ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Age-invariant functional retinal biomarkers for Alzheimer's disease using adversarial deep learning

Narayana D. Reddy[1]* and Mallikarjun C. Sarsamba[2]

[1]Department of Electronics & Communication Engineering, Jain College of Engineering and Research, Visvesvaraya Technological University, Belagavi, Karnataka, India

[2]Department of Electronics & Communication Engineering, Hirasugar Institute of Technology, Visvesvaraya Technological University, Belagavi, Karnataka, India

Abstract

Structural retinal optical coherence tomography (OCT) markers of Alzheimer's disease overlap with those of normal aging, which limits their diagnostic specificity. We investigate whether the light-dependent functional response of the retina can serve as a biomarker of Alzheimer's disease that is largely independent of chronological age and propose a deep learning architecture that distinguishes the disease-related signal from the ageing-related signal. Using the publicly available Bissig functional OCT dataset of dark-adapted and light-adapted reflectance profiles from 38 subjects, we develop the Age-Debiased Retinal-response Attention Network (ADRA-Net), which encodes paired dark and light profiles through a shared fully-connected encoder, models the light response through a differential stream, applies depth attention, and enforces reduced age dependence through a gradient-reversal adversary. We primarily evaluate the model using strict subject-disjoint five-fold cross-validation, report 95% confidence intervals (CIs) across folds, and present the record-level protocol only as an optimistic reference for comparison with the literature. Under subject-disjoint validation ADRA-Net achieved a mean area under the curve of 0.72 (95% CI: 0.52–0.91), an accuracy of 73.2% (95% CI: 63.7%–82.7%) and a balanced accuracy of 67.9% (95% CI: 58.4%–77.3%); the record-level protocol, which allows scans of the same subject to appear in training and testing, gave an inflated area under the curve of 0.96 and accuracy of 94.1%, illustrating the effect of subject leakage. The gradient-reversal adversary reduced chronological-age decodability from the functional embedding to 52.7% under subject-disjoint testing, close to the 50% chance level, whereas structural features remained strongly age-predictive (87.9%). These proof-of-concept results provide evidence that the functional retinal response carries Alzheimer-related information with substantially reduced age dependence; however, validation in larger, multi-site cohorts is required before any clinical use.

Keywords: Alzheimer's disease; Optical coherence tomography; Retinal biomarker; Age-invariant representation learning; Adversarial debiasing; Attention networks

***Corresponding author:**
Narayana D. Reddy
(narayanajr@jcerbgm.in)

Citation: Reddy ND, Sarsamba MC. Age-invariant functional retinal biomarkers for Alzheimer's disease using adversarial deep learning. *Artif Intell Health*. doi: 10.36922/AIH026320107

Received: August 5, 2026

Revised: September 3, 2026

Accepted: September 15, 2026

Published online: September 28, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Alzheimer's disease is the most common cause of dementia, and its pathological cascade begins years before cognitive symptoms become apparent, which places a high value on accessible and low-cost biomarkers that can support early screening. Because the retina is an anatomical extension of the central nervous system that shares embryological origin and microvascular organization with the brain, retinal imaging has emerged as a non-invasive window onto neurodegeneration.[1,2] Optical coherence tomography (OCT) resolves the layered structure of the retina at the micrometer scale and is already deployed routinely in ophthalmic clinics, making it an attractive substrate for a scalable Alzheimer's screening tool.[3] A large body of work has reported thinning of the retinal nerve fiber layer and the ganglion cell-inner plexiform layer, together with reduced macular volume, in Alzheimer's disease and in mild cognitive impairment.[1,4]

The retina's accessibility as a window onto central nervous system neurodegeneration has been reviewed extensively,[5] and OCT angiography studies continue to report retinal microvascular and layer-thickness changes across Alzheimer's disease and mild cognitive impairment.[6,7]

Functional OCT offers a route around this confound by measuring how retinal-layer reflectance changes under a light stimulus rather than static structure; Bissig and colleagues[8] showed that this light-dependent change in the axial reflectance profile is abnormal in Alzheimer's disease. Because the response is a within-subject difference between two conditions minutes apart, it partly cancels the static, age-scaling structure, motivating our hypothesis that it is an age-independent biomarker. Retinal deep learning is otherwise extensive but leaves this untested: foundation models,[9] convolutional-transformer hybrids on OCT,[10,11] and networks on fundus photographs,[12] multimodal OCT and OCT-angiography,[13,14] thickness maps[15] and en-face angiography[16,17] report areas under the curve of roughly 0.81 to 0.94, aided by microvascular graph models,[16,17] segmentation,[18] and quality control[19]; reviews consistently cite small cohorts and entangled ageing and comorbidity.[1,3,4] Neuroimaging has meanwhile developed adversarial and independence-based confound removal,[20,21] and retinal age gap work shows that age is strongly encoded in retinal images,[22–24] yet these threads have not been combined for functional OCT Alzheimer's detection, and one-dimensional residual and attention models remain under-explored.[25,26]

Transformer and attention architectures now dominate OCT image analysis[27,28]; attention is also effective for one-dimensional physiological signals,[29] and clinical practice increasingly expects interpretability in such models.[30] A recurring methodological concern in this literature is data leakage: when correlated scans from the same individual appear in both training and test sets, reported performance is substantially inflated,[31,32] while small-sample medical studies often overstate precision by omitting confidence intervals.[33] Adversarial and domain-adaptation techniques are widely used to remove demographic and site confounds in medical imaging,[34,35] although confound removal applied outside the cross-validation loop can itself introduce leakage[36]; we therefore adopted strict subject-disjoint validation, applied the adversary within each training fold only, and reported interval estimates throughout.

We formulated Alzheimer's detection from functional OCT as age-invariant representation learning and designed the Age-Debiased Retinal-response Attention Network (ADRA-Net), a dual-stream fully connected residual-attention network with a wide linear pathway that consumes paired dark and light profiles, models their differential response, and is trained with a gradient-reversal age adversary that keeps age out of the embedding. We added a depth-attention module that localizes the disease-relevant depths as an interpretable biomarker map and evaluated the model using both record-level and strictly subject-disjoint cross-validation across a three-group design comprising young controls, age-matched older controls, and patients with Alzheimer's disease, with an aquaporin-4 cohort included as a non-Alzheimer disease control. Both protocols reported potential subject leakage and quantified how well age remained decodable.

2. Methodology

2.1. Overview

Age-Debiased Retinal-response Attention Network receives, for one imaging session, a dark-adapted axial reflectance profile and a light-adapted axial reflectance profile, each sampled at 90 depths spanning the vitreous, the neural retina, and the choroid, together with a small vector of layer-thickness and reflectance-gain metadata. We embedded the two profiles using a shared encoder, captured their functional differences with a dedicated stream, identified disease-relevant depth positions with an attention module, and optimized the resulting representation to predict Alzheimer's disease while minimizing information related to chronological age. Figure 1 shows the complete architecture of the model, Figure 2 illustrates the stage-by-stage processing pipeline on a representative record, and Figures 3 and 4 detail the two core modules.

