

Table S4. Summary statistics for the diagnostic performance of the deep learning system on retro-illumination images (CO).

	AUC	Sensitivity	Specificity	Accuracy
Normal vs Others	0.9659 (0.9546 - 0.9772)	92.69% (88.45%, 96.93%)	98.63% (97.94%, 99.31%)	96.21% (94.43%, 97.99%)
Abnormal vs Others	0.9688 (0.9589 - 0.9787)	98.63% (97.94%, 99.31%)	92.69% (88.45%, 96.93%)	96.21% (94.43%, 97.99%)
Overall	0.9680 (0.9579 - 0.9781)	96.94% (95.53%, 98.35%)	96.78% (94.83%, 98.74%)	96.21% (94.43%, 97.99%)
Overall (ResNet18)	0.9621 (0.9522 - 0.9720)	95.73% (93.66%, 97.79%)	94.26% (91.55%, 96.98%)	95.73% (93.66%, 97.79%)
Overall (WideResNet50-2)	0.9581 (0.9485 - 0.9677)	95.22% (93.3%, 97.14%)	93.70% (90.87%, 96.52%)	95.22% (93.3%, 97.14%)
Overall (ResNext50)	0.9533 (0.9423 - 0.9644)	96.14% (94.54%, 97.75%)	94.96% (92.63%, 97.3%)	96.14% (94.54%, 97.75%)

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

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