed through discrete similarity thresholds, implicitly assuming that individuals belong to well-separated risk categories. From a biological perspective, however, neurodegenerative risk evolves along continuous and overlapping trajectories shaped by heterogeneous genetic, vascular, metabolic, and lifestyle factors. Individuals often exhibit partial similarity across multiple phenotypic axes rather than membership in a single homogeneous group. Hard group assignments may therefore introduce artificial boundaries that fail to reflect the graded and spectrum-like nature of disease vulnerability, while preventing the grouping process itself from adapting during representation learning.

In this work, we extend the original REVEAL framework by introducing a differentiable phenotypic weighting formulation for retinal image–clinical narrative alignment. REVEAL++ preserves the same multi-modal alignment setting and encoder backbone choices, while replacing REVEAL's hard threshold-based group assignments with continuous, differentiable phenotypic weights. Rather than treating subject pairs as strictly positive or negative, REVEAL++ allows phenotypic similarity to modulate contrastive supervision in a graded manner. These weights are computed from retinal image embeddings and clinical risk-profile embeddings, then combined through a soft aggregation operator to define a soft multi-positive contrastive objective. By modeling phenotypic relationships as a differentiable attention-like process, the proposed framework enables joint learning of representation alignment and population-level structure, more faithfully capturing the continuous and heterogeneous biological variability underlying neurodegenerative risk. Our contributions are three-fold:

- **Differentiable phenotypic weighting:** We replace hard threshold-based grouping in group-aware contrastive learning with continuous phenotypic similarity weights derived from retinal and clinical embeddings, enabling smooth data-driven cohort modeling that better captures heterogeneous Alzheimer's disease risk.
- **Soft multi-positive contrastive learning:** We introduce a soft-target contrastive objective that incorporates phenotypic similarity into cross-modal alignment, enabling graded multi-positive supervision instead of binary pair assignments.
- **State-of-the-art Alzheimer's risk prediction from retinal imaging:** We achieve new state-of-the-art performance on UK Biobank retinal imaging for incident Alzheimer's disease prediction, outperforming existing vision–language and group-aware contrastive learning methods.

2 Methods

2.1 Overview of REVEAL++

REVEAL++ learns joint image–text representations of retinal fundus images and structured clinical reports under a group-aware contrastive objective. Given a minibatch of N subjects, each subject p is associated with a retinal image and a clinical report. Image and text encoders produce modality-specific embeddings, which are projected into a shared latent space. To incorporate phenotypic structure into contrastive supervision, we compute intra-modality similarity matrices that capture retinal image embedding and risk-profile similarity between subjects. These similarities are transformed into a differentiable phenotypic weighting mask $W \in [0,1]^{N\times N}$, which acts as a soft pairwise target matrix in a multi-positive contrastive loss.

2.2 Clinical Report Generation

To enable alignment between retinal images and systemic risk factors within a vision–language framework, structured questionnaire data were converted into synthetic clinical narratives compatible with pretrained text encoders. Using the LLaMA-3.1 API as the text generation engine, each participant's tabular risk-factor profile was mapped into a standardized clinical-style summary [9]. For each subject, the LLM received a predefined documentation template, the subject's structured demographic, behavioral, cognitive, and lifestyle variables, and explicit instructions to generate a concise report without inferring missing values. The template was adapted from the "Patient Information" section of the CARE clinical case reporting guidelines, ensuring consistency with established medical documentation conventions [8]. To minimize variability and preserve numerical fidelity, the prompt enforced a one-to-one mapping between tabular entries and template fields, with unavailable values explicitly marked rather than imputed. This controlled translation process produces semantically enriched text representations that enable structured health information to be embedded within a shared multi-modal latent space.