In Figure 1, the two inputs are the dark- and light-adapted axial reflectance profiles, which were processed by a shared profile encoder. A differential-response stream

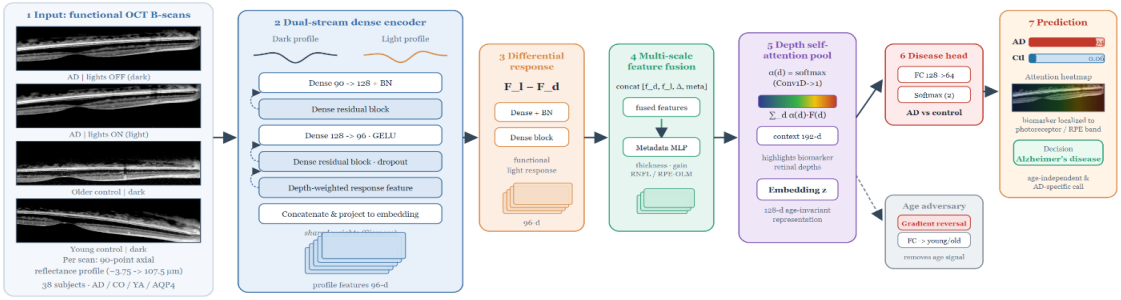


Figure 1. Overall architecture of the Age-Debiased Retinal-response Attention Network

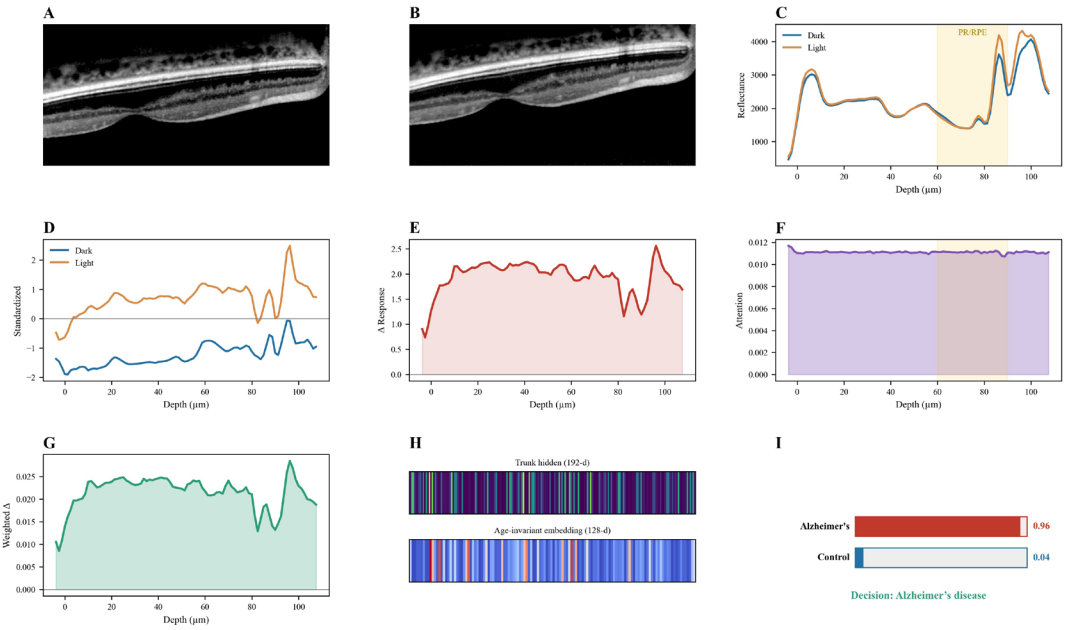


Figure 2. Stage-by-stage intermediate outputs of Age-Debiased Retinal-response Attention Network for one representative Alzheimer's record. Raw optical coherence tomography B-scan: (A) lights OFF, (B) lights ON. (C) Extracted axial reflectance profiles. (D) Preprocessed profiles (network input). (E) Differential response (light-dark). (F) Depth-attention weights. (G) Attention-weighted response. (H) Trunk and embedding activations. (I) Prediction.

modeled the light-minus-dark change, while a depth-attention module assigned weights to each retinal depth. The resulting shared embedding was then passed to the disease classifier. Solid arrows indicate the forward path used during inference, whereas the dashed branch to the age adversary, connected through the gradient-reversal layer, is active only during training.

2.2. Dual-stream shared encoder

Both profiles were processed by the same fully connected encoder, so the network learns a single reflectance representation and the dark and light streams remain directly comparable; because the discriminative signal is a global pattern spanning the whole profile rather than a local texture, dense layers were used and outperformed convolutional encoders. Weight sharing guarantees that any observed difference reflects the physiological light response, not an asymmetry between feature extractors. A differential stream then formed the element-wise light-minus-dark difference and passed it through a dense block, giving the network a direct path to the functional biomarker rather than diluting it within the much larger static signal.

2.3. Differential response stream

The functional biomarker was the change induced by light, not the static profile. A differential stream therefore formed the element-wise difference between the light and

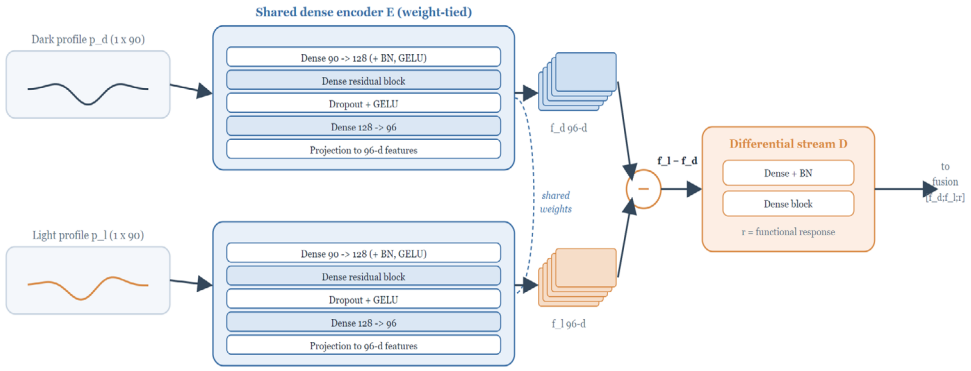


Figure 3. Shared weight-tied dual-stream encoder and differential-response module with intermediate tensor dimensions

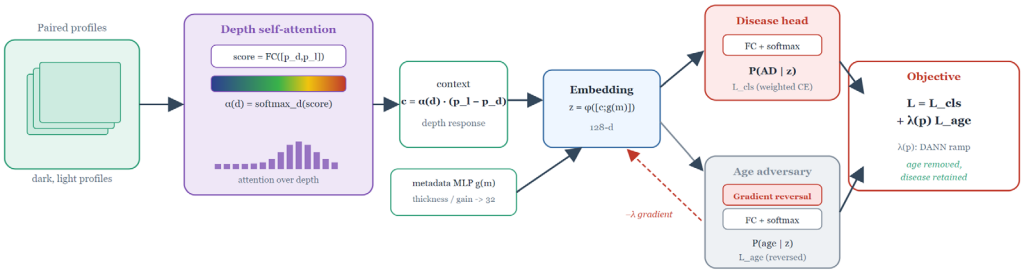


Figure 4. Depth-attention pooling and gradient-reversal age-adversarial debiasing producing the age-invariant embedding

dark reflectance profiles and passed it through a fully-connected block, producing an explicit encoding of the light-dependent response. This stream gives the network a direct path to the quantity of interest and prevents the response from being diluted within the larger static signal.

