Scientific Reports

https://doi.org/10.1038/s41598-026-36518-x

Article in Press

A multimodal retinal aging clock for biological age prediction and systemic health assessment via OCT and fundus imaging

Received: 15 August 2025

Accepted: 13 January 2026

Published online: 28 January 2026

Cite this article as: Ludwig C.A., Salvi A., Mesfin Y. *et al.* A multimodal retinal aging clock for biological age prediction and systemic health assessment via OCT and fundus imaging. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-36518-x

Chase A. Ludwig, Anish Salvi, Yeabsira Mesfin, Leo Arnal, Curtis Langlotz & Vinit Mahajan

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

A Multimodal Retinal Aging Clock for Biological Age Prediction and Systemic Health Assessment via OCT and Fundus Imaging

Chase A. Ludwig MD, MS[1, 2*] Anish Salvi MS[1], Yeabsira Mesfin BS[3], Leo Arnal BS[1], Curtis Langlotz MD, PhD[1, 4], Vinit Mahajan MD, PhD[1, 2*]

Affiliations :

[1]School of Medicine, Stanford University, Palo Alto, CA
[2]Department of Ophthalmology, Byers Eye Institute, Stanford University, Palo Alto, CA
[3]School of Medicine, University of California San Francisco, San Francisco, CA
[4]Department of Radiology, Stanford University, Palo Alto, CA

*** Correspondence:**
Chase A. Ludwig, M.D., M.S.,
2370 Watson Court, Office 100D,
Palo Alto, CA 94303
caludwig@stanford.edu

Vinit B. Mahajan, M.D., Ph.D.,
2452 Watson Court, Office 2286,
Palo Alto, CA 94303
vinit.mahajan@stanford.edu

Running Head: Age Clock for Multimodal Imaging

Funding Declaration: This work was supported by the National Eye Institute K23 Grant, K23EY035741 and E. Matlida Ziegler Foundation for the Blind Grant awarded to Chase A. Ludwig as well as the Stanford P30 Vision Research Core Grant, NEI P30-EY026877, and Research to Prevent Blindness, Inc.

ARTICLE IN PRESS

ARTICLE IN PRESS

Abstract

We develop age-clocks which predict biological age from fundus photography and optical coherence tomography. We evaluated our multimodal models' clinical relevance by examining their associations between predicted biological age and the Charlson Comorbidity Index (CCI). The first viewed how retinal deviations from normal eyes influenced the predicted biological age, and the second determined how including retinal disorders impacts the results. Models were fine-tuned to the imaging dataset to predict biological age. Linear-regressors were trained on chronological and biological features to infer CCI. Gradient-weighted regression activation mapping also generated heatmaps to identify the model's region of focus. Prediction performance improved when trained on both normal and diseased eyes. Predicted biological age showed significantly stronger correlations with CCI than chronological age across both studies, supporting our algorithm's association with this validated measure of mortality. Thus, our algorithm may provide insight into systemic health burdens beyond that of traditional risk assessments.

Keywords: Image Regression, Embeddings, Fundus, Optical Coherence Tomography, Biological Age, Charlson Comorbidity Index, Age Clock, Multimodal, Retina

ARTICLE IN PRESS

ARTICLE IN PRESS

Introduction

Accurately assessing a patient's biological age may offer critical insights into their systemic health and risk for adverse outcomes beyond chronological age alone.[1] This may be related to how biological age predictors consider the physiologic and functional status of a patient's biomarkers, while chronological age refers to the numbers of years since their birth.[2] Thus, biological age predictors can account for the differing functional capacities of patients of the same chronological age by acknowledging the heterogeneity in the rates by which patients age. Since biological age integrates biomarkers to estimate the age of a patient, it has been theorized to be a better predictor for risk stratification.[1] The most commonly used biological clock is based on measurement of DNA methylation from a blood sample.[3] Recently, a proteomic clock based was created to account for organ-specific aging using on eye fluid liquid biopsies.[4] A simpler, noninvasive approach like retinal imaging, that includes fundus photography and optical coherence tomography (OCT), offers a window into vascular, neural, and metabolic health, making it an attractive candidate for biological age inference.

Age clocks have been developed to predict biological age from large imaging cohorts, primarily to link aging features with genetic variation.[5,6] Meanwhile, clinical data from quantifiable imaging modalities, such as microvascular abnormalities from retinal photography and OCT angiography, has already been applied in oculomics to potential indicators of impending stroke, coronary heart disease, and Alzheimer's disease.[7-9] However, few studies have connected imaging-derived biological age to clinical measures of comorbidity or systematically examined how ocular disease impacts biological aging estimates.

The most stringent benchmarks for image-based retinal age clocks still come from two very large population studies. Le Goallec et al. trained Inception-v3 and ResNet variant models on ≈ 175,000 UK Biobank fundus + OCT images and reported a test mean absolute error (MAE) of 2.62 years with R-squared (R^2) = 0.836.[5] Ahadi et al. first fit an EyePACS model (≈ 272,000 images) and then inferred upon ≈ 120,000 UK Biobank

ARTICLE IN PRESS

images, achieving a MAE of 3.30 years.[6] Both groups used their clocks chiefly for genome-wide and exposome-wide association mapping rather than for bedside risk stratification; other single-modality work by Poplin et al. and Hassan et al. showed similar technical accuracy (MAE ≈ 3.2 years, 3.3 years, respectively) but likewise stopped short of linking retinal age to clinical comorbidity.[10,11]

More recent studies, on the other hand, have attempted to address this by utilizing these models for risk stratification. Peng et al., successfully established an association between their RetiAGE system and the risk of chronic obstructive pulmonary disease, reporting a C-statistic of 0.809 (95% CI: 0.789-0.828).[12] Meanwhile, a separate model relying on fundus photography to calculate age displayed a direct association not only with the incidence of chronic kidney disease (CKD) but also with that of cardiovascular disease (CVD) among CKD patients. This group found a 3% increase in the risk of CKD and 10% increase in the risk of CVD for every 1-year difference between the predicted biological age and the chronological age.[13] In a different study, Chang et al., utilized the age predictions of their model to stratify patients' risk of atherosclerosis (OR: 1.55, 95% CI: 1.07-2.23) and all-cause mortality (HR: 2.13, 95% CI: 1.08-4.21).[14] Lastly, Zhu et al., demonstrated how their model's discrepancy from the chronological age is independently associated with the risk of metabolic syndrome and chronic inflammation (OR: 1.10, 95% CI: 1.00, 1.21).[15]

To further the translational application of these models, we aim to explore how four prevalent eye diseases—age related macular degeneration (AMD), diabetes with ophthalmic complications (DM), epiretinal membrane (ERM), and pathologic/high myopia (PM/HM)—influence biological age predictions. The global burden of these four eye disorders is projected to increase in the coming decades. Over the next two decades, AMD and diabetic retinopathy are expected to affect nearly 288 million and 160.5 million patients.[16,17] Meanwhile, within that same timeframe, over 4 billion patients are expected to be diagnosed with myopia, and the increased in life expectancy is expected to concurrently raise the prevalence of ERM.[18-20] Each of these ocular disorders directly affect the structural integrity of the retina.[21-24] Thus, in this study, we examine how deviations from normal retinal structures influence biological age inference across these specific eye diseases, and whether incorporating prior knowledge of eye-disease status can improve biological age predictions and their relationship to systemic health burdens. The Charlson Comorbidity Index (CCI) is a validated measure of overall comorbidity burden and is the most commonly used clinical tool for quantifying mortality. This tool calculates the burden of chronic conditions by assigning disease-

ARTICLE IN PRESS

specific weights that help evaluate each condition's contribution to a patient's 1-year mortality.[25-27] Thus, we developed two multimodal models leveraging non-invasive fundus, OCT imaging, and eye-disease presence to infer biological age and predict the CCI. We aim to demonstrate how routine ocular imaging may offer clinicians a low-cost indicator of systemic health risk, supporting conventional preventative care.

Results

Study 1: Model accuracy and evidence of over-fitting

When trained exclusively on normal retinal images, the multimodal network closely reproduced chronological age in the training partition (MAE = 2.85 years; R^2 = 0.96) (**Table 1**). In the held-out test cohort—containing both normal and diseased eyes—performance fell to MAE = 10.63 years and R^2 = 0.57, demonstrating substantial over-fitting to healthy anatomy (**Supplemental Figure 1**).

