Table S13. Summary statistics for the severity prediction performance (Severity based grading) of the deep learning system on slit-lamp images (PSC).

	AUC	Sensitivity	Specificity	Accuracy
Normal vs Others	0.9362 (0.9218 – 0.9506)	83.13% (75.25%, 91.01%)	88.76% (86.31%, 91.21%)	86.64% (83.28%, 90.01%)
Mild vs Others	0.8878 (0.8581 – 0.9175)	78.74% (68.49%, 88.98%)	98.83% (98.04%, 99.63%)	95.43% (93.42%, 97.44%)
Moderate vs Others	0.7569 (0.7081 – 0.8057)	74.86% (69.52%, 80.19%)	92.58% (88.07%, 97.10%)	86.43% (82.76%, 90.09%)
Severe vs Others	0.9653 (0.9477- 0.9829)	100.0% (100.0%, 100.0%)	93.45% (89.9%, 96.99%)	94.14% (90.98%, 97.31%)
Overall	0.9102 (0.8935 – 0.9269)	81.32% (77.22%, 85.42%)	92.30% (90.4%, 94.20%)	88.82% (86.33%, 91.32%)
Overall (ResNet18)	0.8861 (0.8724 – 0.8997)	79.17% (75.17%, 83.18%)	92.21% (90.32%, 94.11%)	87.73% (85.17%, 90.29%)
Overall (WideResNet50-2)	0.8908 (0.8765 – 0.9052)	80.64% (75.8%, 85.49%)	92.18% (90.26%, 94.1%)	88.33% (85.36%, 91.29%)
Overall (ResNext50)	0.9007 (0.8858 – 0.9155)	79.23% (75.34%, 83.13%)	91.77% (89.94%, 93.61%)	87.77% (85.33%, 90.21%)

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

Abbreviations: PSC = posterior subcapsular opacity; CI = confidence interval; AUC = area under the receiver operating characteristic curve