Dataset	Device name	Image count	Percentage
AlzEye	Topcon 3D OCT-2000SA	1716282	66.44%
	Topcon Triton plus	496344	19.22%
	Topcon FD-OCT	209132	8.10%
	Topcon Triton	76481	2.96%
	Carl Zeiss Meditec CIRRUS photo 600	29158	1.13%
	Topcon 3D OCT-1000MK2	28988	1.12%
	Topcon 3D OCT-2000	26618	1.03%
	Unknown	9	0.00%
UK Biobank	Topcon 3D OCT-1000	171538	100.00%
BRSET	Canon CR	10591	65.11%
	Nikon NF505	5675	34.89%
mBRSET	Phelcom Eyer	5164	100.00%

Tab. 1. Overview of imaging devices used to collect colour fundus photography (CFP) images in the original AlzEye, UK Biobank, BRSET, and mBRSET datasets.

Age subgroup	Mean RAG [95% CI]		OR [95% CI]	P value
	Healthy group	Unhealthy group		
<40	+4.39 [+4.11, +4.67]	+8.14 [+7.41, +8.87]	1.140 [1.112, 1.169]	<0.001
40-50	+3.03 [+2.84, +3.21]	+6.14 [+5.89, +6.39]	1.109 [1.097, 1.121]	<0.001
50-60	+1.14 [+0.97, +1.32]	+3.62 [+3.49, +3.75]	1.101 [1.092, 1.111]	<0.001
60-70	−0.63 [−0.83, −0.43]	+0.76 [+0.66, +0.86]	1.048 [1.041, 1.056]	<0.001
70-80	−2.32 [−2.55, −2.08]	−1.96 [−2.06, −1.87]	1.011 [1.004, 1.018]	0.001
≥80	−4.15 [−4.55, −3.75]	−5.99 [−6.08, −5.89]	0.950 [0.938, 0.962]	<0.001

Tab. 2. Mean RAG and unhealthy odds ratio per +1 year RAG by chronological-age subgroup. Each age subgroup was analysed separately. For each subgroup, mean RAG (± 95% CI) is shown for the healthy and unhealthy groups; the odds ratio per +1 year of RAG is from a logistic regression of unhealthy status on age and RAG fit within that subgroup (OR = exp(β_RAG), with 95% CI and p from the Wald test).

Cohort	MAE [95% CI]	RMSE [95% CI]	Mean RAG [95% CI]
AlzEye internal test set	5.31 [5.28, 5.35]	6.80 [6.76, 6.84]	−0.86 [−0.91, −0.81]

Tab. 3. Retinal age model performance on the AlzEye internal test set. Mean absolute error (MAE), root mean squared error (RMSE), and mean RAG (all in years) are reported with 95% CIs from an image-level bootstrap (1,000 resamples).

Disease category	Disease count	Healthy count	Odds ratio [95% CI]	P value		
Unhealthy	51,603	13,757	1.049 [1.046, 1.053]	<0.001		
Cardiovascular	13,952	13,757	1.041 [1.036, 1.046]	<0.001		
Chronic	44,556	13,757	1.048 [1.045, 1.052]	<0.001		
Ocular	22,797	13,757	1.065 [1.061, 1.069]	<0.001		

Tab. 4. Association between RAG and disease conditions. Logistic regression was performed with RAG as the independent variable and each disease condition as the outcome, adjusted for chronological age. Reported are odds ratios (OR) per 1-year increase in RAG with 95% confidence intervals; p-values are from the Wald test.