Error by ocular-disease subgroup

Prediction error rose with increasing structural disruption (**Table 2**). Normal eyes retained the lowest error (MAE = 7.15 years; R^2 = 0.74). Eye images with diabetes-related changes (MAE = 8.74; R^2 = 0.32), ERM (MAE = 8.76; R^2 = 0.45), and PM/HM (MAE = 12.65; R^2 = 0.33) showed decreases in performance. PM/HM pushed the error higher while AMD proved the most challenging, with MAE = 13.35 years and a negative R^2 of −1.74, signifying that a naïve guess at the cohort's mean age would have outperformed the model for AMD eyes.

Study 1: Calibration across chronological-age decades

Inspection of the test-set histograms revealed systematic bias in specific age bands (**Figure 1**). For the overall cohort, the model tended to over-predict around 60 years and under-predict in patients older than that. In normal eyes, over-prediction was concentrated between 40 and 60 years, whereas younger (<40 years) and older (>60 years) participants were underestimated. AMD eyes were underestimated across the entire age range, consistent with their high per-image noise. Diabetic eyes showed a pocket of over-prediction between 60 and 70 years. ERM eyes were slightly over-predicted just after 60 years and under-predicted beyond that. In PM/HM, younger patients were underestimated, while those aged 30 to 60 years were overestimated. These patterns highlight age-band–specific calibration errors that future modeling can target.

Study 1: Systemic relevance, association with comorbidity burden

Across the full test set, CCI correlated strongly with both chronological age (Pearson p = 1.65E-9, Spearman p = 0.04) and biological age (Pearson p = 5.22E-12,

ARTICLE IN PRESS

Spearman p = 1.20E-8) (**Table 1**). Within subgroups, biologically inferred age retained significance in every disease category for the majority of performance metrics except AMD, where neither Pearson (p = 0.17) nor Spearman (p = 0.08) reached significance (**Table 2)**. Chronological age was not significant in two additional settings: normal eyes (Pearson p = 0.09) and diabetic eyes (Pearson p = 0.14), although Spearman tests in both cases remained significant. Overall, biological age captured comorbidity burden more robustly than chronological age, except in AMD, where severe macular pathology confounded the age signal.

Study 1: CCI prediction using linear models

To test whether the eye-derived biological age improves prediction of systemic disease burden, we compared a baseline linear model using only chronological age with augmented models that also incorporated residual age (biological minus chronological) as well as embeddings via principal component analysis (PCA). Across the entire test cohort, the augmented model reduced the mean absolute CCI error from 1.44 to 1.38 points (**Table 3)**. Similar gains appeared in AMD, ERM, and diabetic eyes, whereas chronological age alone remained marginally superior in high myopia and in normal eyes (**Supplemental Table 1**). Although the narrow CCI range kept R^2 values close to zero, the consistent error reduction shows that retinal biological age captures comorbidity information that chronological age misses.

Study 1: Qualitative model attention

Gradient-weighted regression activation mapping (Grad-RAM)saliency maps showed a condition-independent focus on the optic nerve head in virtually all fundus photographs, suggesting that the network relies heavily on axonal and peripapillary vascular cues when inferring age (**Figure 2**). In the corresponding OCT B-scans, the hot-spots migrated axially: some eyes highlighted the nerve-fiber layer (NFL), others the outer retina and retinal pigment epithelium, and in several cases the signal extended even deeper into the choroid–sclera interface. This layered pattern implies that the model is sampling age-related information along the full thickness of the posterior pole rather than fixating on a single structure.

Study 2: Model accuracy and evidence of over-fitting

With disease labels incorporated, the multimodal network again fit the training data almost perfectly (MAE = 2.74 years; R^2 = 0.97) (**Table 1**). Generalization improved markedly over Study 1: in the mixed-disease test cohort the error fell to MAE = 6.00 years and R^2 = 0.86, cutting the test-set MAE nearly in half (**Supplemental Figure 1**).

ARTICLE IN PRESS

Study 2: Error by ocular-disease subgroup

All five diagnostic groups clustered within a narrow performance band (**Table 2**). AMD showed the lowest error (MAE = 5.43 years; R^2 = 0.44), followed by normal eyes (MAE = 5.78; R^2 = 0.83), DM (MAE = 6.18; R^2 = 0.66), ERM (MAE = 6.23; R^2 = 0.69) and PM/HM (MAE = 6.41; R^2 = 0.83).

Study 2: Calibration across chronological-age decades

Unlike Study 1, the biological-age scatter closely tracked the line of identity across the full age span in every subgroup, indicating that the additional training diversity largely removed the decade-specific over- and under-prediction artefacts observed previously (**Figure 3**).

Study 2: Systemic relevance: association with comorbidity burden

For the entire test set, chronological age showed no significant association with CCI (Pearson $p = 0.07$; Spearman $p = 0.20$), whereas biological age remained significant on both tests (Pearson p = 0.01, Spearman p = 0.02, respectively) (**Table 1**). Within subgroups, significance patterns were mixed. Biological age preserved at least one significant correlation in normal eyes (Spearman $p = 1.02E\text{-}6$) and in pathologic myopia (Pearson p = 0.02 and Spearman p = 0.01), but all correlations were non-significant in AMD and in diabetes (**Table 2**). In ERM, only the biological Spearman correlation reached significance ($p = 0.03$). These results suggest that the biological-age signal still captures systemic burden more robustly than chronological age, yet the strength of that link varies by ocular pathology.

Study 2: CCI prediction using linear models

We repeated the simple regression exercise, now giving both models the one-hot eye-disease codes. Across the full test cohort, the baseline chronological-age model and the augmented biological-feature model tied (MAE = 1.12 CCI points; R^2 = 0.17, **Table 3**). At the subgroup level, however, the augmented model delivered the lower MAE in all categories: AMD, ERM, DM, PM/HM, except for normal eyes (**Supplemental Table 2**). Although the absolute gains were small, the direction of improvement was uniform, suggesting that retinal biological-age features may provide incremental information even when disease status is already known.

Study 2: Qualitative model attention

Grad-RAM maps again converged on the optic-nerve head and peripapillary region in fundus photographs, implying reliance on axonal and vascular cues (**Figure 4**). In OCT B-scans, the network most often highlighted the nasal retinal-nerve-fiber layer but extended attention through the outer retina into the choroid and occasionally the sclera,

ARTICLE IN PRESS

indicating that age-related signals are drawn from multiple depth-resolved structures rather than a single retinal layer.

Discussion

We found that a retinal “age-clock” derived from routine fundus photographs and OCT scans captures systemic health burdens more accurately than chronological age. In both experiments, biologically inferred age, or simple derivatives such as residual age and low-dimensional age embeddings, consistently improved the prediction of CCI in eyes with structural disease and in the mixed-disease test cohort. The advantage disappeared only when biological and chronological ages converged, as in structurally normal retinas, indicating that the clock is most informative when retinal tissue deviates from age-matched expectations. To our knowledge, this is the first multimodal ocular model to demonstrate that image-based biological age provides clinically actionable information on comorbidities beyond what a patient’s chronological age conveys.

First, incorporating explicit disease labels during training (Study 2) nearly halved the test-set MAE—from 10.6 to 6.0 years while preserving biological age’s superiority over chronological age for comorbidity prediction in most diseased subgroups. This result shows that prior diagnostic knowledge can be blended with imaging features to boost raw age accuracy without erasing the downstream systemic signal. Second, calibration analyses revealed that the model’s decade-specific biases—over-estimating 40- to 60-year-olds and under-estimating older patients in Study 1 essentially vanished once diseased images were added to the training set. Thus, exposing the network to a broader anatomic spectrum not only improved error metrics but also produced an age scale that aligns more faithfully with the clinical timeline. Third, Grad-RAM saliency maps converged on the optic-nerve head in fundus photographs and on the nasal nerve-fiber layer through to the choroid in OCT B-scans, indicating that the algorithm draws age cues from axonal health and deeper vascular tissues rather than from macular pigment alone; this anatomic consistency held across all eye diseases. Finally, the retinal age clock retained predictive value even when CCI were compressed toward the low end, suggesting that subtle ocular aging signs may surface before systemic comorbidities cross traditional diagnostic thresholds and could therefore serve as an early-warning biomarker.

