

**Table S6. Summary statistics for the grading prediction performance (LOCS- III based grading) of the deep learning system on slit-lamp images (NO).**

	AUC	Sensitivity	Specificity	Accuracy
<b>Grade 0 vs Others</b>	0.9985 (0.9972 - 0.9998)	95.16% (90.07%, 100.0%)	99.27% (98.53%, 100.0%)	98.58% (97.60%, 99.55%)
<b>Grade 1 vs Others</b>	0.9482 (0.9314 - 0.9650)	73.37% (62.79%, 83.95%)	98.17% (96.78%, 99.57%)	95.24% (93.17%, 97.30%)
<b>Grade 2 vs Others</b>	0.9421 (0.9296 – 0.9547)	84.26% (77.69%, 90.83%)	87.64% (83.14%, 92.15%)	86.72% (83.72%, 90.18%)
<b>Grade 3 vs Others</b>	0.9340 (0.9162 – 0.9518)	75.48% (66.37%, 84.58%)	89.18% (86.70%, 91.66%)	85.52% (82.44%, 88.61%)
<b>Grade 4 vs Others</b>	0.9166 (0.8999 – 0.9333)	73.02% (63.99%, 82.04%)	97.76% (96.41%, 99.12%)	95.11% (93.61%, 96.61%)
<b>Grade 5 vs Others</b>	0.9617 (0.9489 – 0.9745)	73.50% (55.03%, 91.98%)	96.86% (95.06%, 98.66%)	95.74% (94.08%, 97.39%)
<b>Grade 6 vs Others</b>	0.9935 (0.9891 – 1.0000)	95.40% (86.94%, 100.0%)	98.55% (86.94%, 100.0%)	98.45% (97.32%, 99.58%)
<b>Overall</b>	0.9567 (0.9501 – 0.9633)	81.71% (78.49%, 84.93%)	93.10% (91.63%, 94.57%)	91.22% (89.40%, 93.04%)
<b>Overall (ResNet18)</b>	0.9466 (0.9415 – 0.9517)	79.12% (75.51%, 82.72%)	92.35% (91.08%, 93.63%)	90.79% (89.16%, 92.42%)
<b>Overall (WideResNet50-2)</b>	0.9471 (0.9402 – 0.9541)	80.32% (76.68%, 83.95%)	92.49% (91.11%, 93.88%)	90.37% (88.61%, 92.12%)
<b>Overall (ResNext50)</b>	0.9463 (0.9412 – 0.9514)	78.96% (75.07%, 82.85%)	92.75% (91.59%, 93.90%)	90.76% (89.14%, 92.38%)

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