

Supplementary table 1 Different models performance in different cohorts

Cohort	Model	AUC	AUC 95% CI Lower	AUC 95% CI Upper	Accuracy	F1	Sensitivity	Specificity	PPV	NPV
Internal training	Combined Model	0.981085	0.961628	1	0.919811	0.912821	0.917526	0.921739	0.908163	0.929825
Internal training	Logical Score	0.592918	0.515985	0.669851	0.599057	0.484848	0.412371	0.756522	0.588235	0.604167
Internal training	Tumor Score	0.517078	0.438875	0.59528	0.490566	0.5	0.556701	0.434783	0.453782	0.537634
Internal training	Tenon Score	0.648947	0.574414	0.72348	0.580189	0.597285	0.680412	0.495652	0.532258	0.647727
Internal validation	Combined Model	0.974747	0.937844	1	0.901099	0.873239	0.861111	0.927273	0.885714	0.910714
Internal validation	Logical Score	0.692424	0.578654	0.806195	0.681319	0.508475	0.416667	0.854545	0.652174	0.691176
Internal validation	Tumor Score	0.582828	0.461537	0.704119	0.538462	0.543478	0.694444	0.436364	0.446429	0.685714
Internal validation	Tenon Score	0.69899	0.585929	0.812051	0.626374	0.613636	0.75	0.545455	0.519231	0.769231
External validation	Combined Model	0.987205	0.962558	1	0.931373	0.917647	0.95122	0.918033	0.886364	0.965517
External validation	Logical Score	0.52579	0.410773	0.640807	0.617647	0.4	0.317073	0.819672	0.541667	0.641026
External validation	Tumor Score	0.583766	0.469685	0.697848	0.558824	0.536082	0.634146	0.508197	0.464286	0.673913
External validation	Tenon Score	0.70032	0.594198	0.806441	0.637255	0.602151	0.682927	0.606557	0.538462	0.74

2 AUC: area under curve. CI: Confidence Interval. PPV: Positive predict value. NPV: Negative predict value.

Supplementary table 2 Calibration performance in different cohort

Cohort	Model	Brier score	HL p-value	Calibration intercept	Calibration slope
Internal training	Combined model	0.053159237	0.228280305	0.004947035	1.002457344
Internal training	Logical Score	0.241958313	0.001521464	0.214546564	1.442510534
Internal training	Tumor Score	0.247919493	0.000000478	0.462016401	1.392027217
Internal training	Tenon Score	0.228487887	0.049813624	-0.256470011	1.530215185
Internal validation	Combined model	0.061141014	0.669602028	0.001334471	1.001444851
Internal validation	Logical Score	0.222182631	0.000456993	0.245843947	2.608065132
Internal validation	Tumor Score	0.240967821	0.002414427	0.335548805	4.739660185
Internal validation	Tenon Score	0.210623272	0.014237474	-0.206374517	1.287954034
External validation	Combined model	0.045184209	0.650432209	-0.000959001	1.007731584
External validation	Logical Score	0.242671229	0.004756305	-0.29574794	0.462503937
External validation	Tumor Score	0.241107247	0.053050407	0.456403858	5.0121627
External validation	Tenon Score	0.210868022	0.003548492	-0.346218392	1.40391525

Supplementary table 3 DCA performance in different cohort(Combined model as reference)

Cohort	Compare (Reference=Combined model)	NRI	NRI 95%CI Lower	NRI 95%CI Upper	IDI	IDI 95%CI Lower	IDI 95%CI Upper
Internal training	Logical Score	- 1.68278	-1.81316	-1.53303	- 0.66039	-0.71436	-0.60555
Internal training	Tumor Score	-1.6855	-1.81597	-1.54318	- 0.70554	-0.7551	-0.64972

Internal training	Tenon Score	- 1.56027	-1.72252	-1.38813	- 0.59383	-0.64842	-0.53359
Internal validation	Logical Score	- 1.57551	-1.83127	-1.29584	- 0.61725	-0.71357	-0.5079
Internal validation	Tumor Score	- 1.57879	-1.83591	-1.32652	- 0.70796	-0.79213	-0.61196
Internal validation	Tenon Score	- 1.31982	-1.61131	-0.99283	- 0.55665	-0.66352	-0.44115
External validation	Logical Score	- 1.70251	-1.88549	-1.49006	- 0.68879	-0.76526	-0.6051
External validation	Tumor Score	- 1.66887	-1.85351	-1.44419	- 0.71431	-0.78243	-0.63532
External validation	Tenon Score	- 1.50379	-1.72004	-1.26139	- 0.59646	-0.66796	-0.51208

---

NRI: Net Reclassification Improvement. IDI: Integrated Discrimination Improvement. CI: Confidence Interval