The convergence of our saliency maps on the optic nerve head, retinal nerve fiber layer, and choroid underscore the clinical significance of these anatomical landmarks in the aging eye. Increased collagen deposition along the extracellular matrix of the posterior pole, astrocyte basement membrane, and lamina cribrosa are hallmarks of

ARTICLE IN PRESS

aging.[28,29] This thickens the sclera and underlying tissue layers, reducing its compliance and raises the susceptibility of the optic nerve to bariatric insults from intraocular pressure spikes.[28] Similarly, the resultant stiffening of the lamina cribrosa, which normally dissipates pressure differences between intravitreal and retrobulbar compartments, exacerbates this susceptibility to glaucomatous damage as the eye ages.[29,30] Nutrients from capillary beds must also travel through the basement membrane of astrocytes before reaching the optic nerve head. Thus, as it stiffens, the subsequent reduction of nutrients further compromises axonal health as patients age.[31] Meanwhile, as the structural integrity of the optic nerve head weakens over time, the thickness of the retinal nerve fiber layer and choroid also decrease.[32,33] One study found the reduction of the retinal nerve fiber layer thickness occurs at a rate of 0.16 μm/year, while the choroid atrophies at a rate of 2.98 μm/year subfoveally.[32,34] Our model's localization to these anatomical landmarks substantiates these prior findings. By converging on the optic nerve head, retinal nerve fiber layer, and choroid, the model reinforces how the degenerative changes of these structures can serve as biomarkers of age.

Of note, our saliency maps did not consistently converge on the macula when predicting biological age, despite the clinical significance of macular changes in aging eyes.[35-38] This omission of the macula may have contributed to our model's poorer performance for AMD and DM patients, given how the pathophysiology of these conditions impact the macula.[39,40] However, our Grad-RAM heatmaps revealed that the focus shifted from the optic disc slightly towards the macula when comparing the overall results for Study 1 to Study 2. This reveals that certain retinal pathologies, like AMD, require disease specific training to capture the disease's biomarkers. As a clinical implication, the difference between biological age and chronological age for AMD as per MAE was greater for Study 1 as compared to the same for Study 2. Consistent with clinical knowledge, the model itself potentially identified disease afflicted anatomy as per the observable trends across the results.

This study has several caveats. First, though the CCI is the most utilized clinical tool for assessing mortality, its calculations are based on a limited number of comorbidities and are derived from patient data from the 1980s. Thus, utilizing it as a reference for our algorithm may have limited the generalizability of our findings for current patient populations and to those with additional comorbidities that may contribute to mortality but are omitted by the CCI.[41] Additionally, the testing and training dataset utilized in this study was limited to a single clinical data warehouse, the

ARTICLE IN PRESS

Stanford Research Repository (STARR), which may also limit our algorithm's generalizability. Because all images and diagnoses came from a single academic center, disease counts, particularly AMD and pathologic myopia in Study 1, were modest, so some subgroup correlations may be imprecise and will benefit from larger, multi-site cohorts for further external validation. Removing the age component of the CCI kept age from "double-counting" but also compressed most scores to 0–1; a more finely graded comorbidity scale could sharpen the signal. Model outputs were not explicitly constrained to non-negative values, occasionally yielding implausible negative ages or CCI estimates, a technical detail that can be corrected with a soft-plus or post-hoc clipping layer. In Study 2, concatenating one-hot disease labels alongside images may have let the network glean age clues from diagnosis prevalence rather than retinal structure alone, so future external tests should mask these labels at inference. Finally, Grad-RAM saliency maps, while reassuringly centered on expected anatomy, remain a coarse interpretability tool and cannot specify the exact tissue changes driving the age signal.

Priority follow-up studies should validate the clock in multi-center cohorts that use diverse imaging hardware and coding practices, track patients longitudinally to determine whether increases in retinal biological age precede incident comorbidities and constrain model outputs while masking disease labels to exclude artefactual cues. Adding complementary data—anterior-segment images, blood biomarkers, or genomics—could extend the method into a multi-organ ageing panel, and finer-scale saliency or histologic work is needed to pinpoint the retinal changes that drive the signal.

Across both experiments, retinal biological age consistently added clinical value beyond chronological age. In Study 1, where the model had no prior knowledge of ocular disease, feature sets built on biological age (either residual age or age-embeddings with six PCA components) cut CCI error for the overall test cohort and for eyes with AMD, ERM, and diabetes, whereas normal and high-myopic eyes showed little gain because their biological and chronological ages already overlapped closely. In Study 2, after we supplied one-hot disease labels, the same pattern re-emerged: biological features again improved CCI prediction in AMD (via residual age), ERM, diabetic eyes (via age-embeddings with 16 PCA components), and pathologic myopia (via residual age), while the advantage vanished in normal eyes. Taken together, these results show that a multimodal retinal "age clock" can flag systemic comorbidity even without explicit

ARTICLE IN PRESS

disease codes and that adding disease context further sharpens its predictive reach in structurally abnormal eyes.

Ultimately, our study continues to fill the translational gap between the prediction of biological age and its clinical application. With only 12,741 ultrawide-field fundus photographs and 7,750 OCT B-scans from 2,467 patients, roughly 1/14 the image volume of the UK Biobank efforts, we still reached a test MAE of 6.00 years and $R^2 = 0.86$ when eye-disease codes were included (Study 2). More importantly, this modest yet clinically rich dataset allowed us to show that biologically inferred age outperforms chronological age for gauging the CCI, especially in structurally abnormal eyes, a link not demonstrated by the larger screens. The single-network architecture also generalized across five distinct ocular diseases, suggesting a robust, pathology-aware age signal. Anatomically, our Grad-RAM maps concentrate on the optic-nerve head, nasal retinal-nerve-fiber layer, and choroid—sites of early microvascular and axonal ageing that lie outside the macula-centric fields used in most prior clocks. By coupling such depth-resolved retinal cues to systemic comorbidity, we shift the paradigm from viewing the eye merely as a “calendar” of chronological ageing to recognizing it as a barometer of latent, whole-body disease burden that may aid in early health risk stratification and personalized patient management.

Methods

Materials and Equipment

A convolutional neural network (CNN), Inception-v4, was adapted for the image regression task.[42] We chose the CNN architecture over that of the vision transformer (ViT) because ViTs require larger datasets while ours was comparatively small.[43] The improved performance of CNNs for smaller datasets has been attributed to their robust inductive bias to extract local spatial features. We specifically chose Inception-v4 because past models were built upon Inception-v3 and InceptionResNet-v2. Via ordinary least squares (OLS), we developed CCI linear regression models built upon chronological age or biological features derived from biological age. When developing CCI linear regression models underpinned on biological age embeddings, we applied principal component analysis (PCA) for dimensionality reduction. To help identify potential biomarkers contributing to biological age, we adapted gradient-weighted class activation mapping (Grad-CAM) to Grad-RAM.[44] Thus, we produced heatmaps which highlight the important regions of anatomy within an image with respect to biological age. This was accomplished by leveraging the gradient of the final convolutional layer

ARTICLE IN PRESS

within a CNN which contains important spatial and semantic information with respect to biological age.

Data Acquisition and Image Sources

Clinical data was obtained from the Stanford Research Repository (STARR), a de-identified clinical data warehouse at Stanford Medicine. The study protocol was reviewed and approved by the Stanford University Institutional Review Board (IRB #39760). Because this was a retrospective analysis of de-identified data, the requirement to obtain informed consent from individual subjects and/or their legal guardians was waived by the Institutional Review Board, in accordance with institutional policy and the Declaration of Helsinki. Fundus images were sourced from the manufacturer OPTOS (Dunfermline, United Kingdom) as red-green channel images of size 4000x4000x3 (12719 images) and 3072x3900x3 (22 images). OCTs were sourced from the manufacturers Carl Zeiss Meditec (Jena, Germany) of size 5x1024x1024 (6389 images) and size 5x1024x4096 (173 images) and Heidelberg Engineering (Heidelberg, Germany) of size 7x496x1536 (1188 images). The eye-disease status of each patient image was determined by internal classification of disease (ICD) 9 & 10 codes, specifically: AMD, ERM, DM, PM/HM, and plaquenil use (normal) without toxic maculopathy (**Supplemental Table 3**).