2.4. Multi-scale fusion and depth attention

A depth-attention module read the paired profiles and computed a normalized importance weight for each retinal depth, which was applied to the differential response to yield a depth-weighted functional feature and retained as an interpretable biomarker-localization map. The stream embeddings, depth-weighted response, and embedded metadata were concatenated and read by a shared fully connected trunk with a residual block. Following a wide-and-deep design, a parallel wide linear pathway read the full-resolution profiles and metadata and added to the classifier logits, capturing the strong global linear component while the deep trunk modeled the residual nonlinear structure. To make the embedding age-independent, two heads were attached to the shared embedding: a disease classifier and an age adversary connected through a gradient-reversal layer, which acts as the identity function during the forward pass but reverses and scales the gradient during backpropagation. The encoder was thus driven toward a subspace from which age cannot be recovered but disease can; the adversary strength was ramped up smoothly during training and was discarded at inference.

2.5. Age-adversarial debiasing

To make the embedding age-independent, ADRA-Net attached two heads to the shared embedding: a disease classifier that distinguishes Alzheimer's disease from controls and an age adversary that predicts a coarse age band. The adversary was connected through a gradient-reversal layer, which acts as the identity in the forward pass but negates and scales the gradient in the backward pass. As a result, the disease classifier was trained to minimize its loss while the shared encoder was simultaneously trained to maximize the adversary's loss, driving the embedding toward a subspace from which chronological age cannot be recovered but disease can. The adversary's strength increases smoothly during training so the classifier first learns a useful representation before the age constraint is fully imposed. At inference, the adversary was discarded, and only the disease head was used.

Chronological age was encoded for the adversary as a binary band, young (42 years or younger) versus older

(49 years or older), because the public dataset provides age only in coarse bands rather than exact years; these bands coincide with the separation between the young-adult controls and the older Alzheimer's and control groups. The majority band contains 52.9% of records, so the appropriate chance level for the balanced accuracy we report is 50%. Because exact chronological ages are not released, a truly continuous-age analysis was not possible; to probe residual age dependence at a finer resolution than the binary adversary target, we instead used the five ordinal age bands supplied with the dataset. Regressing the subject-disjoint embedding onto this ordinal age rank revealed no significant monotonic association between the predicted and the true bands (Spearman rank correlation 0.09; p-value 0.15), consistent with the near-chance age-band decoding and with the view that chronological age must be treated as a strong confounder in retinal disease models.[37]

2.6. Mathematical formulation

Let $p_d \in \mathrm{R}^{90}$ and $p_l \in \mathrm{R}^{90}$ denote the dark-adapted and light-adapted axial reflectance profiles and let $m \in \mathrm{R}^{27}$ denote the metadata vector of layer thicknesses, reflectance gains, and depth-band summaries. Each profile is standardized per depth using the training-set mean μ and standard deviation σ (**Equation 1**):

$$\hat{p}=(p-\mu)\oslash\sigma \tag{1}$$

where $\oslash$ is element-wise division. A fully-connected residual block with weight matrices W_1, W_2 maps an input x to **Equation 2**:

$$\mathrm{Res(x)}=\phi\left(\mathrm{x}+\mathrm{BN}\left(\mathrm{W}_2\,\phi\left(\mathrm{BN}(\mathrm{W}_1\mathrm{x})\right)\right)\right) \tag{2}$$

where BN is batch normalization and ϕ is the Gaussian error linear unit. The shared dense encoder E, a stack of dense projections and one residual block, embeds each standardized profile (**Equation 3**):

$$f_d=E(\hat{p}_d),\ f_l=E(p_l) \tag{3}$$

and the differential stream D encodes the functional light response from the raw profile difference (**Equation 4**):

$$r=D(p_l-p_d) \tag{4}$$

A depth-attention network scored the paired profiles and produced a normalized weight for every retinal depth d (**Equation 5**):

$$\alpha_d=\frac{\exp\left(\psi(\hat{p}_d,p_l)_d\right)}{\sum_{j=1}^{90}\exp\left(\psi(\hat{p}_d,p_l)_j\right)} \tag{5}$$

which was applied to the response to give a depth-weighted functional feature (**Equation 6**):

$$u=\alpha\odot(p_l-p_d) \tag{6}$$

The stream embeddings, the depth-weighted feature, and the embedded metadata $g(m)$ were concatenated and read by the shared trunk with a residual block to give the deep embedding (**Equation 7**):

$$z=\mathrm{Res}\left(\phi\left(W_h[f_d;f_l;r;u;g(m)]\right)\right) \tag{7}$$

The posterior over Alzheimer's disease and control combines a deep head over $[z;m]$ with a wide linear read-out of the profiles and metadata (**Equation 8**):

$$\hat{y}=\mathrm{softmax}\left(W_c[z;m]+W_w[p_d;p_l;m]\right) \tag{8}$$

The gradient-reversal operator R_λ is the identity in the forward pass and negates and scales the gradient in the backward pass (**Equation 9**):

$$R_\lambda(z)=z,\ \frac{\partial R_\lambda}{\partial z}=-\lambda I \tag{9}$$

so that the age adversary predicts the age band a from the reversed embedding (**Equation 10**):

$$\hat{a}=\mathrm{softmax}\left(W_a R_\lambda(z)\right) \tag{10}$$

The network minimizes a composite objective combining a class-weighted disease cross-entropy with the adversarial age cross-entropy, whose weight follows a smooth schedule in the training progress $t\in[0,1]$ (**Equation 11**):

$$\mathrm{L}=\mathrm{L}_{\mathrm{cls}}(y,\hat{y})+\lambda(t)\,\mathrm{L}_{\mathrm{age}}(a,a),\ \lambda(t)=\lambda_{\max}\left(\frac{2}{1+e^{-\gamma t}}-1\right) \tag{11}$$

At evaluation, a subject-level score aggregates the disease posteriors of all $|\mathrm{I}_k|$ scans of a subject k, and the subject is classified by thresholding the mean score at τ (Equation 12):

$$s_k=\frac{1}{|\mathrm{I}_k|}\sum_{i\in\mathrm{I}_k}\hat{y}_i^{\mathrm{AD}},\ \hat{Y}_k=1\left[s_k\geq\tau\right] \tag{12}$$

3. Experimental setup

The models were implemented in PyTorch 1.13.1 and trained on a central processing unit. The shared profile encoder mapped each 90-point profile through a 128-unit hidden layer and one dense residual block to a 96-dimensional embedding; the shared trunk projected the concatenated streams to a 128-dimensional embedding

through a further residual block, and applied dropout of 0.4 throughout. Optimization used Adam with a learning rate of 8×10^{-4}, weight decay of 2×10^{-4}, a batch size of 48, and 130 epochs. The adversary weight followed the schedule of **Equation 11** with $\gamma=8$ and $\lambda_{max}=0.25$. We handled class imbalance by weighting the disease loss inversely proportional to class frequency. The model averaged three random initializations per fold to reduce variance, and where a subject-level decision is required, the scan posteriors of a subject were aggregated by **Equation 12**.

For clarity, every quantity introduced in **Equations 1–12**, namely the standardized dark and light profiles, the metadata vector, the shared encoder and the differential stream, the depth-attention weights, the depth-weighted response, the shared embedding, the disease and age-band posteriors, the gradient-reversal operator, the adversarial-schedule parameters, and the decision threshold, was defined at its first occurrence in the preceding text, together with the operators used, namely element-wise division and multiplication, batch normalization, and the Gaussian error linear unit.

