

## Supplemental Material For - Artificial Intelligence (AI) to Stratify Severity of Age-related Macular Degeneration (AMD) and Predict Risk of Progression to Late AMD

The tables are in the order they appear in the main paper.

**eTable 1: Number of images in training and validation sets for building deep learning based 12-stage AMD severity level classifier, as defined by AREDS study.**

AMD Severity Scale Level	Training set	Validation Set
1	27521	10674
2	7743	3135
3	3309	1358
4	6230	2565
5	3957	1751
6	5165	2373
7	4765	1948
8	4038	1672
9	1466	618
10	2349	697
11	9866	1897
12	752	123

**eTable 2: Confusion matrix for test data with each of the four AMD classes (No AMD, Early, Intermediate and Advanced) along with precision and accuracy reported.**

Confusion Matrix	No AMD	Early	Intermediate	Advanced	Accuracy
No AMD	2762	132	1	0	95.41%
Early	135	2114	26	0	92.62%
Intermediate	0	47	2983	2	98.38%
Advanced	0	0	10	1263	99.21%
Precision	95.34%	92.19%	98.74%	99.84%	

**eTable 3: Performance of the AMD screening system model with accuracy, kappa, sensitivity, specificity and AUC reported for four-class classification system (No AMD, Early, Intermediate and Advanced) along with accuracy, sensitivity, specificity and AUC for 2-class (No AMD and Early AMD as one class, and Intermediate and Advanced as the other class) AMD screening system.**

Metric	Value (95% CI)
Overall Accuracy (4-class)	96.1% (n/a)
Kappa (Unweighted) (4-class)	94.9% (n/a)
Kappa (Quadratic Weighted) (4-class)	98.3% (n/a)
Two-class Accuracy (classes 1 or 2 and 3 or 4)	99.2% (99.02% to 99.39%)
Two-class Sensitivity (classes 3 or 4 as positive)	98.9% (98.64% to 99.66%)
Two-class Specificity (classes 1 or 2 as negative)	99.5% (98.85% to 99.80%)
AUC (two-class)	99.0% (98.9% to 99.15%)

**eTable 4: Confusion matrix for the test data on 1-year risk of late AMD prediction system with converted (incident) cases as positive and non-converted cases as negative.**

AMD 1-year Prediction	Predicted Positive (Converted)	Predicted Negative (Non-Converted)
True Positive (Converted)	245	25
True Negative (Non-converted)	130	722

**eTable 5: Confusion matrix for the test data on 2-year risk of late AMD prediction system with converted (incident) cases as positive and non-converted cases as negative.**

AMD 2-year Prediction	Predicted Positive (Converted)	Predicted Negative (Non-Converted)
True Positive (Converted)	256	21
True Negative (Non-converted)	133	719

**eTable 6: Confusion matrix for test data on 2-year dry AMD prediction system with converted (incident) cases as positive and non-converted cases as negative.**

Dry AMD <b>2-year</b> Prediction	Predicted Positive (dry-AMD)	Predicted Negative
True Positive/dry-AMD	74	33
True Negative	331	661

**eTable 7: Confusion matrix for test data on 2-year wet AMD prediction system with converted (incident) cases as positive and non-converted cases as negative.**

Wet AMD <b>2-year</b> Prediction	Predicted Positive (wet-AMD)	Predicted Negative
True Positive (wet-AMD)	100	40
True Negative	321	638

**eTable 8: Confusion matrix for test data on 1-year dry AMD prediction system with converted (incident) cases as positive and non-converted cases as negative.**

Dry AMD <b>1-year</b> Prediction	Predicted Positive (dry-AMD)	Predicted Negative
True Positive/dry-AMD	75	32
True Negative	333	659

**eTable 9: Confusion matrix for test data on 1-year wet AMD prediction system with converted (incident) cases as positive and non-converted cases as negative.**