Clinical Labels and Dataset Construction

Chronological age was computed by subtracting the image acquisition date by the patient birth date **(Supplemental Table 4)**. Comorbidities include, for example, congestive heart failure, peripheral vascular disease, diabetes with chronic complications, and metastatic solid tumors (**Supplemental Table 3**). The CCI was calculated as a weighted sum of the comorbidities the patient presented at the acquisition date (**Supplemental Table 5**). In cases where comorbidities overlap in disease, priority was given to the more severe comorbidity (i.e., CCI points were given) and the less severe comorbidity was ignored (i.e., CCI points were not given) (**Supplemental Table 6**). Our CCI calculation did not include the age component to avoid any bias from chronological age. A dataset was constructed of 2467 patients and 7228 samples with 12741 fundus and 7750 OCT images (**Supplemental Table 7**). The distribution of patient ages by disease were also plotted (**Supplemental Figure 2**). For each sample, the data contained: 1) multiple fundus and OCT images taken within one year of each other; 2) one exclusive eye-disease; 3) chronological age rounded to the nearest whole year; 4) and laterality (i.e., right or left eye), and acquisition specific to

ARTICLE IN PRESS

the patient. After collecting CCIs, we observed a high prevalence of low CCI values (**Supplemental Table 8**).

Image Pre-processing

Each fundus image was preprocessed as follows: 1) taking the first two channels (i.e., red-green), 2) cropping to the region-of-interest (i.e., elimination of background), 3) interpolating to 299x299 pixels, and 4) normalizing to mean 0.50 and standard deviation 0.50. Each OCT image was preprocessed by: 1) taking the center image depth-wise, 2) cropping to the region-of-interest (i.e., anatomy center), 3) interpolating to 299x299 pixels, and 4) normalizing channel-wise to mean 0.50 and standard deviation 0.50.[5]

Deep-Learning Model Training and Cross-Validation

We applied transfer learning with no freezing of weights for the deep learning regression task. Pretrained on ImageNet weights, Inception-v4 was fine-tuned to the multimodal dataset.[45] The dataset was segmented into clusters for chronological age, CCI, images per sample, OCT manufacturer, and disease status (Study 2 only). For stratification, OCT manufacturer and disease status were already represented as categorical values. Continuous quantities such as chronological age, CCI, and images per sample were segmented into different groups based upon cutoff intervals to generate categorical values. These separate categorical values were combined for stratification (proportional splitting between folds). Cross-validation was applied by stratifying clusters while each fold was an independent patient cohort. The eye-diseases and chronological ages were normalized by the mean and standard deviation of the training sets. Steps to reduce overfitting included: 1) early stopping, 2) learning rate scheduler, and 3) image augmentation. The model was fit according to mean-squared error (MSE) loss. If multiple fundus or OCT images were taken for a single sample, the fundus and OCT pair was selected at random during the training process. Via 7-fold cross validation, the first study's model was trained on normal and tested on normal and disease fundus and OCT images with the normalized chronological age as the target. Via 5-fold cross-validation, the second study's model was trained on normal and disease samples of fundus, OCT, and normalized one-hot eye-disease encodings and tested on the same with the normalized chronological age as the target. Study 1 had an increased number of folds due to its smaller sample size. For Study 1, 7-folds were implemented because the training dataset size was relatively small, being restricted to only normal cases. Each fold would have roughly 86% training v. 14% testing. For Study 2, 5-folds were implemented because the dataset size was much

ARTICLE IN PRESS

larger, including disease conditions. Each fold would have roughly 80% training v. 20% testing. After training, the model predicted age is referred to as the biological age (**Figure 5**).

Linear Regression for CCI Prediction

For the linear regression exercise, two groups of predictors were defined as experimental and control. The experimental group contained features derived from biological age. The control group contained features derived from chronological age, exclusively. Feature sets for biological inference were built on biological age embeddings and the residual age (the biological age minus the chronological age). Embeddings of vector length 256 were extracted from the dense layer of the deep learning model and PCA was performed. Via OLS linear regression models were trained on the experimental or control group features and fit to CCI. For Study 1, the number of PCA components were 2, 6, and 10. For Study 2, the number of PCA components were 8, 12, and 16. Study 2 had a greater PCA values due to additional training samples.

Grad-RAM Heat-map Generation

Heatmaps were produced via an adaptation of Grad-CAM, which we call *Grad-RAM*. Gradients were taken from the model's last convolutional layer and pooled across the channels. Activations were extracted from the last convolutional layer as well. Activations were weighted by the corresponding pooled gradients by channel. The channels of the activations were averaged to generate the heatmap. ReLU was applied to the heatmap to identify the most pertinent spatial and semantic contributions with respect to biological age. Each heatmap was normalized to a maximum value of 1 and interpolated to 299x299 pixels, the same as the input fundus and OCT. For smoothing, the heatmap underwent gaussian filtering.

Inference Procedure and Evaluation Metrics

Under the assumption that the first fundus taken on a specific acquisition date was centered on the fovea, the first fundus by acquisition time was used for image inference. Under the assumption that the last OCT taken on a specific acquisition date was of high quality, the last OCT by acquisition time was used for image inference. While the inference selection criteria may introduce bias, we ensure clinically meaningful results by following the advice of a vitreoretinal surgeon. For each fold, the deep learning regressors predicted the biological age logits from the fundus, OCT, and eye-disease encoding (Study 2 only). These logits were converted into biological age as per the mean and standard deviation of the chronological age obtained from the training set per each fold. The mean squared error (MSE), root mean squared error (RMSE), MAE, and R-

ARTICLE IN PRESS

squared (R^2) were computed between the chronological age and biological age. The same process was repeated for disease subgroups. The distributions of the overall test set and by disease of the biological vs. chronological age were quantified as well. The Pearson's r and Spearman's r correlation coefficients were computed between 1) CCI and chronological age and 2) CCI and biological age. The Pearson's r quantifies linear relationships whereas the Spearman's r quantifies monotonic relationships (i.e., is nonparametric). A significant correlation was characterized as $p < 0.05$. For the linear regression exercise, the metrics of MAE and R^2 were implemented as well. Results were computed independently for each fold before being averaged across training and test sets for disease. The heatmaps were generated by Grad-RAM for the fundus and OCT encoders, independently. The overall results were compared to findings in literature (**Supplemental Table 9**).

ARTICLE IN PRESS

ARTICLE IN PRESS

References

1 Rollandi, G. A. *et al.* Biological age versus chronological age in the prevention of age associated diseases. *OBM Geriatrics* **3**, 1–11 (2019).

2 Salih, A., Nichols, T., Szabo, L., Petersen, S. E. & Raisi-Estabragh, Z. Conceptual overview of biological age estimation. *Aging and disease* **14**, 583 (2023).

3 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844 (2016).

4 Wolf, J. *et al.* Liquid-biopsy proteomics combined with AI identifies cellular drivers of eye aging and disease in vivo. *Cell* **186**, 4868–4884. e4812 (2023).

5 Le Goallec, A., Diai, S., Collin, S., Vincent, T. & Patel, C. J. Deep learning of fundus and optical coherence tomography images enables identification of diverse genetic and environmental factors associated with eye aging. *medRxiv*, 2021.2006. 2024.21259471 (2021).

6 Ahadi, S. *et al.* Longitudinal fundus imaging and its genome-wide association analysis provide evidence for a human retinal aging clock. *Elife* **12**, e82364 (2023).

7 Wong, T. Y. *et al.* Retinal microvascular abnormalities and incident stroke: the Atherosclerosis Risk in Communities Study. *The Lancet* **358**, 1134–1140 (2001).

8 Wong, T. Y. *et al.* Retinal arteriolar narrowing and risk of coronary heart disease in men and women: the Atherosclerosis Risk in Communities Study. *Jama* **287**, 1153–1159 (2002).

9 Salobrar-Garcia, E. *et al.* Ocular vascular changes in mild Alzheimer's disease patients: foveal avascular zone, choroidal thickness, and ONH hemoglobin analysis. *Journal of personalized medicine* **10**, 231 (2020).

10 Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature biomedical engineering* **2**, 158–164 (2018).

11 Hassan, O. N. *et al.* in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI).* 238–242 (IEEE).

12 Peng, Q. *et al.* Predictive Potential of Retina-Based Biological Age in Assessing Chronic Obstructive Pulmonary Disease Risk. *Clinical & Experimental Ophthalmology* (2025).