2.3 Image and Text Encoders

Let $\mathbf{x}_p$ denote the retinal image and $\mathbf{t}_p$ the associated clinical report for subject p. The image encoder $E_I(\cdot)$ and text encoder $E_T(\cdot)$ produce modality-specific embeddings. In our implementation, we instantiate E_I as RETFound [21] and E_T as GatorTron [19]. Each encoder is followed by a projection head, denoted P_I and P_T, to map features into a shared embedding space of dimension d.

$$\mathbf{h}_p^I = E_I(\mathbf{x}_p), \qquad \mathbf{h}_p^T = E_T(\mathbf{t}_p). \tag{1}$$

The encoder outputs are projected into the shared embedding space and ℓ_2-normalized:

$$\hat{\mathbf{z}}_p^I = \frac{P_I(\mathbf{h}_p^I)}{\|P_I(\mathbf{h}_p^I)\|_2}, \qquad \hat{\mathbf{z}}_p^T = \frac{P_T(\mathbf{h}_p^T)}{\|P_T(\mathbf{h}_p^T)\|_2}. \tag{2}$$

2.4 Intra-Modality Similarity for Phenotypic Grouping

To capture phenotypic similarity between subjects, we construct intra-modality similarity matrices based on cosine similarity between normalized projected representations. Let $\hat{\mathbf{z}}_p^I$ denote the normalized projected image embedding for subject p, and let $\hat{\mathbf{z}}_p^T$ denote the normalized projected text-derived risk-profile embedding for subject p.

$$S_{ii}(p,q) = \langle \hat{\mathbf{z}}_p^I, \hat{\mathbf{z}}_q^I \rangle \tag{3}$$

$$S_{tt}(p,q) = \langle \hat{\mathbf{z}}_p^T, \hat{\mathbf{z}}_q^T \rangle \tag{4}$$

Here, $S_{ii}(p,q)$ captures similarity between the learned retinal image embeddings of subjects p and q, while $S_{tt}(p,q)$ captures similarity between their clinical report embeddings.

2.5 Differentiable Phenotypic Weighting

We transform these similarities into soft membership signals using sigmoid gating with thresholds τ_F, τ_T and learnable sharpness parameters g_F, g_T:

$$a_F(p,q) = \sigma\left(\frac{S_{ii}(p,q) - \tau_F}{g_F}\right), \qquad a_T(p,q) = \sigma\left(\frac{S_{tt}(p,q) - \tau_T}{g_T}\right), \tag{5}$$

where $\sigma(\cdot)$ denotes the logistic sigmoid function. Finally, we combine the two signals using a differentiable probabilistic union operator to obtain the phenotypic weighting score:

$$W_{pq} = 1 - \big(1 - a_F(p,q)\big)\big(1 - a_T(p,q)\big), \qquad W_{pq} \in [0,1]. \tag{6}$$

Pairs with larger W_{pq} are treated as more strongly aligned in phenotype space and receive higher weight as positives in the multi-positive contrastive objective.

2.6 Phenotypic Similarity-Weighted Multi-Positive Contrastive Loss

Cross-modal similarity between image and text embeddings is defined as:

$$S_{it}(p,q) = \left\langle \hat{\mathbf{z}}_p^I, \hat{\mathbf{z}}_q^T \right\rangle. \tag{7}$$

Logits are computed using temperature scaling, where the learnable temperature is parameterized as $T_c = \exp(-s)$, together with a learnable bias term β:

$$\ell_{pq} = \frac{S_{it}(p,q)}{T_c} - \beta, \qquad T_c = \exp(-s). \tag{8}$$

We optimize a soft-target multi-positive contrastive objective:

$$\mathcal{L}_{\mathrm{MP}} = \frac{1}{N^2} \sum_{p=1}^{N} \sum_{q=1}^{N} \left[W_{pq} \, \log\big(1 + \exp(-\ell_{pq})\big) + \big(1 - W_{pq}\big) \, \log\big(1 + \exp(\ell_{pq})\big) \right]. \tag{9}$$

Table 1. Cohort characteristics across data splits.

	Train (n=30,462)	**Validation** (n=3,384)	**Test** (n=5,396)
Gender: (male %)	45.10	45.41	45.10
Age: mean (s.d)	55.53 (8.24)	55.78 (8.12)	55.52 (8.17)
Ethnicity: (British %)	84.08	83.51	88.51

When W_{pq} approaches 1, the pair (p, q) is treated as a positive match, when W_{pq} approaches 0, it is treated as a negative pair. Intermediate values allow soft supervision based on phenotypic similarity.