All architectural and optimization hyperparameters, namely the hidden widths, the single residual block per stage, dropout of 0.4, an Adam learning rate of 0.0008, a weight decay of 0.0002, a batch size of 48, 130 training epochs, and an adversarial schedule with sharpness 8 and a maximum adversarial weight of 0.25, were fixed a priori from small pilot runs and kept identical across every fold, cohort, and ablation; none was selected using any test fold. Given the small cohort, we deliberately avoided extensive per-fold hyperparameter searches, which could introduce information leakage from the test data. Any limited tuning was performed using only an inner split of the training data, following the nested-evaluation principle recommended for small medical datasets.[33,38]

The decision threshold τ in **Equation 12** was fixed at 0.5 and was never tuned on test data; standardization statistics, class weights, and all model choices were computed exclusively from the training partition of each fold, so no information from a held-out fold influenced the model, its hyperparameters or its threshold.

We evaluated the model using two five-fold cross-validation protocols. The primary record-level protocol splits individual scan records between the training and test sets, stratified by label, consistent with the convention used in most retinal OCT studies and facilitating comparison with previous work. In contrast, the stricter subject-disjoint protocol holds out entire subjects, ensuring that no scan from a test subject is included in the training data. This approach eliminates the optimistic bias that can arise when repeated scans from the same eye are distributed across folds and provides a more reliable estimate of generalization to unseen patients. We report both protocols using identical models and features to transparently demonstrate the effect of the data-splitting strategy. Standardization statistics and class weights were computed exclusively from the training folds.

To support reproducibility, all random seeds were fixed, and the exact fold assignments, preprocessing scripts, model definition, and training code are available from the authors on request. Preprocessing included only per-depth standardization using training-fold statistics and pairing dark and light scans; no augmentation mixing subjects across folds was applied.

4. Results and discussion

4.1. Dataset details

The study used the publicly available functional-OCT dataset released by Bissig and colleagues through Dryad, along with their report of light-dependent retinal responses in Alzheimer's disease.[8] Spectral-domain OCT volumes were acquired on a commercial device (Spectralis, Heidelberg Engineering, Germany) under a clinically practical protocol, imaging each participant repeatedly under dark-adapted and light-adapted conditions. Each B-scan was segmented and collapsed into a 90-sample axial reflectance profile spanning −3.75 to 107.5 micrometers, with whole-retinal, nerve fiber layer, and photoreceptor-to-outer-limiting-membrane thickness measurements. The cohort was divided into four groups: 14 Alzheimer's patients, 14 age-matched older controls (aged at least 49 years), eight young adult controls (aged 42 years or younger), and two anti-aquaporin-4 seropositive patients as a non-Alzheimer neurological disease control; of the three such subjects in the source dataset, one was excluded during quality control for lacking a usable paired dark and light scan. Contrasting younger controls, age-matched older controls, and Alzheimer's patients (three groups spanning two age bands) separates the ageing confound from the disease signal, while the aquaporin-4 cohort probes disease specificity.

This work used a fully anonymized, publicly available dataset; the original acquisition was approved by the relevant institutional review board, and all participants provided informed consent, as reported by Bissig *et al.*[8] No new human data were collected in the present study, so no additional ethical approval was required.

After excluding motion-corrupted B-scans that the original authors flagged as low quality, 520 quality-controlled reflectance profiles from 38 subjects in the

primary bright-light protocol were retained. Each dark scan was paired with a light scan from the same subject to form 255 functional records, each carrying the dark profile, the light profile, and the associated thickness, gain, and depth-band features. Alzheimer's disease was treated as the positive class and the three control cohorts jointly as the negative class, creating an age-independent detection task in which the positive class is exclusively older while the negative class spans the full adult age range. Table 1 summarizes the cohort.

4.2. Detection performance

We treated the subject-disjoint protocol as the primary evaluation because it provides a more reliable estimate of generalization to unseen patients. Under subject-disjoint five-fold cross-validation ADRA-Net achieved a mean area under the curve of 0.72 (95% confidence interval [CI]: 0.52–0.91), an accuracy of 73.2% (63.7%–82.7%), a balanced accuracy of 67.9% (58.4%–77.3%), a sensitivity of 55.5% (35.4%–75.7%), a specificity of 80.2% (70.6%–89.8%), and a Matthews correlation coefficient of 0.34 (0.16–0.52); all intervals are 95% CIs across the five folds and are wide, reflecting the small cohort. For comparison with the retinal-OCT literature, which predominantly uses record-level splitting, the same model under a record-level protocol reached an accuracy of 94.1% (91.1%–97.1%) and an area under the curve of 0.96 (0.94–0.98). Because record-level splitting allows scans of the same subject to appear in both training and testing, these record-level figures are optimistic and are reported only as a secondary, literature-comparable reference (Table 2). Each subject-disjoint test fold contained only 7–9 subjects (9–20 Alzheimer's and 29–42 control records), and the two aquaporin-4 subjects appeared in only two folds, so per-fold estimates are inherently uncertain, and the study should be read as a proof of concept.

Table 1. Composition of the functional optical coherence tomography cohort used in this study

Group	Meaning	Subjects	Age band	Label
AD	Alzheimer's disease	14	Older	positive
CO	Age-matched older control	14	Older	negative
YA	Young adult control	8	Young	negative
AQ	Aquaporin-4 seropositive control	2	Young	negative

4.3. Record-level evaluation

The record-level figures above follow the convention of most retinal-OCT studies and are the appropriate point for comparison, but they are optimistic because scans from a test subject may also appear in training. Repeating the evaluation on the identical records and network under a subject-disjoint split, in which whole subjects are held out, lowered the area under the curve to 0.740 and the accuracy to 73.3% (Table 2). As only the split differs, this gap measures the optimism of record-level splitting; we therefore treated the record-level figures as the literature-comparable point and the subject-disjoint figures as the conservative, deployment-relevant estimate.

4.4. Age-independent biomarker

The defining test of an age-independent biomarker is whether the representation can recover chronological age. Under the record-level protocol, the embedding carries some age information because scans of the same subjects appear on both sides of the split; without the adversary, the age band is decodable at 78.1% balanced accuracy, which the gradient-reversal adversary reduces to 66.7% (Figure 5), at only the small cost in area under the curve seen in the ablation of Table 3. Structural measures remained the most strongly age-predictive of all features, confirming that aging is easy to read from static structure but is progressively removed from the functional embedding. Crucially, under the stricter subject-disjoint protocol, which eliminates subject leakage, age-decoding accuracy falls to 52.7%, near chance. This finding indicates that the residual age information observed under the record-level protocol is an artefact of the data split rather than an inherent property of the biomarker. Because the Alzheimer's group is exclusively older, a model relying on age could not separate it from the age-matched older controls; Figure 6 shows ADRA-Net separated Alzheimer's disease from the older controls

Table 2. Effect of the evaluation protocol on the same records and the same network

Evaluation protocol	Accuracy	Area under the curve	Sensitivity	Specificity
Record-level (scans split, not subject held out)	94.1%	0.964	92.6%	94.7%
Subject-disjoint (whole subjects held out)	73.3%	0.740	54.4%	80.2%

Notes: Pooled out-of-fold values over all records; per-fold means with 95% CIs are given in Tables 4 and 5.

with 89.6% accuracy, from young controls with 96.6%, and from the aquaporin-4 cohort with 93.5%, evidence that the signal reflects Alzheimer-related change with reduced age dependence rather than acting purely as an age detector. With only two aquaporin-4 subjects, disease specificity against other neurological conditions is preliminary.