13 Wu, G. *et al.* Association of retinal age gap with chronic kidney disease and subsequent cardiovascular disease sequelae: a cross-sectional and longitudinal study from the UK Biobank. *Clinical Kidney Journal* **17**, sfae088 (2024).

14 Chang, J. *et al.* in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 1179–1181 (IEEE).

15 Zhu, Z. *et al.* The Association of Retinal age gap with metabolic syndrome and inflammation. *Journal of diabetes* **15**, 237–245 (2023).

16 Wong, W. L. *et al.* Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health* **2**, e106–e116 (2014).

ARTICLE IN PRESS

17 Teo, Z. L. *et al.* Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology* **128**, 1580–1591 (2021).

18 Holden, B. A. *et al.* Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology* **123**, 1036–1042 (2016).

19 Bosworth, B. P. & Keys, B. Increased life expectancy: a global perspective. *Coping With Methuselah: The Impact of Molecular Biology on Medicine and Society*, 247–283 (2004).

20 Xiao, W., Chen, X., Yan, W., Zhu, Z. & He, M. Prevalence and risk factors of epiretinal membranes: a systematic review and meta-analysis of population-based studies. *BMJ open* **7**, e014644 (2017).

21 Cho, B.-J., Shin, J. Y. & Yu, H. G. Complications of pathologic myopia. *Eye & contact lens* **42**, 9–15 (2016).

22 Garcia-Garcia, J. *et al.* Pathophysiology of age-related macular degeneration: implications for treatment. *Ophthalmic Research* **65**, 615–636 (2022).

23 Ożóg, M. K., Nowak-Wąs, M. & Rokicki, W. Pathophysiology and clinical aspects of epiretinal membrane–review. *Frontiers in Medicine* **10**, 1121270 (2023).

24 Wang, W. & Lo, A. C. Diabetic retinopathy: pathophysiology and treatments. *International journal of molecular sciences* **19**, 1816 (2018).

25 D'Hoore, W., Bouckaert, A. & Tilquin, C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *Journal of clinical epidemiology* **49**, 1429–1433 (1996).

26 Charlson, M. E., Carrozzino, D., Guidi, J. & Patierno, C. Charlson comorbidity index: a critical review of clinimetric properties. *Psychotherapy and psychosomatics* **91**, 8–35 (2022).

27 Bannay, A. *et al.* The best use of the Charlson comorbidity index with electronic health care database to predict mortality. *Medical care* **54**, 188–194 (2016).

28 Coleman-Belin, J. *et al.* Aging effects on optic nerve neurodegeneration. *International Journal of Molecular Sciences* **24**, 2573 (2023).

29 Albon, J., Karwatowski, W., Avery, N., Easty, D. L. & Duance, V. C. Changes in the collagenous matrix of the aging human lamina cribrosa. *British journal of ophthalmology* **79**, 368–375 (1995).

30 Wang, J. Glaucoma and myopia: risk factors, pathophysiology, and treatment. *Canadian Eye Care Today*, 30–35–30–35 (2022).

31 Downs, J. C. Optic nerve head biomechanics in aging and disease. *Experimental eye research* **133**, 19–29 (2015).

32 Parikh, R. S. *et al.* Normal age-related decay of retinal nerve fiber layer thickness. *Ophthalmology* **114**, 921–926 (2007).

33 Chirco, K., Sohn, E., Stone, E., Tucker, B. & Mullins, R. Structural and molecular changes in the aging choroid: implications for age-related macular degeneration. *Eye* **31**, 10–25 (2017).

34 Wakatsuki, Y., Shinojima, A., Kawamura, A. & Yuzawa, M. Correlation of aging and segmental choroidal thickness measurement using swept source optical coherence tomography in healthy eyes. *PloS one* **10**, e0144156 (2015).

ARTICLE IN PRESS

35 Ikuno, Y., Kawaguchi, K., Nouchi, T. & Yasuno, Y. Choroidal thickness in healthy Japanese subjects. *Investigative ophthalmology & visual science* **51**, 2173–2176 (2010).

36 Nivison-Smith, L. *et al.* Normal aging changes in the choroidal angioarchitecture of the macula. *Scientific reports* **10**, 10810 (2020).

37 Gartner, S. & Henkind, P. Aging and degeneration of the human macula. 1. Outer nuclear layer and photoreceptors. *British Journal of Ophthalmology* **65**, 23–28 (1981).

38 Hogan, M. J. & Alvarado, J. Studies on the human macula: IV. Aging changes in Bruch's membrane. *Archives of ophthalmology* **77**, 410–420 (1967).

39 Bhagat, N., Grigorian, R. A., Tutela, A. & Zarbin, M. A. Diabetic macular edema: pathogenesis and treatment. *Survey of ophthalmology* **54**, 1–32 (2009).

40 Gottlieb, J. L. Age-related macular degeneration. *Jama* **288**, 2233–2236 (2002).

41 Drosdowsky, A. & Gough, K. The Charlson Comorbidity Index: problems with use in epidemiological research. *Journal of clinical epidemiology* **148**, 174–177 (2022).

42 Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. in *Proceedings of the AAAI conference on artificial intelligence.*

43 Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

44 Selvaraju, R. R. *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization. *International journal of computer vision* **128**, 336–359 (2020).

45 Deng, J. *et al.* in *2009 IEEE conference on computer vision and pattern recognition.* 248–255 (Ieee).

ARTICLE IN PRESS

Funding Declaration: This work was supported by the National Eye Institute K23 Grant, K23EY035741 and E. Matlida Ziegler Foundation for the Blind Grant awarded to Chase A. Ludwig as well as the Stanford P30 Vision Research Core Grant, NEI P30-EY026877, and Research to Prevent Blindness, Inc.

Author Contributions: CAL contributed to the conception, design, data acquisition, analysis, and interpretation of the present work, and the writing and revision of the manuscript; AS contributed to the design, analysis, and interpretation of the present work and writing and revision of the manuscript; YM contributed to the interpretation of the present work and the writing and revision of the manuscript; LA contributed to the revision of manuscript; CL contributed to the writing and revision of the manuscript; VM contributed to the design and writing and revision of the manuscript

ARTICLE IN PRESS

Data Availability: The data analyzed in the current study is available from the corresponding author on reasonable request.

Competing Interests: The authors declare no competing interests.

ARTICLE IN PRESS

Figure Captions

Figure 1. Study 1 histograms comparing biological-age predictions with chronological age for (a) the full test cohort (n = 7228 samples), and for eyes with (b) plaquenil use - normal (n = 1943 samples), (c) age-related macular degeneration (n = 1269 samples), (d) diabetes with ophthalmic complications (n = 1465 samples), (e) epiretinal membrane (n = 1253 samples), and (f) pathologic/high myopia (n = 1298 samples); disease-specific predictions are the average across cross-validation folds.

Figure 2. For Study 1, test set eyes with Grad-RAM saliency for red-channel fundus (left), green-channel fundus (center), and greyscale OCT B-scan (right).

Figure 3. Study 2 histograms juxtaposing biological-age predictions with chronological age for (a) the full test cohort (n = 7228 samples) and for eyes with (b) plaquenil use – normal (n = 1943 samples), (c) age-related macular degeneration (n = 1269 samples), (d) diabetes with ophthalmic complications (n = 1465 samples), (e) epiretinal membrane (n = 1253 samples), and (f) pathologic/high myopia (n = 1298 samples).

Figure 4. For Study 2, test set eyes with Grad-RAM saliency for red-channel fundus (left), green-channel fundus (center), and greyscale OCT B-scan (right).

Figure 5. Fundus photographs and optical-coherence-tomography B-scans are encoded and fused—together with one-hot eye-disease labels in Study 2 only (red box)—to produce a biological-age estimate. This output, along with dense-layer embeddings, is related to chronological age and the Charlson Comorbidity Index, while Grad-weighted regression activation maps highlight image regions driving the prediction.

