

**Table S8. Summary statistics for the grading prediction performance (LOCS- III based grading) of the deep learning system on retro-illumination images (CO).**

	AUC	Sensitivity	Specificity	Accuracy
<b>Grade 0 vs Others</b>	0.9757 (0.9671 – 0.9848)	91.94% (87.57%, 96.32%)	90.50% (86.56%, 96.32%)	91.10% (88.73%, 93.47%)
<b>Grade 1 vs Others</b>	0.7905 (0.7505- 0.8305)	73.79% (62.11%, 85.47%)	89.40% (86.83%, 91.98%)	87.62% (84.88%, 90.37%)
<b>Grade 2 vs Others</b>	0.7394 (0.7066 – 0.7722)	55.15% (44.37%, 65.92%)	92.85% (90.39%, 95.32%)	85.55% (82.15%, 88.95%)
<b>Grade 3 vs Others</b>	0.8350 (0.8098 – 0.8603)	68.17% (60.88%, 75.45%)	91.14% (87.04%, 95.25%)	87.14% (83.69%, 90.59%)
<b>Grade 4 vs Others</b>	0.9679 (0.9560 – 0.9799)	90.56% (80.92%, 100.0%)	94.73% (92.34%, 97.12%)	94.35% (92.14%, 96.56%)
<b>Grade 5 vs Others</b>	0.9667 (0.9525 – 0.9809)	83.49% (73.91%, 93.07%)	100.0% (100.0%, 100.0%)	99.45% (98.97%, 99.93%)
<b>Overall</b>	0.9044 (0.8958 – 0.9129)	78.02% (74.36%, 81.69%)	94.05% (92.78%, 95.32%)	91.33% (89.66%, 93.01%)
<b>Overall (ResNet18)</b>	0.8879 (0.8800 – 0.8958)	76.99% (73.46%, 80.52%)	93.76% (92.60%, 94.92%)	90.57% (88.78%, 92.35%)
<b>Overall (WideResNet50-2)</b>	0.8899 (0.8804 – 0.8994)	74.31% (70.51%, 78.10%)	92.79% (91.41%, 94.16%)	89.78% (88.02%, 91.54%)
<b>Overall (ResNext50)</b>	0.8948 (0.8842 – 0.9054)	77.50% (73.61%, 81.40%)	93.14% (91.87%, 94.42%)	91.15% (89.39%, 92.90%)

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; CO = cortical opacity; CI = confidence interval; AUC = area under the receiver operating characteristic curve