Disease category	β_0 (Intercept)	β_1 (Age)	β_2 (Group)	β_3 (Age × Group)	Healthy RAG-age slope	Disease RAG-age slope
Unhealthy	−1.346 [−1.459, −1.232], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+0.932 [+0.808, +1.057], p<0.001	−0.130 [−0.138, −0.122], p<0.001	−0.184 [−0.191, −0.177]	−0.314 [−0.318, −0.310]
Cardiovascular	−1.346 [−1.459, −1.233], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+1.423 [+1.266, +1.579], p<0.001	−0.161 [−0.172, −0.150], p<0.001	−0.184 [−0.191, −0.177]	−0.345 [−0.353, −0.336]
Chronic	−1.346 [−1.459, −1.232], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+0.884 [+0.759, +1.010], p<0.001	−0.136 [−0.144, −0.128], p<0.001	−0.184 [−0.191, −0.177]	−0.320 [−0.324, −0.316]
Ocular	−1.346 [−1.461, −1.231], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+1.768 [+1.629, +1.907], p<0.001	−0.161 [−0.171, −0.152], p<0.001	−0.184 [−0.191, −0.177]	−0.345 [−0.352, −0.339]
Disease count =1	−1.346 [−1.457, −1.235], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+0.451 [+0.322, +0.581], p<0.001	−0.106 [−0.114, −0.097], p<0.001	−0.184 [−0.191, −0.177]	−0.290 [−0.295, −0.284]
Disease count =2	−1.346 [−1.459, −1.232], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+1.519 [+1.378, +1.659], p<0.001	−0.173 [−0.183, −0.163], p<0.001	−0.184 [−0.191, −0.177]	−0.357 [−0.364, −0.350]
Disease count =3	−1.346 [−1.459, −1.233], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+2.289 [+2.078, +2.501], p<0.001	−0.206 [−0.221, −0.191], p<0.001	−0.184 [−0.191, −0.177]	−0.390 [−0.404, −0.376]

Tab. 5. Linear regression coefficients of the RAG-age relationship by health status, for each disease category. For each disease category exact disease-category count group, RAG was regressed on chronological age, group status (0: healthy, 1: disease category or count group), and their interaction: $RAG = \beta_0 + \beta_1 \cdot Age + \beta_2 \cdot Group + \beta_3 \cdot (Age \times Group)$. Age was mean-centered at 68.38 years (the original age minus the mean age). Disease-category count was defined across four disease categories: cardiovascular, chronic, ocular, and neurodegenerative disease. Results are shown for three exact count groups: disease count = 1, disease count = 2, and disease count = 3. Results are shown for four disease groups: unhealthy, cardiovascular, chronic, and ocular disease. Each coefficient is given with its 95% CI and p-value, together with the implied healthy (β_1) and disease-group ($\beta_1 + \beta_3$) RAG-age slopes.

Condition	β_0 (Intercept)	β_1 (Age)	β_2 (Group)	β_3 (Age × Group)	Healthy RAG-age slope	Unhealthy RAG-age slope
Raw	−1.346 [−1.459, −1.232], p<0.001	−0.184 [−0.191, −0.177], p<0.001	+0.932 [+0.808, +1.057], p<0.001	−0.130 [−0.138, −0.122], p<0.001	−0.184 [−0.191, −0.177]	−0.314 [−0.318, −0.310]
Linear calibration	+0.259 [+0.145, +0.373], p<0.001	+0.017 [+0.011, +0.024], p<0.001	+0.932 [+0.808, +1.057], p<0.001	−0.130 [−0.138, −0.122], p<0.001	+0.017 [+0.011, +0.024]	−0.113 [−0.117, −0.109]
Quadratic calibration	+0.273 [+0.159, +0.387], p<0.001	+0.018 [+0.011, +0.025], p<0.001	+0.953 [+0.828, +1.077], p<0.001	−0.119 [−0.127, −0.111], p<0.001	+0.018 [+0.011, +0.024]	−0.101 [−0.105, −0.097]
Balanced training	−1.420 [−1.530, −1.310], p<0.001	−0.190 [−0.196, −0.183], p<0.001	+1.096 [+0.976, +1.216], p<0.001	−0.154 [−0.162, −0.147], p<0.001	−0.190 [−0.196, −0.183]	−0.344 [−0.348, −0.340]

Tab. 6. Linear regression coefficients of the RAG-age relationship by health status, for different conditions. RAG was regressed on chronological age, health-status group (0: healthy, 1: unhealthy), and their interaction: RAG = $\beta_0 + \beta_1 \cdot$Age + $\beta_2 \cdot$Group + $\beta_3 \cdot$(Age × Group). Age was mean-centered at 68.38 years (the original age minus the mean age). Results are shown for four conditions: raw RAG, linear age-calibrated RAG, quadratic age-calibrated RAG, and RAG from the model retrained on a balanced age distribution. Each coefficient is given with its 95% CI and p-value, together with the implied healthy (β_1) and unhealthy ($\beta_1 + \beta_3$) RAG-age slopes.

