

**Table S7. Summary statistics for the grading prediction performance (LOCS- III based grading) of the deep learning system on slit-lamp images (NC).**

	AUC	Sensitivity	Specificity	Accuracy
<b>Grade 0 vs Others</b>	0.9981 (0.9958 – 1.0000)	94.84% (88.53%, 100.0%)	99.61% (98.75%, 100.0%)	98.36% (97.12%, 99.6%)
<b>Grade 1 vs Others</b>	0.9711 (0.9337 – 1.0000)	72.15% (59.33%, 84.96%)	94.42% (91.94%, 96.72%)	91.51% (88.44%, 94.58%)
<b>Grade 2 vs Others</b>	0.9436 (0.8906 – 0.9966)	77.16% (69.41%, 84.91%)	88.60% (83.94%, 92.99%)	85.79% (81.72%, 89.86%)
<b>Grade 3 vs Others</b>	0.9162 (0.8624 – 0.9700)	66.22% (55.43%, 77.00%)	92.71% (90.43%, 95.82%)	86.02% (83.47%, 88.58%)
<b>Grade 4 vs Others</b>	0.9251 (0.8463- 1.0000)	71.33% (56.96%, 85.71%)	95.59% (93.65%, 97.10%)	92.65% (90.45%, 94.85%)
<b>Grade 5 vs Others</b>	0.9560 (0.8628 – 1.0000)	78.65% (65.00%, 92.30%)	96.33% (93.86%, 98.53%)	95.29% (92.95%, 97.62%)
<b>Grade 6 vs Others</b>	0.9920 (0.9748 – 1.0000)	72.04% (54.07%, 90.02%)	99.31% (98.40%, 99.80%)	98.36% (97.21%, 99.51%)
<b>Overall</b>	0.9650 (0.9509 – 0.9792)	75.69% (71.94%, 79.44%)	93.71% (92.62%, 94.80%)	90.26% (88.64%, 91.88%)
<b>Overall (ResNet18)</b>	0.9541 (0.9346 – 0.9737)	72.89% (67.22%, 78.56%)	92.44% (90.58%, 94.3%)	89.11% (86.72%, 91.51%)
<b>Overall (WideResNet50-2)</b>	0.9379 (0.9101 – 0.9656)	71.46% (66.66%, 76.26%)	92.94% (91.38%, 94.49%)	88.42% (86.23%, 90.61%)
<b>Overall (ResNext50)</b>	0.9585 (0.9424 – 0.9748)	75.34% (69.8%, 80.88%)	92.97% (91.09%, 94.85%)	89.68% (87.05%, 92.3%)

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

Abbreviations: LOCS-III = Lens Opacities Classification System III; NC = nuclear color; CI = confidence interval; AUC = area under the receiver operating characteristic curve