ARTICLE IN PRESS

Tables

Table 1. The performance metrics across training and test sets via 7-fold cross validation for Study 1 and 5-fold cross-validation for Study 2 (n = 7228 samples). Chronological age served as the ground truth while biological age inferred from the deep learning regressor were used to compute mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2). Pearson r and Spearman r were computed between CGI and chronological age as well as CGI and

Study	1		2	
Dataset	Training	Test	Training	Test
MSE [95% CI]	14.87 [14.38, 15.39]	193.69 [190.15, 197.36]	13.64 [13.40, 13.89]	61.62 [59.46, 63.96]
RMSE [95% CI]	3.71 [3.59, 3.84]	13.91 [13.66, 14.18]	3.60 [3.54, 3.67]	7.81 [7.53, 8.10]
MAE [95% CI]	2.85 [2.73, 2.96]	10.63 [10.40, 10.87]	2.74 [2.68, 2.80]	6.00 [5.74, 6.26]
R^2 [95% CI]	0.96 [0.95, 0.96]	0.57 [0.55, 0.58]	0.97 [0.97, 0.97]	0.86 [0.85, 0.88]
Chr. Pearsonr	p = 6.21E-14	p = 1.65E-09	p = 1.56E-15	p = 0.07
Chr. Spearmanr	p = 7.26E-37	p = 0.04	p = 3.61E-14	p = 0.2
Bio. Pearsonr	p = 9.62E-15	p = 5.22E-12	p = 1.19E-15	p = 0.01

ARTICLE IN PRESS

Table 2. The performance metrics across training and test sets by disease with normal (n = 1943 samples), AMD (n = 1269 samples), DM (n = 1465 samples), ERM (n = 1253 samples), and PM/HM (n = 1298 samples) via 7-fold cross validation for Study 1 and 5-fold cross-validation for Study 2 (n = 7228 samples). Chronological age served as the ground truth while biological age inferred from the deep learning regressor were used to compute mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2). Pearson r and Spearman r were computed between CCI

	Study 1: Test					Study 1: Training
	Normal	**AMD**	**DM**	**ERM**	**PM/HM**	**Normal**
MSE [95% CI]	91.37 [84.39, 99.62]	260.47 [250.72, 271.02]	125.20 [120.82, 129.90]	130.16 [125.26, 135.47]	288.48 [277.80, 300.03]	14.87 [14.38, 15.39]
RMSE [95% CI]	9.36 [8.64, 10.21]	16.10 [15.50, 16.75]	11.17 [10.78, 11.59]	11.40 [10.97, 11.86]	16.97 [16.34, 17.65]	3.71 [3.59, 3.84]
MAE [95% CI]	7.15 [6.45, 7.86]	13.35 [12.86, 13.85]	8.74 [8.38, 9.10]	8.76 [8.36, 9.16]	12.65 [12.03, 13.26]	2.85 [2.73, 2.96]
R^2 [95% CI]	0.74 [0.68, 0.78]	-1.74 [NA, NA]	0.32 [0.28, 0.36]	0.45 [0.41, 0.49]	0.33 [0.29, 0.37]	0.96 [0.95, 0.96]
Chr. Pearsonr	p = 0.09	p = 0.62	p = 0.14	p = 5.86E-05	p = 2.82E-10	p = 6.21E-14
Chr. Spearmanr	p = 0.002	p = 0.92	p = 0.001	p = 7.79E-16	p = 1.47E-22	p = 7.26E-37
Bio. Pearsonr	p = 0.09	p = 0.17	p = 0.001	p = 0.0004	p = 0.0003	p = 9.62E-21 15
Bio. Spearmanr	p = 0.02	p = 0.08	p = 6.85E-05	p = 9.77E-09	p = 1.25E-10	p = 1.10E-35

ARTICLE IN PRESS

ARTICLE IN PRESS

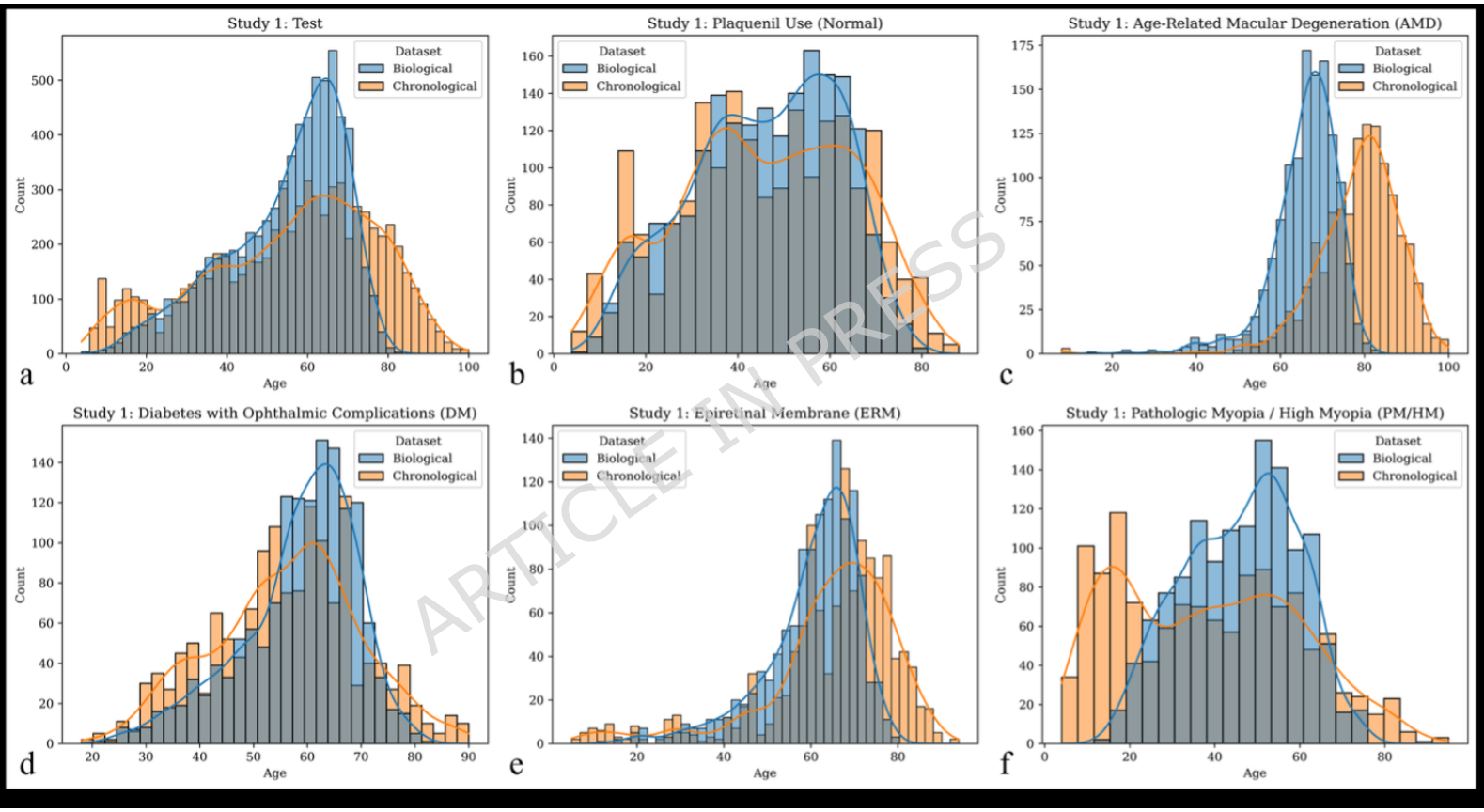
Table 3. The performance metrics of the CCI linear regressor for the training and test sets via 7-fold cross validation for Study 1 and 5-fold cross-validation for Study 2 (n = 7228 samples). The feature set involved chronological age (chr), biological age (bio), residual age (res) which is chr subtracted from bio, age embeddings (emb), and principal component analysis (pca) usage with 6 or 16 components. Greater predictive power was determined by having a lower test set mean absolute error

