Table S11. Summary statistics for the severity prediction performance (Severity based grading) of the deep learning system on slit-lamp images (NC).

	AUC	Sensitivity	Specificity	Accuracy
Normal vs Others	0.9990 (0.9963 – 1.0000)	95.57% (90.65%, 100.0%)	99.55% (98.85%, 100.0%)	98.88% (97.85%, 99.91%)
Mild vs Others	0.9777 (0.9500 – 1.0000)	89.2% (84.62%, 93.78%)	91.06% (87.43%, 94.69%)	90.35% (87.54%, 93.16%)
Moderate vs Others	0.9579 (0.9178 – 0.9981)	85.22% (79.87%, 90.56%)	92.16% (89.25%, 95.06%)	89.57% (86.62%, 92.52%)
Severe vs Others	0.9903 (0.9755 – 1.0000)	84.73% (72.14%, 97.32%)	95.16% (93.08%, 97.24%)	94.36% (92.24%, 96.49%)
Overall	0.9842 (0.9711 – 0.9974)	88.44% (85.33%, 91.55%)	93.21% (91.38%, 95.03%)	91.70% (89.63%, 93.77%)
Overall (ResNet18)	0.9744 (0.9604 – 0.9885)	87.22% (83.98%, 90.45%)	91.54% (89.28%, 93.8%)	90.39% (88.09%, 92.69%)
Overall (WideResNet50-2)	0.9764 (0.9532 – 0.9996)	87.41% (84.43%, 90.38%)	91.96% (90.03%, 93.89%)	90.43% (88.32%, 92.54%)
Overall (ResNext50)	0.9739 (0.9552 – 0.9926)	85.73% (82.21%, 89.25%)	91.51% (89.34%, 93.68%)	89.86% (87.49%, 92.23%)

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

Abbreviations: NC = nuclear color; CI = confidence interval; AUC = area under the receiver operating characteristic curve