Calibrator	Form	Coefficients		
Linear	$\alpha\_l + \beta\_l \cdot Age$	$\alpha\_l = 12.159$	$\beta\_l = -0.201$	
Quadratic	$\alpha\_q + \beta_1 \cdot Age + \beta_2 \cdot Age^2$	$\alpha\_q = 9.899$	$\beta_1 = -0.122$	$\beta_2 = -0.00066$

Tab. 7. Age-calibration coefficients. Two age-calibrators were fit on the healthy subset of the downstream training set by regressing RAG on chronological age: a linear calibrator, $Calibrator(Age) = \alpha\_l + \beta\_l \cdot Age$, and a quadratic calibrator, $Calibrator(Age) = \alpha\_q + \beta_1 \cdot Age + \beta_2 \cdot Age^2$. The calibrated RAG used in the slope analyses was defined as $RAG\_cal = RAG - Calibrator(Age)$. Coefficients are on the original (un-centred) age scale.

Chronological Age	Unhealthy		Cardiovascular		Chronic		Ocular	
	Odds ratio [95% CI]	P value	Odds ratio [95% CI]	P value	Odds ratio [95% CI]	P value	Odds ratio [95% CI]	P value
30	1.116 [1.110, 1.122]	<0.001	1.135 [1.126, 1.144]	<0.001	1.118 [1.112, 1.124]	<0.001	1.168 [1.161, 1.175]	<0.001
40	1.096 [1.092, 1.100]	<0.001	1.111 [1.104, 1.118]	<0.001	1.098 [1.093, 1.102]	<0.001	1.138 [1.133, 1.144]	<0.001
50	1.076 [1.073, 1.079]	<0.001	1.088 [1.083, 1.093]	<0.001	1.077 [1.074, 1.081]	<0.001	1.110 [1.106, 1.114]	<0.001
60	1.057 [1.055, 1.059]	<0.001	1.065 [1.061, 1.068]	<0.001	1.058 [1.055, 1.060]	<0.001	1.082 [1.079, 1.085]	<0.001
70	1.038 [1.035, 1.040]	<0.001	1.042 [1.039, 1.045]	<0.001	1.038 [1.036, 1.041]	<0.001	1.054 [1.052, 1.057]	<0.001
80	1.020 [1.016, 1.023]	<0.001	1.020 [1.016, 1.024]	<0.001	1.019 [1.016, 1.022]	<0.001	1.028 [1.024, 1.031]	<0.001
90	1.001 [0.996, 1.005]	0.687	0.998 [0.993, 1.003]	0.513	1.000 [0.996, 1.005]	0.931	1.002 [0.997, 1.007]	0.455
100	0.983 [0.978, 0.989]	<0.001	0.977 [0.970, 0.984]	<0.001	0.982 [0.976, 0.988]	<0.001	0.977 [0.970, 0.983]	<0.001

Tab. 8. Odds ratio per +1 year of RAG for different disease categories, by chronological age. From the Age + RAG + Age × RAG logistic regression fit on AlzEye dataset, the odds ratio in RAG at a given age is OR = exp(β_RAG + β_Age×RAG · age). Values are shown at 10-year intervals (30-100 years) with 95% CIs from the model covariance.

Disease	Model	Overall	<50	50-80	>80
Unhealthy	Age	0.731 [0.726, 0.736]	0.580 [0.567, 0.594]	0.622 [0.616, 0.629]	0.585 [0.566, 0.603]
	Age + RAG	0.737 [0.732, 0.742]	0.674 [0.662, 0.686]	0.641 [0.634, 0.648]	0.494 [0.475, 0.512]
	Age + RAG + Age × RAG	0.741 [0.736, 0.746]	0.685 [0.672, 0.697]	0.641 [0.635, 0.649]	0.564 [0.546, 0.582]
Cardiovascular	Age	0.806 [0.800, 0.810]	0.638 [0.611, 0.664]	0.676 [0.668, 0.684]	0.620 [0.602, 0.638]
	Age + RAG	0.809 [0.804, 0.814]	0.697 [0.667, 0.727]	0.692 [0.684, 0.700]	0.556 [0.539, 0.575]
	Age + RAG + Age × RAG	0.813 [0.808, 0.818]	0.689 [0.655, 0.721]	0.695 [0.687, 0.703]	0.606 [0.588, 0.624]
Chronic	Age	0.727 [0.722, 0.732]	0.586 [0.572, 0.600]	0.614 [0.607, 0.621]	0.579 [0.560, 0.595]
	Age + RAG	0.733 [0.728, 0.738]	0.682 [0.669, 0.695]	0.633 [0.627, 0.640]	0.485 [0.467, 0.505]
	Age + RAG + Age × RAG	0.737 [0.732, 0.742]	0.692 [0.679, 0.706]	0.634 [0.628, 0.641]	0.558 [0.539, 0.576]
Ocular	Age	0.758 [0.752, 0.763]	0.537 [0.517, 0.559]	0.656 [0.649, 0.664]	0.588 [0.570, 0.608]
	Age + RAG	0.769 [0.764, 0.775]	0.723 [0.705, 0.742]	0.684 [0.677, 0.691]	0.508 [0.489, 0.529]
	Age + RAG + Age × RAG	0.774 [0.769, 0.780]	0.760 [0.742, 0.779]	0.685 [0.678, 0.692]	0.571 [0.554, 0.590]