Study 1					
Inference	**Features**	**Training**		**Test**	
		MAE [95% CI]	**R^2 [95% CI]**	**MAE [95% CI]**	**R^2 [95% C**
chronological	chr	1.28 [1.21, 1.36]	0.04 [0.03, 0.06]	1.44 [1.41, 1.48]	-0.02 [-0.02 0.01]
biological	chr+bio	1.28 [1.21, 1.35]	0.04 [0.03, 0.06]	1.42 [1.38, 1.46]	0.00 [0.00 0.01]
	bio	1.28 [1.21, 1.36]	0.04 [0.03, 0.06]	1.41 [1.38, 1.45]	0.01 [0.01 0.02]
	res	1.32 [1.25, 1.40]	0.00 [0.00, 0.01]	1.38 [1.34, 1.42]	-0.06 [-0.07 0.05]
	chr+bio+emb+pca 6	1.28 [1.21, 1.35]	0.05 [0.03, 0.08]	1.45 [1.41, 1.49]	-0.01 [-0.02 0.01]
	chr+bio+res+emb+pca 6	1.28 [1.21, 1.35]	0.05 [0.03, 0.08]	1.45 [1.41, 1.49]	-0.01 [-0.02 0.01]
	chr+emb+pca 6	1.28 [1.21, 1.35]	0.05 [0.03, 0.08]	1.45 [1.41, 1.49]	-0.01 [-0.02 0.01]
Greater Predictive Power		Biological			
Study 2					
		Training		**Test**	