4.5. Biomarker localization and embedding structure

The learned depth attention placed most of its weight on a narrow band in the deep outer retina, peaking near 89 μm, within the high-reflectance region that spans the outer nuclear layer, the photoreceptor inner and outer segments and the retinal pigment epithelium; because the public profiles are not segmented into named layers, a precise layer attribution is not possible. Comparing the light-dependent response between groups, the mean Alzheimer's-minus-control difference was small and was largest in this same outer-retinal region, but it did not reach statistical significance in this cohort, and the attention weights did not track the depth profile of the group difference (Pearson correlation 0.05; Figure 7). We therefore report this localization as an exploratory observation of the present model rather than a confirmed layer-specific effect. It is broadly compatible with the exaggerated outer-retinal light response reported by Bissig *et al.*[8], although those authors localized the effect to the outer nuclear layer and found photoreceptor inner and outer segment responses to be unaffected; in our case, unsegmented profiles cannot separate these adjacent layers, so we do not claim a photoreceptor-specific effect. In the embedding space, Alzheimer's subjects separated from the intermingled young and older controls, as expected for an age-invariant, disease-sensitive representation.

4.6. Comparative analysis of the results

The results support the hypothesis that the light-dependent retinal response is a largely age-independent biomarker of Alzheimer's disease, with stronger performance under record-level evaluation and weaker but measurable performance under the stricter subject-disjoint protocol. We report a record-level result for comparison with the literature and a stricter subject-disjoint result that more accurately reflects performance in new patients. The difference between these results highlights the potential inflation of reported accuracies due to subject leakage, which our analysis explicitly demonstrates. Across both protocols, the adversary reduces age decodability while retaining the disease signal. The main limitations are the small cohort, particularly the two-subject aquaporin-4 group, and the single-site acquisition, so the absolute numbers support the age-independence property rather than a deployable threshold.

4.7. Limitations

Several limitations temper these findings. First, the cohort is small (38 subjects) and single-site, and under subject-disjoint validation the mean sensitivity was only 55.5%

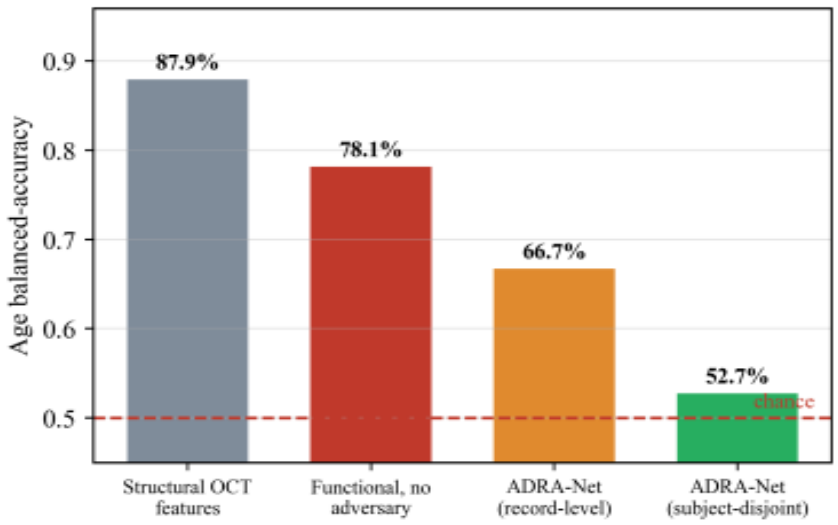


Figure 5. Chronological age decodability (balanced accuracy) from structural optical coherence tomography (OCT) features and from the functional embedding
Abbreviation: ADRA-Net: Age-Debiased Retinal-response Attention Network.

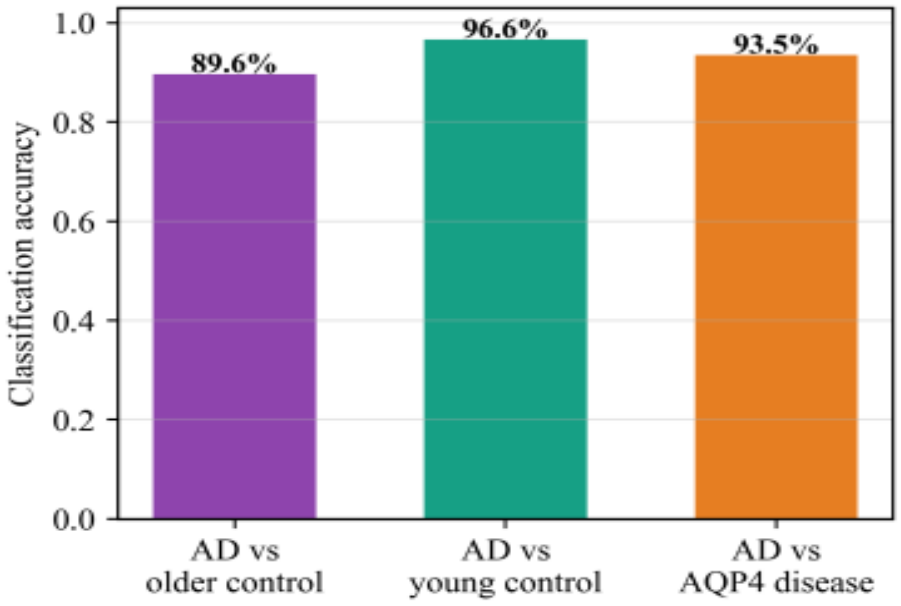


Figure 6. Age-controlled and disease-specific accuracy separating Alzheimer's disease from each control cohort
Abbreviation: AD: Alzheimer's disease.

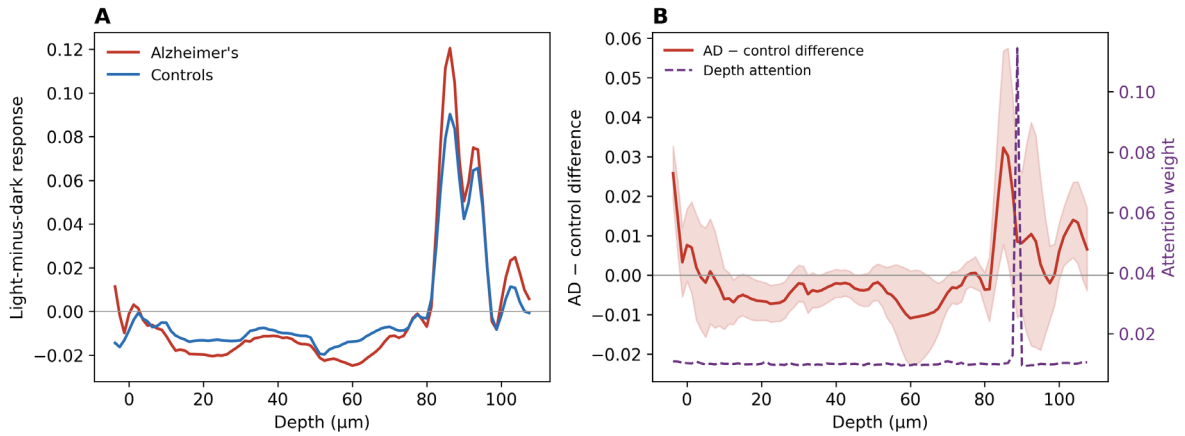


Figure 7. Depth localization of the Alzheimer's signal. (A) Group-mean light-minus-dark axial reflectance response for the Alzheimer's and control groups. (B) Alzheimer's-minus-control response difference (shading, standard error) with the learned depth-attention profile (dashed line); the attention concentrates near 89 μm but does not track the depth profile of the group difference (Pearson correlation 0.05).