Tab. 9. AUC of disease classification models within different age subgroups. For each disease (overall unhealthy, cardiovascular, chronic, and ocular disease), three logistic models (Age; Age + RAG; and Age + RAG + Age × RAG) were fit on the downstream training set and evaluated on the internal test set. AUC is reported overall and within three age subgroups (<50, 50-80, and >80 years), each with a 95% CI from a bootstrap with 1,000 resamples.

Cohort	MAE [95% CI]	RMSE [95% CI]	Mean RAG [95% CI]
UKB	3.80 [3.76, 3.83]	4.87 [4.83, 4.92]	+1.48 [+1.43, +1.54]
BRSET	7.34 [7.17, 7.50]	9.37 [9.15, 9.56]	−0.92 [−1.19, −0.68]
mBRSET	8.40 [8.17, 8.62]	10.52 [10.19, 10.85]	−5.70 [−6.03, −5.37]

Tab. 10. Retinal age model performance on three external cohorts (UK Biobank, BRSET, mBRSET). Mean absolute error (MAE), root mean squared error (RMSE), and mean RAG (all in years) are reported with 95% CIs from an image-level bootstrap (1,000 resamples). The model was trained on AlzEye and applied off-the-shelf on external data.

Chronological age	UKB		BRSET		mBRSET	
	OR [95% CI]	P value	OR [95% CI]	P value	OR [95% CI]	P value
30	—	—	1.052 [1.039, 1.066]	<0.001	1.008 [0.984, 1.032]	0.541
40	1.026 [1.012, 1.041]	<0.001	1.048 [1.036, 1.060]	<0.001	1.017 [0.998, 1.037]	0.083
50	1.027 [1.019, 1.035]	<0.001	1.043 [1.032, 1.054]	<0.001	1.027 [1.011, 1.044]	<0.001
60	1.027 [1.021, 1.034]	<0.001	1.038 [1.027, 1.049]	<0.001	1.037 [1.021, 1.053]	<0.001
70	1.028 [1.016, 1.039]	<0.001	1.034 [1.022, 1.045]	<0.001	1.048 [1.029, 1.066]	<0.001
80	—	—	1.029 [1.016, 1.042]	<0.001	1.058 [1.035, 1.081]	<0.001
90	—	—	1.024 [1.009, 1.039]	<0.001	—	—
100	—	—	—	—	—	—

Tab. 11. Odds ratio per +1 year of RAG for unhealthy conditions, by chronological age, in three external cohorts. For each cohort, from Age + RAG + Age × RAG logistic model fit on that cohort's training split, the odds ratio per one-year increase in RAG at a given age is OR = exp(β_RAG + β_Age×RAG · Age), with 95% CI from the model covariance and p from the Wald test. Values are shown at 10-year intervals; ages outside a cohort's observed range are left blank.