Wet AMD <b>1-year</b> Prediction	Predicted Positive (wet-AMD)	Predicted Negative
True Positive (wet-AMD)	103	37
True Negative	313	646

**eTable 10: Performance comparison on input data/parameter types for 2-year incident of late AMD system with converted (incident) cases as positive and non-converted cases as negative.**

	Sensitivity	Specificity
AMD Prediction with retinal and socio-demographic parameters	92.42% (88.64% to 95.25%)	84.39% (81.78% to 86.76%)
AMD Prediction with only retinal image parameters	88.79% (85.84% to 90.62%)	84.71% (80.49% to 88.32%)

**eTable 11: The confusion matrix for incident of late AMD (either dry or wet i.e., any) in NAT-2 dataset.**

<b>NAT-2 Validation</b>	Predicted Positive	Predicted Negative	Total Actual
Actual Positive	27	3	30
Actual Negative	11	47	58
Total Predicted	38	50	-

**eTable 12. Number of subjects in NAT-2 dataset used for dry and wet AMD prediction system validation.**

Total subjects in validation set (in placebo group)	88
Normal (non-convert)	58
Incident Wet AMD cases	26
Incident Dry AMD cases	6
Incident Wet and Dry AMD cases	2
Incident Wet or Dry AMD cases	30 (26+6-2)

**eTable 13: Confusion matrix for incident late wet AMD in NAT-2 dataset.**

<b>Incident Wet AMD</b>	<b>Predicted wet AMD</b>	<b>Predicted Negative</b>
Actual wet AMD	19	7
Actual Negative	24	38

**eTable 14: Baseline Characteristics and on baseline data grouped by incident AMD status (final AMD status) on AREDS dataset.**

<b>Characteristics (Means or Prevalences)</b>	<b>Late AMD (Case)</b>	<b>No Late AMD (Control)</b>	<b>P Value</b>
Age, years $\pm$ SD	70.9 $\pm$ 5	68.6 $\pm$ 5	0.01
Male, n (%)	597 (46)	1267 (44)	0.1
Race-White, n (%)	1276 (99)	2686 (94)	0.02
SBP, mmHg $\pm$ SD	140 $\pm$ 19	136 $\pm$ 18	0.9
DBP, mmHg $\pm$ SD	79 $\pm$ 10	79 $\pm$ 9.6	0.03
Smoker, n (%)	833 (65)	1508 (53)	0.06
Diabetes, n (%)	116 (9)	250 (9)	0.6
Hypertension, n (%)	547 (42)	1079 (38)	0.1
AMD Category			<0.0001
No AMD, n (%)	5 (0.3)	975 (34)	
Early AMD, n (%)	35 (2.7)	866 (30)	
Intermediate AMD, n (%)	467 (36.2)	931 (33)	
Advanced AMD, n (%)	781 (60.6)	79 (3)	
Body mass index, kg/m <sup>2</sup> $\pm$ SD	28 $\pm$ 5	27 $\pm$ 5	0.03
Treatment Category-Placebo, n (%)	315 (24)	961 (34)	0.2

eTable 15: Performance evaluation of exploratory model building based on two classifiers predicting risk of dry and wet AMD individually based on images only without any specific retinal features being added as input parameters.

Measure	1-year		2-year	
	Wet AMD	Dry AMD	Wet AMD	Dry AMD
<b>Sensitivity</b>	58.97% (51.72% to 65.95%)	62.84% (54.52% to 70.63%)	60.00% (52.76% to 66.93%)	64.86% (56.60% to 72.52%)
<b>Specificity</b>	68.07% (63.78% to 72.15%)	68.99% (64.92% to 72.85%)	70.08% (65.85% to 74.07%)	68.99% (64.92% to 72.85%)
<b>Accuracy</b>	65.51% (61.84% to 69.05%)	67.68% (64.05% to 71.15%)	67.24% (63.61% to 70.73%)	68.11% (64.50% to 71.57%)