

Table S10. Summary statistics for the severity prediction performance (Severity based grading) of the deep learning system on slit-lamp images (NO).

	AUC	Sensitivity	Specificity	Accuracy
Normal vs Others	0.9990 (0.9982 – 0.9999)	96.37% (91.45%, 100.0%)	99.55% (98.84%, 100.0%)	98.90% (97.93%, 99.86%)
Mild vs Others	0.9727 (0.9634 – 0.9820)	90.97% (86.35%, 95.59%)	94.53% (91.66%, 97.41%)	93.23% (90.69%, 95.78%)
Moderate vs Others	0.9478 (0.9314 – 0.9642)	88.35% (83.3%, 93.4%)	92.02% (88.35%, 95.70%)	90.53% (87.32%, 93.74%)
Severe vs Others	0.9860 (0.9794 – 0.9926)	96.16% (92.55%, 99.76%)	98.01% (96.63%, 99.40%)	97.87% (96.52%, 99.22%)
Overall	0.9789 (0.9733 – 0.9845)	91.10% (88.18%, 94.02%)	94.69% (92.88%, 96.5%)	93.51% (91.39%, 95.63%)
Overall (ResNet18)	0.9587 (0.9516 – 0.9657)	88.38% (85.53%, 91.23%)	92.73% (91.0%, 94.45%)	91.29% (89.26%, 93.33%)
Overall (WideResNet50-2)	0.9749 (0.9686 – 0.9813)	90.22% (87.76%, 92.69%)	94.24% (92.68%, 95.81%)	92.85% (91.08%, 94.63%)
Overall (ResNext50)	0.9698 (0.9646 – 0.9751)	89.2% (85.79%, 92.6%)	93.67% (91.71%, 95.63%)	92.34% (90.04%, 94.64%)

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

Abbreviations: NO = nuclear opalescence; CI = confidence interval; AUC = area under the receiver operating characteristic curve