Cohort	Evaluation	Model	AUC [95% CI]	P value (ΔAUC)	F1 [95% CI]	P value (ΔF1)	Sensitivity [95% CI]	Specificity [95% CI]
AlzEye	Off-the-shelf	Age	0.731 [0.706, 0.756]	<0.001	0.807 [0.740, 0.830]	<0.001	0.751 [0.632, 0.805]	0.588 [0.528, 0.708]
		Age + RAG + Age × RAG	0.741 [0.715, 0.767]		0.831 [0.806, 0.842]		0.791 [0.744, 0.810]	0.573 [0.530, 0.639]
UKB	Off-the-shelf	Age	0.684 [0.674, 0.693]	0.812	0.479 [0.466, 0.490]	<0.001	0.413 [0.401, 0.426]	0.813 [0.805, 0.821]
		Age + RAG + Age × RAG	0.685 [0.675, 0.694]		0.521 [0.510, 0.532]		0.494 [0.482, 0.507]	0.760 [0.751, 0.768]
	Externally calibrated	Age	0.684 [0.674, 0.693]	<0.001	0.576 [0.562, 0.595]	<0.001	0.661 [0.616, 0.761]	0.620 [0.519, 0.665]
		Age + RAG + Age × RAG	0.688 [0.679, 0.697]		0.589 [0.566, 0.599]		0.717 [0.610, 0.736]	0.572 [0.552, 0.678]
BRSET	Off-the-shelf	Age	0.662 [0.644, 0.680]	0.508	0.594 [0.572, 0.614]	0.002	0.558 [0.532, 0.585]	0.670 [0.648, 0.692]
		Age + RAG + Age × RAG	0.664 [0.646, 0.682]		0.574 [0.553, 0.596]		0.510 [0.484, 0.536]	0.725 [0.703, 0.745]
	Externally calibrated	Age	0.662 [0.644, 0.680]	0.012	0.626 [0.576, 0.650]	0.002	0.633 [0.525, 0.688]	0.600 [0.544, 0.710]
		Age + RAG + Age × RAG	0.671 [0.652, 0.689]		0.607 [0.577, 0.637]		0.570 [0.514, 0.621]	0.683 [0.633, 0.742]
mBRSET	Off-the-shelf	Age	0.637 [0.582, 0.696]	0.15	0.659 [0.622, 0.697]	<0.001	0.535 [0.491, 0.581]	0.646 [0.560, 0.734]
		Age + RAG + Age × RAG	0.614 [0.559, 0.665]		0.536 [0.496, 0.575]		0.390 [0.353, 0.429]	0.736 [0.668, 0.805]
	Externally calibrated	Age	0.637 [0.582, 0.696]	0.938	0.790 [0.664, 0.851]	<0.001	0.732 [0.535, 0.850]	0.512 [0.381, 0.717]
		Age + RAG + Age × RAG	0.638 [0.583, 0.695]		0.836 [0.694, 0.858]		0.822 [0.583, 0.851]	0.419 [0.358, 0.654]

Tab. 12. Disease (unhealthy) classification performance across cohorts. AUC, sensitivity, specificity, and F1 (each with a 95% CI from a patient-level cluster bootstrap) are reported for two logistic models (Age-only; Age + RAG + Age × RAG) on the AlzEye internal test set and three external cohorts' test splits. Within each cohort and evaluation setting, p values indicate the paired difference between the two models for AUC or F1. Sensitivity, specificity and F1 are evaluated at an operating threshold selected by the Youden criterion on the relevant training set. For external cohorts, two evaluations are shown: off-the-shelf, applying the AlzEye-fit logistic and its Youden threshold (both derived on the AlzEye downstream training set) directly, and externally calibrated, refitting the logistic and re-deriving the Youden threshold on each cohort's own training split.

Disease category	Disease	AlzEye (ICD-10)	UK Biobank	
			ICD-10	Self-report
Cardiovascular	Atrial fibrillation	I48	I48	1471
	Heart failure	I110, I130, I132, I255, I420, I425, I428, I429, I500, I501, I509	I110, I130, I132, I255, I420, I425, I428, I429, I500, I501, I509	1076
	Myocardial infarction	I21, I22, I23, I241, I252	I21, I22, I23, I241, I252	1075
	Intracerebral haemorrhage	I61	I61	1491
	Ischaemic stroke	I63, I64	I63, I64	1583
	Aortic stenosis	I350	I350	1490
	Aortic regurgitation	I351	I351	1587
	Mitral stenosis	I342, I050	I342, I050	1489
	Mitral regurgitation	I340	I340	1585
Chronic systemic	Diabetes mellitus	E08, E09, E10, E11, E13	E08, E09, E10, E11, E13	1220, 1222, 1223
	Hypertension	I10, I11, I12, I13, I15	I10, I12, I15	1065, 1072
Ocular	Glaucoma	H40, H41, H42	H40	1277
	Diabetic retinopathy	E103, E113, H360	H360	1276
	Age-related macular degeneration	H353	—	—
	Cataract	—	H25, H26	1278
Neurodegenerative	Dementia	F01, F02, F03	F01, F02, F03	1263
	Parkinson's disease	G20, G21	G20, G21	1262

Tab. 13. Disease definitions for AlzEye and UK Biobank. For both cohorts, disease labels were derived from linked hospital records using International Classification of Diseases, 10th Revision (ICD-10) codes, supplemented in UK Biobank by self-reported codes (Data-Field 20002).