3 Experiments

3.1 Dataset and Preprocessing

A comprehensive set of demographic, behavioral, cognitive, and lifestyle variables was extracted from the UK Biobank [5] baseline assessment and compiled as candidate risk factors based on established epidemiological and biomarker evidence linking modifiable exposures to Alzheimer's disease and dementia risk [14,2,18,10,11,16]. These include factors associated with amyloid and tau pathology, sleep disturbance, cardiometabolic health, and other modifiable determinants of neurodegeneration.

Color fundus photographs (CFPs) from the initial UK Biobank assessment visit were used for image-based modeling. Images underwent automated quality control to exclude low-quality scans, retaining only high-quality CFPs for downstream analysis [22]. Preprocessed CFPs are input into a RETFound-initialized vision encoder, which is fine-tuned during training [21]. Each image was resized to match the input resolution of the pretrained RETFound encoder and normalized using standard channel-wise mean and standard deviation values consistent with its pretraining setup. To ensure consistent anatomical orientation across subjects, right-eye images were horizontally flipped prior to encoding.

Downstream incident AD prediction was formulated as a subject-level binary classification task. A total of 39,242 participants with high-quality CFPs were included and allocated into training (n=30,462), validation (n=3,384), and test (n=5,396) sets. Participants free of prevalent and incident AD were split between each set, and each participant who later developed AD was put into the test pool. The training set was used for representation alignment, the validation set for hyperparameter optimization, and the test set for downstream incident AD prediction. To form the downstream evaluation cohort, controls were sampled from the non-AD test-set participants to obtain an approximate 8% AD prevalence while maintaining age- and gender-matched distributions. From this cohort (n=1,163), 931 subjects (862 controls, 69 AD) were used to train SVM models with 5-fold cross-validation. The remaining 232 subjects (215 controls, 17 AD) were held out as an independent test set for AD prediction.

3.2 Implementation Details

RETFound and GatorTron were used as image and text encoders. The vision encoder is initialized with RETFound weights and fine-tuned end-to-end, while the text encoder is kept frozen. Linear projections map both modalities into a shared $d = 1024$-dimensional space. Embeddings were ℓ_2-normalized prior to similarity computation. The batch size was 128. Optimization used AdamW with hyperparameters selected via Optuna [1]. The final learning rate was 2.42×10^{-4}, $\epsilon = 8.61 \times 10^{-7}$, and weight decay 0.0232. Phenotypic similarity thresholds were initialized from empirical intra-modality cosine similarity distributions computed on the development set, restricting the search to the upper interquartile range.

4 Results

To evaluate the proposed framework, we compared our method against strong retinal and biomedical foundation models. We included RETFound [21], a large-scale retinal image foundation model; RET-CLIP [7], a retinal image–text contrastive pretraining framework; and KeepFIT-CFP [17], a knowledge-enhanced retinal vision–language model introduced with MM-Retinal. We additionally evaluated two general biomedical multi-modal foundation models: PMC-CLIP [15] and BiomedCLIP [20]. Because RETFound is an image-only encoder, we paired it with GatorTron [19] to construct multi-modal representations, concatenating image and text embeddings for downstream classification. In addition to vision–language baselines, we trained tabular SVM models using structured clinical variables and CFP-derived image features to assess whether performance gains stem from semantic narrative modeling or solely from image foundation representations. All methods were evaluated under an identical multi-modal SVM protocol. Each experiment was repeated across 10 random seeds, and we report mean ± standard deviation performance.

In the incident AD prediction task (Table 2), our phenotypic-weighted multi-positive contrastive framework consistently outperformed all comparison methods, indicating that soft, differentiable phenotypic alignment leads to more coherent multi-modal representations. Rather than relying on single positive pairs or hard grouping, the proposed formulation allows subjects with similar risk profiles to contribute proportionally during training, yielding stronger downstream discrimination. While pretrained vision–language baselines such as RETFound+GatorTron and RET-CLIP capture meaningful retinal–text correspondences, they do not explicitly model phenotypic structure, which may be important for long-horizon neurodegenerative risk prediction.