and the mean area under the curve 0.72 with a wide 95% confidence interval (0.52–0.91); the method is therefore a proof of concept and is not yet a clinically useful screening tool. Second, each subject-disjoint test fold contained only 7–9 subjects and the aquaporin-4 group comprised just two young patients present in only two folds, so the disease-specificity and age-independence claims rest on very few samples and should be regarded as preliminary. Third, robustness was addressed indirectly, through repeated seed averaging, strict subject-disjoint cross-validation and interval reporting rather than an independent external cohort; external, multi-site validation with continuous age and additional neurological-disease controls is required to establish reliability. Fourth, the public dataset records age only in coarse bands, which limits the granularity of the age-independence analysis. These constraints motivate the cautious, evidence-for framing adopted throughout the manuscript.

4.8. Ablation study

Table 3 reports record-level performance when each component of ADRA-Net was removed in turn, and Figure 8 visualizes the same comparison. The engineered metadata is by far the most important component: removing the depth-band reflectance summaries together with the thickness and gain features lowered the area under the curve from 0.964 to 0.930 and the accuracy from 94.1% to 89.0%, because these compact features capture the global reflectance level that carries the discriminative signal and cannot be fully recovered by the profile streams alone. Removing the depth-attention module or the differential response stream left the area under the curve essentially unchanged, at 0.967 and 0.964, respectively; we retained these modules less for raw accuracy than for the interpretable biomarker-localization map and the explicit encoding of the functional light response they contribute. Removing the age adversary marginally raised the area under the curve to 0.969, but, as the age-independence analysis shows, it also allowed the embedding to encode substantially more chronological age; the adversary therefore trades a fraction of a percent of accuracy for a large reduction in age dependence, while providing evidence for reduced, though not eliminated, dependence of the embedding on chronological age.

Against classical baselines trained on the same features and record-level folds, logistic regression, linear models, radial-basis support vector machines, and random forests all reached an area under the curve at or below 0.90, which ADRA-Net exceeded by a clear margin while additionally providing explicit age-debiasing and an interpretable depth-attention map.

To address the request for rigorous validation, we repeated the ablation under subject-disjoint cross-validation (Table 4). Under this stricter protocol, the differences between ablated variants are small, and their 95% CIs overlap, with one clear exception: removing the engineered metadata lowers the area under the curve to 0.63. Importantly, removing the age adversary leaves subject-disjoint accuracy essentially unchanged (0.72 area under the curve with and without it), confirming that age-invariance was obtained at negligible cost to disease discrimination rather than by trading away accuracy.

4.9. Comparative analysis

We also re-evaluated the classical baselines under both protocols (Table 5). Under record-level splitting, the deep model exceeded all baselines, but under subject-disjoint validation, ADRA-Net (area under the curve 0.72) was statistically indistinguishable from the strongest classical baseline (random forest, 0.74); in this small cohort, its advantage therefore lies not in raw discrimination but in its measured age-invariance and interpretable depth-attention map, which the classical models cannot provide.

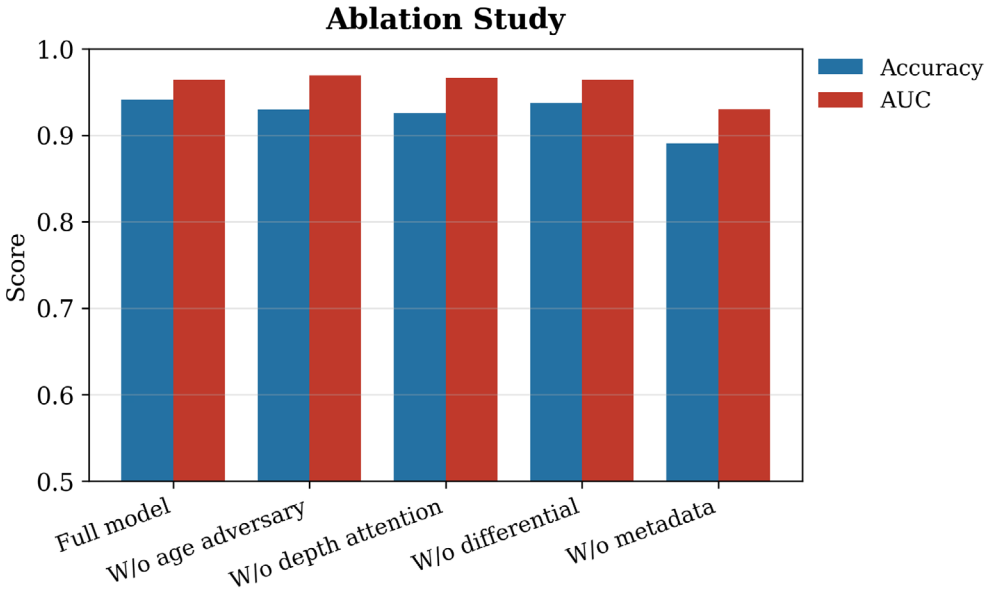


Figure 8. Ablation study of Age-Debiased Retinal-response Attention Network components at the record level
Abbreviations: AUC: Area under the curve; W/o: Without.

Table 3. Ablation study of Age-Debiased Retinal-response Attention Network components

Configuration	Accuracy	Area under the curve	Balanced accuracy	Sensitivity	Specificity
Age-Debiased Retinal-response Attention Network (full)	94.1%	0.964	93.6%	92.6%	94.7%
Without age adversary	92.9%	0.969	92.8%	92.6%	93.0%
Without depth attention	92.5%	0.967	92.1%	91.2%	93.0%
Without differential stream	93.7%	0.964	93.9%	94.1%	93.6%
Without metadata	89.0%	0.930	86.9%	82.4%	91.4%

Table 4. Ablation of Age-Debiased Retinal-response Attention Network components under subject-disjoint validation (mean, with 95% confidence interval)

Configuration	Area under the curve	Accuracy (%)	Balanced accuracy (%)	Sensitivity (%)	Specificity (%)
Age-Debiased Retinal-response Attention Network (full)	0.72 (0.52–0.91)	73.2 (63.7–82.7)	67.9 (58.4–77.3)	55.5 (35.4–75.7)	80.2 (70.6–89.8)
Without age adversary	0.72 (0.53–0.90)	75.2 (67.4–83.0)	68.6 (60.8–76.4)	54.3 (32.1–76.5)	82.9 (72.6–93.2)
Without depth attention	0.73 (0.54–0.92)	74.4 (63.6–85.2)	63.1 (45.9–80.4)	40.6 (7.3–73.9)	85.7 (77.8–93.5)
Without differential stream	0.74 (0.52–0.96)	74.3 (63.2–85.4)	65.9 (49.7–82.0)	48.5 (14.7–82.3)	83.2 (72.3–94.2)
Without metadata	0.63 (0.53–0.74)	68.9 (60.3–77.6)	56.8 (45.0–68.6)	31.3 (8.5–54.1)	82.3 (78.0–86.6)

