

Table S12. Summary statistics for the severity prediction performance (Severity based grading) of the deep learning system on slit-lamp images (CO).

	AUC	Sensitivity	Specificity	Accuracy
Normal vs Others	0.9690 (0.9596 – 0.9784)	94.37% (91.01%, 97.74%)	93.53%, (90.88%, 96.18%)	93.86%, (91.69%, 96.02%)
Mild vs Others	0.8672 (0.8420 – 0.8924)	53.01% (39.50%, 66.52%)	98.07%, (96.84%, 99.29%)	93.07%, (90.78%, 95.36%)
Moderate vs Others	0.9230 (0.9093 – 0.9367)	88.84% (84.03%, 93.65%)	92.61%, (89.12%, 96.10%)	91.26%, (88.50%, 94.02%)
Severe vs Others	0.9927 (0.9889 – 0.9965)	98.20% (95.53%, 100.0%)	98.29%, (96.96%, 99.63%)	98.29%, (97.04%, 99.54%)
Overall	0.9395 (0.9323 – 0.9467)	88.06% (84.76%, 91.37%)	94.26%, (92.74%, 95.77%)	93.30%, (91.51%, 95.10%)
Overall (ResNet18)	0.9186 (0.9100 – 0.9272)	87.10% (84.02%, 90.19%)	92.93% (91.34%, 94.53%)	92.53% (90.59%, 94.47%)
Overall (WideResNet50-2)	0.9294 (0.9187 – 0.9400)	87.38% (84.4%, 90.36%)	93.55% (91.86%, 95.24%)	92.53% (90.62%, 94.43%)
Overall (ResNext50)	0.9246 (0.9160 – 0.9333)	83.52% (79.99%, 87.04%)	93.15% (91.21%, 95.09%)	90.55% (88.41%, 92.68%)

Asymptotic two-sided 95% CIs, adjusted for clustering by patients, were calculated and presented as proportions (sensitivity, specificity, accuracy) and the area under the curve (AUC).

Abbreviations: CO = cortical opacity; CI = confidence interval; AUC = area under the receiver operating characteristic curve