These findings suggest that modeling phenotypic similarity as a continuous, differentiable weighting mechanism enables smoother transitions between positive and negative supervision, leading to more coherent multi-modal embedding spaces and improved long-horizon neurodegenerative risk prediction.

Table 2. Comparison of multi-modal and baseline methods for incident AD prediction. Results are reported as mean ± standard deviation across 10 random seeds. **Bold** and Underline represent the best and the second best results.

	AUROC	Balanced Accuracy	F1-Score	MCC
Baseline SVM	0.593±0.068	0.574±0.083	0.140±0.089	0.076±0.099
KeepFIT-CFP	0.490±0.063	0.505±0.0412	0.099±0.034	0.002±0.046
BiomedCLIP	0.525±0.064	0.522±0.060	0.121±0.052	0.023±0.054
RETCLIP	0.558±0.076	0.527±0.042	0.106±0.069	0.028±0.051
PMC-CLIP	0.471±0.049	0.484±0.020	0.076±0.023	-0.022±0.023
RETFound + GatorTron	0.642±0.052	0.581±0.069	0.185±0.099	0.119±0.101
REVEAL (no GACL)	0.654±0.092	0.602±0.075	0.205±0.096	0.144±0.105
REVEAL (with GACL)	0.658±0.090	0.609±0.079	0.207±0.100	0.146±0.111
REVEAL++	**0.678±0.061**	**0.613±0.048**	**0.236±0.079**	**0.168±0.088**

5 Discussion

Alzheimer's disease is increasingly understood not as a binary condition but as a long-term neurodegenerative process that evolves over years prior to diagnosis. Pathological changes including amyloid deposition, tau accumulation, vascular dysfunction, and systemic metabolic dysregulation emerge progressively and interact across multiple biological scales before cognitive symptoms become apparent [6,16]. Retinal microvascular alterations and structural remodeling likewise develop along a continuum, reflecting cumulative exposure to systemic and neurodegenerative risk factors. Therefore, similarity between individuals along disease-relevant dimensions is continuous rather than discretely separable.

Hard similarity thresholds impose artificial boundaries on this biological continuum by assigning subjects to fixed phenotypic groups. While such grouping can strengthen contrastive supervision, it implicitly assumes well-defined subtype partitions that may not reflect the underlying progression of the disease. However, such discrete subtype boundaries may not exist during the preclinical stages of neurodegenerative disease, where pathological processes evolve gradually and heterogeneously across individuals. This mismatch can limit the ability of representation learning methods to capture subtle transitions in risk states. In contrast, the proposed differentiable phenotypic weighting mechanism allows phenotypic similarity to modulate supervision strength continuously. Participants with partially overlapping risk profiles or subtly similar retinal signatures contribute proportionally during training, enabling smoother organization of the shared embedding space while preserving meaningful inter-subject variation.

This formulation more closely reflects the pathophysiology of preclinical AD, where risk accumulates gradually and manifests heterogeneously across individuals [12]. By relaxing discrete grouping into continuous supervision, the model is able to represent intermediate phenotypic states that may correspond to early pathological changes. The resulting embedding geometry reflects a continuum of

risk rather than rigid clusters, providing a representation that is both biologically plausible and better suited for early risk stratification.

Although REVEAL++ improves over REVEAL and other retinal vision–language baselines, the observed AUROC should not be interpreted as sufficient for clinical deployment. Rather, these results demonstrate improved representation learning for a difficult, low-prevalence, long-horizon incident AD prediction task.