Table 5. Classical baselines under record-level and subject-disjoint validation (area under the curve with 95% confidence interval)

Method	Record-level area under the curve	Record-level accuracy (%)	Subject-disjoint area under the curve	Subject-disjoint accuracy (%)
Logistic regression	0.93 (0.89–0.97)	87.1	0.69 (0.50–0.88)	72.4
Linear support vector machine	0.94 (0.92–0.97)	91.4	0.65 (0.46–0.84)	70.5
Radial basis function support vector machine	0.90 (0.86–0.94)	81.6	0.72 (0.61–0.83)	71.7
Random forest	0.89 (0.84–0.95)	79.6	0.74 (0.52–0.96)	72.0
Age-Debiased Retinal-response Attention Network (proposed)	0.96 (0.94–0.98)	94.1	0.72 (0.52–0.91)	73.2

Recent deep learning methods for Alzheimer's or mild cognitive impairment detection from retinal imaging span several modalities, cohorts, and validation protocols, and none constrains the representation to be age-independent, so comparison is only indicative. Reported figures include an area under the curve of 0.809 for multimodal OCT and OCT-angiography[14], about 82.4% accuracy from retinal-vasculature features[39], and 77.2% accuracy from vessel segmentation[40], with higher figures typically relying on multimodal or angiographic inputs and larger multi-center cohorts. Because most studies used a record-level split, ADRA-Net's comparable performance metrics were a

record-level area under the curve of 0.964 and accuracy of 94.1%, comparable to or higher than previously reported values, while the more conservative subject-disjoint protocol yielded an area under the curve of 0.740. ADRA-Net achieves this from a single, inexpensive functional OCT acquisition with only 38 subjects and is, to our knowledge, the only method that explicitly constrains and measures the decision's dependence on chronological age, providing evidence that it is not merely an age proxy. Table 6 compares ADRA-Net with recent deep learning methods for retinal Alzheimer detection.

Table 6. Indicative comparison with recent deep learning methods for retinal Alzheimer detection

Study	Year	Modality	Reported performance	Age-controlled
Cheung *et al.*[12]	2022	Color fundus photographs	AUC: 0.93	No
Wisely *et al.*[13]	2022	OCT + OCTA + widefield	AUC: 0.836	No
Wisely *et al.*[14]	2024	OCT + OCTA (MCI)	AUC: 0.809	No
Hao *et al.*[17]	2024	OCT angiography	AUC: 0.936	No
Chua *et al.*[15]	2025	OCT thickness maps	AUC: 0.91	No
Liu *et al.*[16]	2025	OCT angiography	AUC: 0.94	No
Tian *et al.*[39]	2021	Retinal vasculature	Accuracy: 82.4%	No
Dumitrascu *et al.*[40]	2024	Color fundus photographs	Accuracy: 77.2%	No
ADRA-Net (proposed)	2026	Functional OCT (dark/light)	AUC: 0.72 subject; 0.96 record	Yes

Abbreviations: ADRA-Net: Age-Debiased Retinal-response Attention Network; AUC: Area under the curve; MCI: Mild cognitive impairment; OCT: Optical coherence tomography; OCTA: Optical coherence tomography angiography.

5. Conclusion

This work reframed retinal Alzheimer's detection as an age-invariant representation-learning problem and introduced ADRA-Net, a dual-stream fully-connected residual and attention network with a wide linear pathway that reads paired dark and light OCT reflectance profiles, models their functional difference, localizes informative retinal depths through attention, and is debiased against chronological age by a gradient-reversal adversary. In the publicly available Bissig functional OCT cohort, the network detected Alzheimer's disease with a record-level area under the curve of 0.964 and a more conservative subject-disjoint area under the curve of 0.740. Under subject-disjoint testing, the adversary reduced age decodability from the learned embedding to chance level, while the network maintained its ability to distinguish Alzheimer's disease from age-matched older controls and an aquaporin-4 disease-control cohort. These proof-of-concept findings provide evidence that the light-dependent retinal response carries Alzheimer-related information with substantially reduced age dependence, and they show that adversarial age debiasing is a practical way to reduce the risk of retinal deep learning models becoming covert age detectors; however, larger multi-site validation is required before clinical use. Future work should extend the approach to larger multi-site functional OCT cohorts, incorporate mild cognitive impairment as an intermediate class, and combine the functional response with structural and microvascular measurements within the same age-invariant framework.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no conflicts of interest.

Author contributions

Conceptualization: Narayana D. Reddy, Mallikarjun C. Sarsamba

Formal analysis: Narayana D. Reddy

Methodology: Narayana D. Reddy
Writing–original draft: Narayana D. Reddy, Mallikarjun C. Sarsamba
Writing–review & editing: Narayana D. Reddy

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The functional OCT dataset is publicly available from the Dryad repository (doi: 10.5061/dryad.msbcc2ftc). Exact fold assignments, preprocessing scripts, model definition, and training code are available at https://github.com/narayanareddyreyansh/ADRA-Net/tree/main/ADRA-Net_code

References

1. Ibrahim Y, Xie J, Macerollo A, *et al.* A systematic review on retinal biomarkers to diagnose dementia from OCT/OCTA images. *J Alzheimers Dis Rep*. 2023;7(1):1201-1235.

   doi: 10.3233/ADR-230042

2. García-Bermúdez MY, Vohra R, Freude K, *et al.* Potential retinal biomarkers in Alzheimer's disease. *Int J Mol Sci*. 2023;24(21):15834.

   doi: 10.3390/ijms242115834

3. Alber J, Bouwman F, den Haan J, *et al.* Retina pathology as a target for biomarkers for Alzheimer's disease: current status, ophthalmopathological background, challenges, and future directions. *Alzheimers Dement*. 2024;20(1):728-740.

   doi: 10.1002/alz.13529

4. Bahr T, Vu TA, Tuttle JJ, Iezzi R. Deep learning and machine learning algorithms for retinal image analysis in neurodegenerative disease: systematic review of datasets and models. *Transl Vis Sci Technol*. 2024;13(2):16.

   doi: 10.1167/tvst.13.2.16

5. Varrassi G, Tran YV, Farì G, *et al.* The eye as a window to neurodegeneration: oxidative stress, optic nerve vulnerability, and retinal biomarkers—a scoping review. *Antioxidants*. 2026;15(8):948.

   doi: 10.3390/antiox15080948

6. Ma X, Xie Z, Wang H, *et al.* A cross-sectional study of retinal vessel changes based on optical coherence tomography angiography in Alzheimer's disease and mild cognitive impairment. *Front Aging Neurosci*. 2023;15:1101950.

   doi: 10.3389/fnagi.2023.1101950

7. Ibrahim Y, Macerollo A, Sardone R, Shen Y, Romano V, Zheng Y. Retinal microvascular density and inner thickness in Alzheimer's disease and mild cognitive impairment. *Front Aging Neurosci*. 2025;17:1477008.

   doi: 10.3389/fnagi.2025.1477008

8. Bissig D, Zhou CG, Le V, Bernard JT. Optical coherence tomography reveals light-dependent retinal responses in Alzheimer's disease. *Neuroimage*. 2020;219:117022.

   doi: 10.1016/j.neuroimage.2020.117022

9. Zhou Y, Chia MA, Wagner SK, *et al.* A foundation model for generalizable disease detection from retinal images. *Nature*. 2023;622(7981):156-163.