ARTICLE IN PRESS

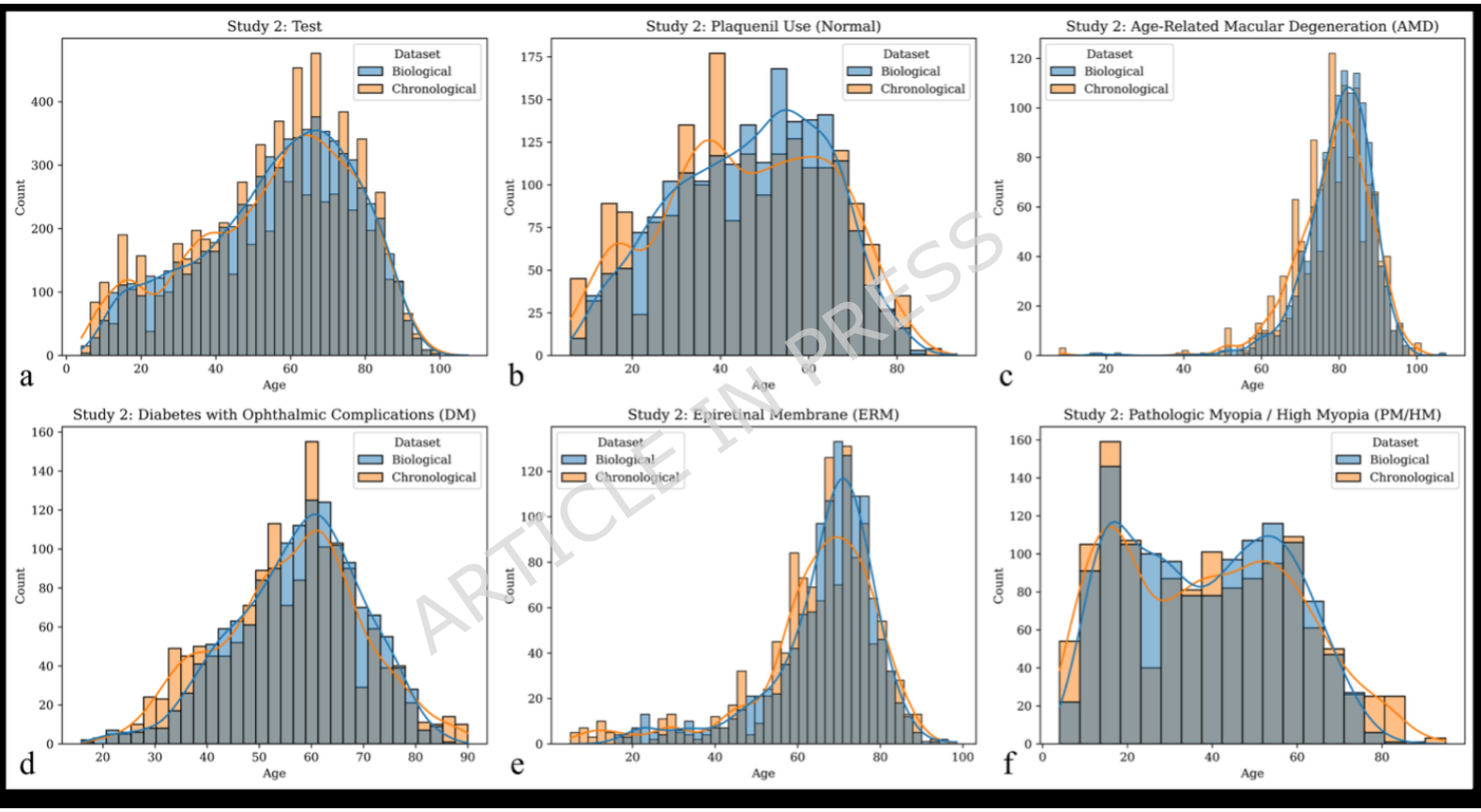
ARTICLE IN PRESS

ARTICLE IN PRESS



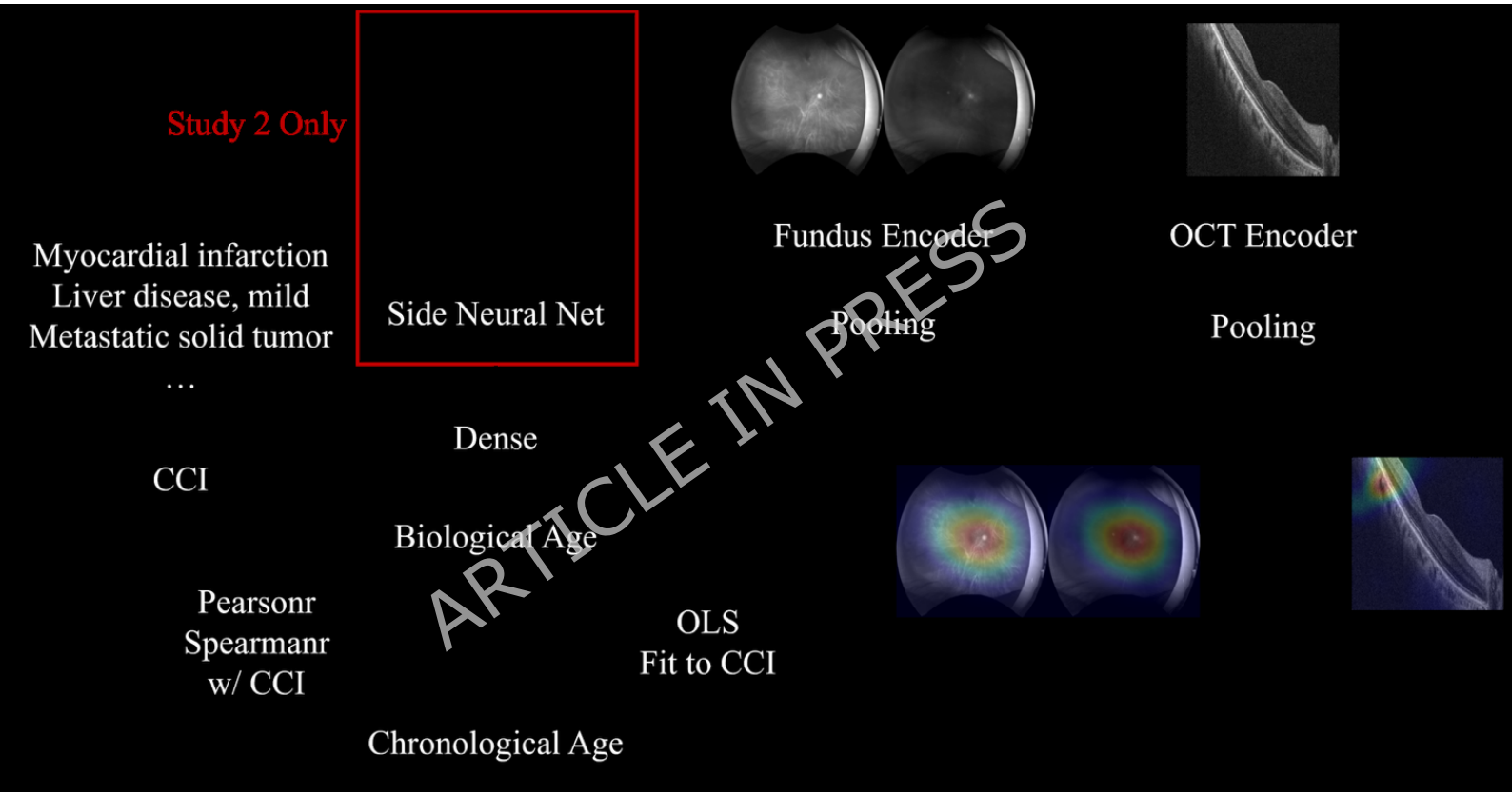
ARTICLE IN PRESS

ARTICLE IN PRESS



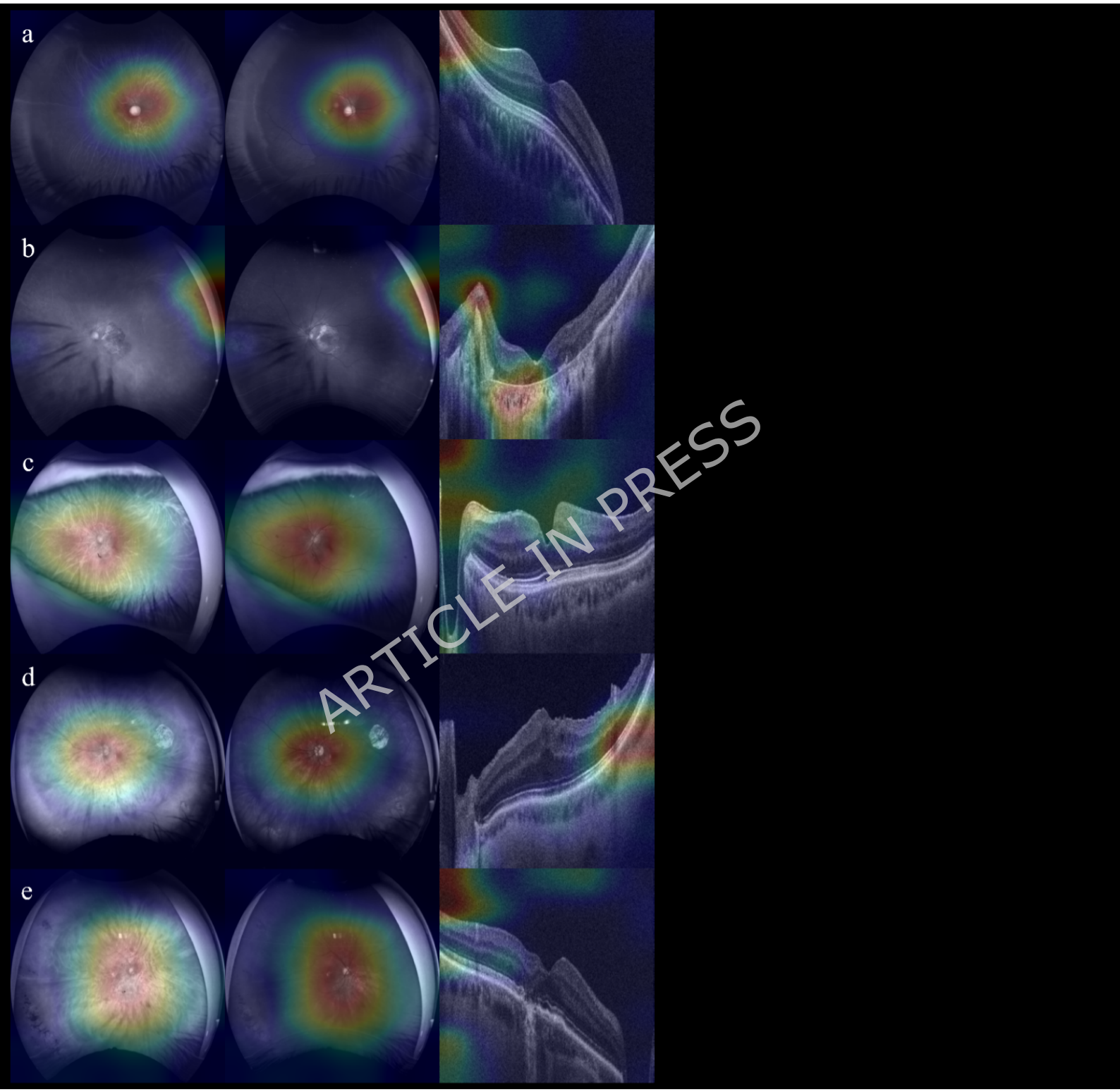
ARTICLE IN PRESS

ARTICLE IN PRESS



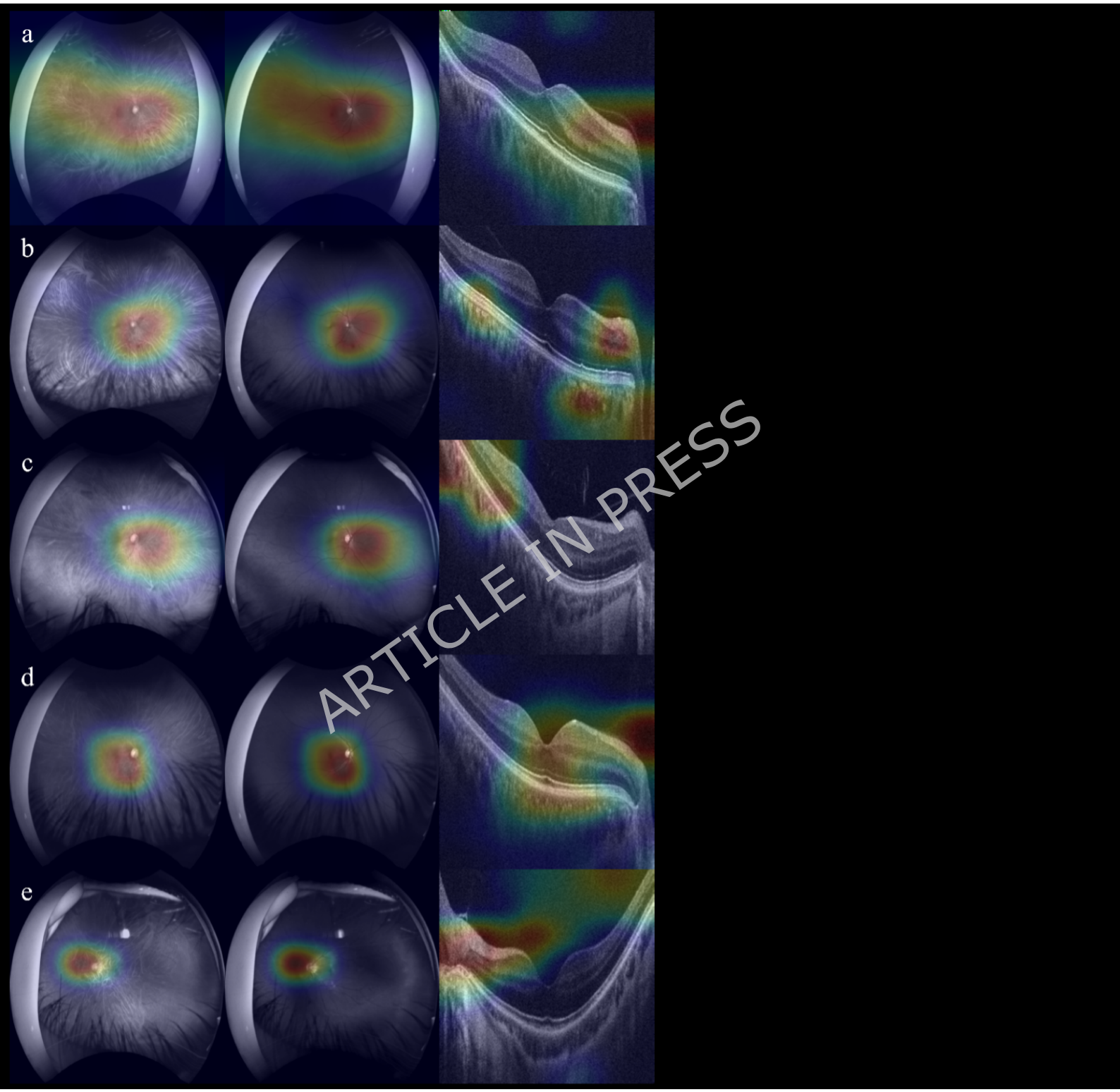
ARTICLE IN PRESS

ARTICLE IN PRESS



ARTICLE IN PRESS

ARTICLE IN PRESS



ARTICLE IN PRESS