6 Conclusion

We presented REVEAL++, a differentiable phenotypic alignment framework for multi-modal learning from retinal imaging and clinical risk narratives in preclinical Alzheimer's disease prediction. By replacing discrete threshold-based grouping with continuous similarity-driven supervision, the proposed approach enables phenotypic relationships to be learned jointly with representation alignment, allowing population structure to emerge directly from data. This formulation better captures the gradual and heterogeneous nature of neurodegenerative disease progression and leads to improved risk prediction performance. More broadly, differentiable phenotypic alignment offers a strategy for modeling structured variability in multi-modal biomedical data, with potential applications spanning chronic disease risk prediction, precision medicine, longitudinal health modeling, and large-scale population health analytics across diverse clinical domains.

Acknowledgments and Disclosure of Interests This material is based upon work supported by the National Science Foundation under Grant No. (NSF 2123809). This research has been conducted using the UK Biobank Resource under Application Number 48388. The authors declare that they have no competing interests.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). pp. 2623–2631. ACM, New York, NY, USA (2019). https://doi.org/10.1145/3292500.3330701
2. Aktan Süzgün, M., Tang, Q., Stefani, A.: Sleep abnormalities and risk of alzheimer's disease. Current Neurology and Neuroscience Reports **25**(1), 67 (2025). https://doi.org/10.1007/s11910-025-01451-5
3. Banna, H., Slayo, M., Armitage, J., Del Rosal, B., Vocale, L., Spencer, S.: Imaging the eye as a window to brain health: frontier approaches and future directions. Journal of Neuroinflammation **21**(1), 309 (2024). https://doi.org/10.1186/s12974-024-03304-3
4. Bueno Lopez, C., Iona, A., Avery, D., Turnbull, I., Yang, L., Du, H., Chen, Y., Zhang, N., Chen, J., Pei, P., Lv, J., Yu, C., Sun, D., Li, L., Bennett, D., van Duijn,

C., Clarke, R., Chen, Z., Bragg, F.: Cardiometabolic health and risk of dementia and brain atrophy: a community-based prospective cohort study of 0.5 million adults in china. The Lancet Regional Health – Western Pacific **64**, 101743 (2025). https://doi.org/10.1016/j.lanwpc.2025.101743

5. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J.: The uk biobank resource with deep phenotyping and genomic data. Nature **562**(7726), 203–209 (2018). https://doi.org/10.1038/s41586-018-0579-z
6. Chow, K.H.M., Abel, T.: Neurodevelopmental origins of age-related neurodegenerative diseases. eBioMedicine **124**, 106151 (2026). https://doi.org/10.1016/j.ebiom.2026.106151
7. Du, J., Guo, J., Zhang, W., Yang, S., Liu, H., Li, H., Wang, N.: Ret-clip: A retinal image foundation model pre-trained with clinical diagnostic reports. arXiv preprint arXiv:2405.14137 (2024)
8. Gagnier, J.J., Kienle, G., Altman, D.G., Moher, D., Sox, H., Riley, D., Group, C.: The care guidelines: Consensus-based clinical case reporting guideline development. Global Advances in Health and Medicine **2**(5), 38–43 (September 2013). https://doi.org/10.7453/gahmj.2013.008
9. Grattafiori, A., Dubey, A., Jauhri, A., et al.: The llama 3 herd of models. arXiv preprint arXiv:2407.21783 (2024), https://arxiv.org/abs/2407.21783
10. Hayden, K.M., Mielke, M.M., Evans, J.K., Neiberg, R., Molina-Henry, D., Culkin, M., Marcovina, S., Johnson, K.C., Carmichael, O.T., Rapp, S.R., Sachs, B.C., Ding, J., Shappell, H., Wagenknecht, L., Luchsinger, J.A., Espeland, M.A.: Association between modifiable risk factors and levels of blood-based biomarkers of alzheimer's and related dementias in the look ahead cohort. JAR Life **13**, 1–21 (January 2024). https://doi.org/10.14283/jarlife.2024.1
11. Huszár, Z., Solomon, A., Engh, M.A., Koszovácz, V., Terebessy, T., Molnár, Z., Hegyi, P., Horváth, A., Mangialasche, F., Kivipelto, M., Csukly, G.: Association of modifiable risk factors with progression to dementia in relation to amyloid and tau pathology. Alzheimer's Research & Therapy **16**, 238 (October 2024). https://doi.org/10.1186/s13195-024-01602-9
12. Jack, C.R.J., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R.: Nia-aa research framework: Toward a biological definition of alzheimer's disease. Alzheimer's & Dementia **14**(4), 535–562 (2018). https://doi.org/10.1016/j.jalz.2018.02.018
13. Leem, S., Gu, L., You, C., Gong, K., Fang, R.: REVEAL: Multimodal vision–language alignment of retinal morphometry and clinical risks for incident AD and dementia prediction. In: Medical Imaging with Deep Learning (2026), https://openreview.net/pdf?id=aOKAXRHXVw, accepted by MIDL 2026. Proceedings of Machine Learning Research (PMLR)
14. Leshner, A.I., Landis, S., Stroud, C., Downey, A. (eds.): Preventing Cognitive Decline and Dementia: A Way Forward. National Academies Press, Washington, DC (September 2017). https://doi.org/10.17226/24782
15. Lin, W., Zhao, Z., Zhang, X., Wu, C., Zhang, Y., Wang, Y., Xie, W.: Pmc-clip: Contrastive language-image pre-training using biomedical documents. arXiv preprint arXiv:2303.07240 (2023)

