Table S3. **Summary statistics for the diagnostic performance of the deep learning system on slit-lamp images (NC).**

	AUC	Sensitivity	Specificity	Accuracy
Normal vs Others	0.9990 (0.9983 - 0.9998)	90.82% (83.74%, 97.9%)	100.0% (100.0%, 100.0%)	98.51% (97.39%, 99.62%)
Abnormal vs Others	0.9991 (0.9985 - 0.9998)	100.0% (100.0%, 100.0%)	90.82% (83.74%, 97.9%)	98.51% (97.39%, 99.62%)
Overall	0.9994 (0.9989 - 0.9998)	98.51% (97.39%, 99.62%)	92.31% (86.27%, 98.35%)	98.51% (97.39%, 99.62%)
Overall (ResNet18)	0.9992 (0.9987 - 0.9997)	98.10% (96.86%, 99.35%)	90.50% (84.39%, 96.6%)	98.10% (96.86%, 99.35%)
Overall (WideResNet50-2)	0.9990 (0.9983 - 0.9997)	98.21% (97.01%, 99.42%)	90.97% (85.03%, 96.9%)	98.21% (97.01%, 99.42%)
Overall (ResNext50)	0.9988 (0.9981 - 0.9996)	98.48% (97.4%, 99.56%)	94.17% (89.4%, 98.94%)	98.48% (97.4%, 99.56%)

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

Abbreviations: NC = nuclear color; CI = confidence interval; AUC = area under the receiver operating characteristic curve