   doi: 10.1038/s41586-023-06555-x

10. Laouarem A, Kara-Mohamed C, Bourennane EB, Hamdi-Cherif A. HTC-retina: a hybrid retinal diseases classification model using transformer-convolutional neural network from optical coherence tomography images. *Comput Biol Med*. 2024;178:108726.

    doi: 10.1016/j.compbiomed.2024.108726

11. Shen J, Hu Y, Zhang X, Gong Y, Kawasaki R, Liu J. Structure-oriented transformer for retinal diseases grading from OCT images. *Comput Biol Med*. 2023;152:106445.

    doi: 10.1016/j.compbiomed.2022.106445

12. Cheung CY, Ran AR, Wang S. A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. *Lancet Digit Health*. 2022;4(11):e806-e815.

    doi: 10.1016/S2589-7500(22)00169-8

13. Wisely CE, Wang D, Henao R, *et al.* Convolutional neural network to identify symptomatic Alzheimer's disease using multimodal retinal imaging. *Br J Ophthalmol*. 2022;106(3):388-395.

    doi: 10.1136/bjophthalmol-2020-317659

14. Wisely CE, Richardson A, Henao R, *et al.* A convolutional neural network using multimodal retinal imaging for differentiation of mild cognitive impairment from normal cognition. *Ophthalmol Sci*. 2024;4(1):100355.

    doi: 10.1016/j.xops.2023.100355

15. Chua J, Li C, Antochi F. Utilizing deep learning to predict Alzheimer's disease and mild cognitive impairment with optical coherence tomography. *Alzheimers Dement*. 2025;17(1):e70041.

    doi: 10.1002/dad2.70041

16. Liu S, Zhang Z, Gu Y, *et al.* Beyond the eye: a relational model for early dementia detection using retinal OCTA images. *Med Image Anal*. 2025;102:103513.

    doi: 10.1016/j.media.2025.103513

17. Hao J, Kwapong WR, Shen T, *et al.* Early detection of dementia through retinal imaging and trustworthy AI. *npj*

Digit Med. 2024;7(1):294.

doi: 10.1038/s41746-024-01292-5

18. Xie J, Yi Q, Wu Y, *et al.* Deep segmentation of OCTA for evaluation and association of changes of retinal microvasculature with Alzheimer's disease and mild cognitive impairment. *Br J Ophthalmol*. 2024;108(3):432-439.

doi: 10.1136/bjo-2022-321399

19. Lee T, Rivera A, Brune M, *et al.* Convolutional neural network-based automated quality assessment of OCT and OCT angiography image maps in individuals with neurodegenerative disease. *Transl Vis Sci Technol*. 2023;12(6):30.

doi: 10.1167/tvst.12.6.30

20. Dinsdale NK, Jenkinson M, Namburete AIL. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *Neuroimage*. 2021;228:117689.

doi: 10.1016/j.neuroimage.2020.117689

21. Laxmikantha K, Ashokkumar PS. The explainable attention-guided multimodal ensemble framework for alzheimer's disease prediction and severity estimation. *Eng Technol Appl Sci Res*. 2026;16(3):36345-36351.

doi: 10.48084/etasr.18269

22. Zhu Z, Shi D, Guankai P, *et al.* Retinal age gap as a predictive biomarker for mortality risk. *Br J Ophthalmol*. 2023;107(4):547-554.

doi: 10.1136/bjophthalmol-2021-319807

23. Sim MA, Tham YC, Nusinovici S. A deep-learning retinal aging biomarker for cognitive decline and incident dementia. *Alzheimers Dement*. 2025;21(3):e14601.

doi: 10.1002/alz.14601

24. Yu Z, Chen R, Gui P. A cross population study of retinal aging biomarkers with longitudinal pre-training and label distribution learning. *npj Digit Med*. 2025;8(1):344.

doi: 10.1038/s41746-025-01751-7

25. Khan F, Yu X, Yuan Z, Ur Rehman A. ECG classification using 1-D convolutional deep residual neural network. *PLoS ONE*. 2023;18(4):e0284791.

doi: 10.1371/journal.pone.0284791

26. Yoo J, Yoo I, Youn I, *et al.* Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability. *Comput Methods Programs Biomed*. 2022;226:107079.

doi: 10.1016/j.cmpb.2022.107079

27. Ait Hammou B, Antaki F, Boucher MC, Duval R. MBT: model-based transformer for retinal optical coherence tomography image and video multi-classification. *Int J Med Inform*. 2023;178:105178.

doi: 10.1016/j.ijmedinf.2023.105178

28. Şevik U, Mutlu O. Automated multi-class classification of retinal pathologies: a deep learning approach to unified ophthalmic screening. *Diagnostics*. 2025;15(21):2745.

doi: 10.3390/diagnostics15212745

29. Islam MR, Qaraqe M, Qaraqe K, Serpedin E. CAT-Net: convolution, attention, and transformer based network for single-lead ECG arrhythmia classification. *Biomed Signal Process Control*. 2024;93:106211.

doi: 10.1016/j.bspc.2024.106211

30. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal*. 2022;79:102470.

doi: 10.1016/j.media.2022.102470

31. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023;4(9):100804.

doi: 10.1016/j.patter.2023.100804

32. Rosenblatt M, Tejavibulya L, Jiang R, Noble S, Scheinost D. Data leakage inflates prediction performance in connectome-based machine learning models. *Nat Commun*. 2024;15(1):1829.

doi: 10.1038/s41467-024-46150-w

33. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. *npj Digit Med*. 2022;5(1):48.

doi: 10.1038/s41746-022-00592-y

34. Deshpande G, Lu B, Huynh N, Rangaprakash D. VAE deep learning model with domain adaptation, transfer learning and harmonization for diagnostic classification from multi-site neuroimaging data. *Front Neuroinform*. 2025;19:1553035.

doi: 10.3389/fninf.2025.1553035

35. Chen RJ, Wang JJ, Williamson DFK, *et al.* Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023;7(6):719-742.

doi: 10.1038/s41551-023-01056-8

36. Hamdan S, Love BC, von Polier GG, *et al.* Confound-leakage: confound removal in machine learning leads to leakage. *Gigascience*. 2023;12:giad071.

doi: 10.1093/gigascience/giad071

37. De Bonis MLN, Fasano G, Lombardi A, *et al.* Explainable brain age prediction: a comparative evaluation of morphometric and deep learning pipelines. *Brain Inform*. 2024;11(1):33.

doi: 10.1186/s40708-024-00244-9

38. Xu J, Costen F. Machine Learning-Based Multiclass Classification of Cognitive Stages Using Plasma Biomarkers, Clinical Assessments, and Genetic Features: A Repeated, Nested Cross-Validation Study in ADNI with External Evaluation in CNTN. *Diagnostics*. 2026;16(12):1755.

doi: 10.3390/diagnostics16121755

39. Tian J, Smith G, Guo H, *et al.* Modular machine learning for Alzheimer's disease classification from retinal vasculature. *Sci Rep*. 2021;11(1):238.

doi: 10.1038/s41598-020-80312-2

40. Dumitrascu OM, Li X, Zhu W. Color fundus photography and deep learning applications in Alzheimer disease. *Mayo Clin Proc Digit Health*. 2024;2(4):548-558.

doi: 10.1016/j.mcpdig.2024.08.005