16. Livingston, G., Huntley, J., Liu, K.Y., Costafreda, S.G., Selbæk, G., Alladi, S., Ames, D., Banerjee, S., Burns, A., Brayne, C., Fox, N.C., Ferri, C.P., Gitlin, L.N., Howard, R., Kales, H.C., Kivimäki, M., Larson, E.B., Nakasujja, N., Rockwood, K., Samus, Q., Shirai, K., Singh-Manoux, A., Schneider, L.S., Walsh, S., Yao, Y., Sommerlad, A., Mukadam, N.: Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. The Lancet **404**(10452), 572–628 (August 2024). https://doi.org/10.1016/S0140-6736(24)01296-0
17. Wu, R., Zhang, C., Zhang, J., Zhou, Y., Zhou, T., Fu, H.: Mm-retinal: Knowledge-enhanced foundational pretraining with fundus image-text expertise. arXiv preprint arXiv:2405.11793 (2024)
18. Xiong, J., Bhimani, R., Carney-Anderson, L.: Review of risk factors associated with biomarkers for alzheimer disease. Journal of Neuroscience Nursing **55**(3), 103–109 (June 2023). https://doi.org/10.1097/JNN.0000000000000705
19. Yang, X., Chen, A., PourNejatian, N., Shin, H.C., Smith, K.E., Parisien, C., Compas, C., Martin, C., Costa, A.B., Flores, M.G., Zhang, Y., Magoc, T., Harle, C.A., Lipori, G., Mitchell, D.A., Hogan, W.R., Shenkman, E.A., Bian, J., Wu, Y.: A large language model for electronic health records. npj Digital Medicine **5**(1), 1–9 (December 2022). https://doi.org/10.1038/s41746-022-00742-2
20. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., Wong, C., Tupini, A., Wang, Y., Mazzola, M., Shukla, S., Liden, L., Gao, J., Crabtree, A., Piening, B., Bifulco, C., Lungren, M.P., Naumann, T., Wang, S., Poon, H.: Biomedclip: A multimodal biomedical foundation model pretrained from scientific image-text pairs. arXiv preprint arXiv:2303.00915 (2025)
21. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., Kihara, Y., Altmann, A., Lee, A.Y., Topol, E.J., Denniston, A.K., Alexander, D.C., Keane, P.A.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (October 2023). https://doi.org/10.1038/s41586-023-06555-x
22. Zhou, Y., Wagner, S.K., Chia, M.A., Zhao, A., Woodward-Court, P., Xu, M., Struyven, R., Alexander, D.C., Keane, P.A.: Automorph: Automated retinal vascular morphology quantification via a deep learning pipeline. Translational Vision Science & Technology **11**(7), 12 (July 2022). https://doi.org/10.1167/tvst.11.7.12, https://doi.org/10.1167/tvst.11.7.12