Supplemental Information for:

Automated diagnosis of 7 retinal diseases with convolutional neural networks in a dataset of 2,234 eye images.

Neal Jain[*]

Monta Vista High School, 21840 McClellan Rd, Cupertino, CA 95014

Supplementary Tables

Table S1. Model performance on each of the five models.

Model	Test F1 Score Cataract	Test F1 Score Pathological Myopia
Resnet 18	90.1%	84.3%
Resnet 34	89.0%	77.6%
Resnet 50	89.4%	72.5%
Resnet 101	87.0%	82.8%
Resnet 152	89.3%	80.8%

5 different ResNet models were trained individually with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data.

Table S2. Model performance for each of the learning rates.

Learning Rate	Test F1 Score Cataract	Test F1 Score Pathological Myopia
0.1	32.5%	47.0%
0.05	35.5%	50.2%
0.01	50.6%	52.5%
0.005	44.4%	63.4%
0.001	85.5%	81.3%
0.0005	94.4%	86.0%
0.0001	94.8%	94.0%

A ResNet 18 model was trained individually on five different learning rates with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data.

Supplementary Figures

Figure S1. Model performance on each of the five models for pathological myopia. Five models were trained individually on pathological myopia with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data. The best F1 score after the testing process was completed was by the ResNet18 model, with a F1-score of 84.3%.
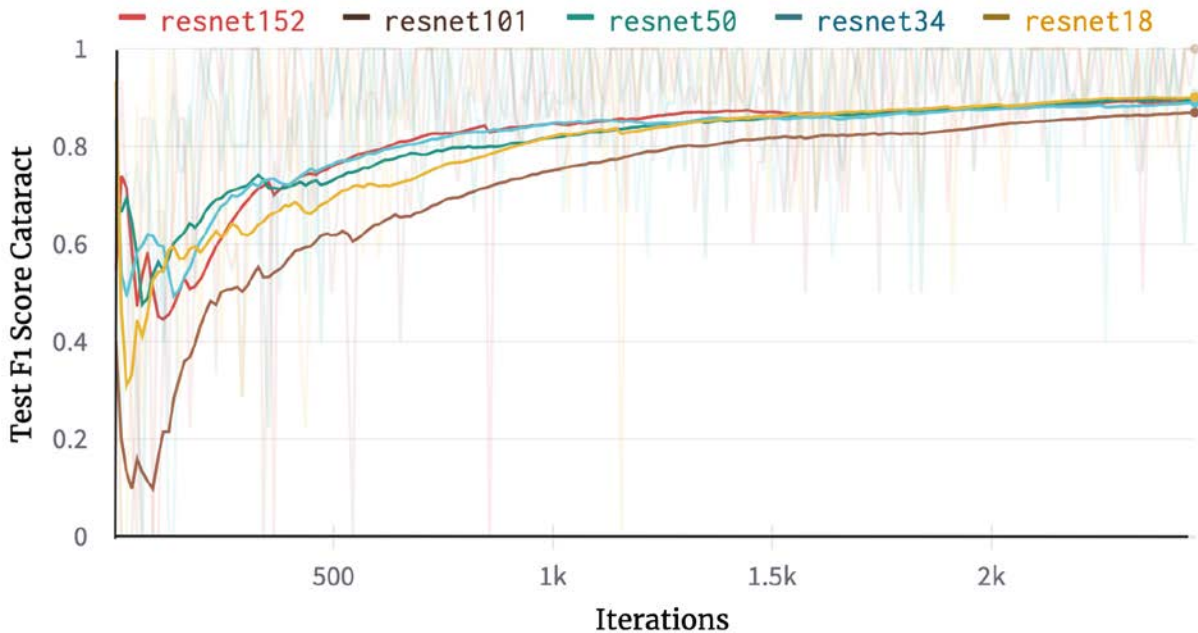

Figure S2. Model performance on each of the five models for cataract. Five models were trained individually on cataracts with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data. The best F1 score after the testing process was completed was by the ResNet18 model, with a F1-score of 90.1%.
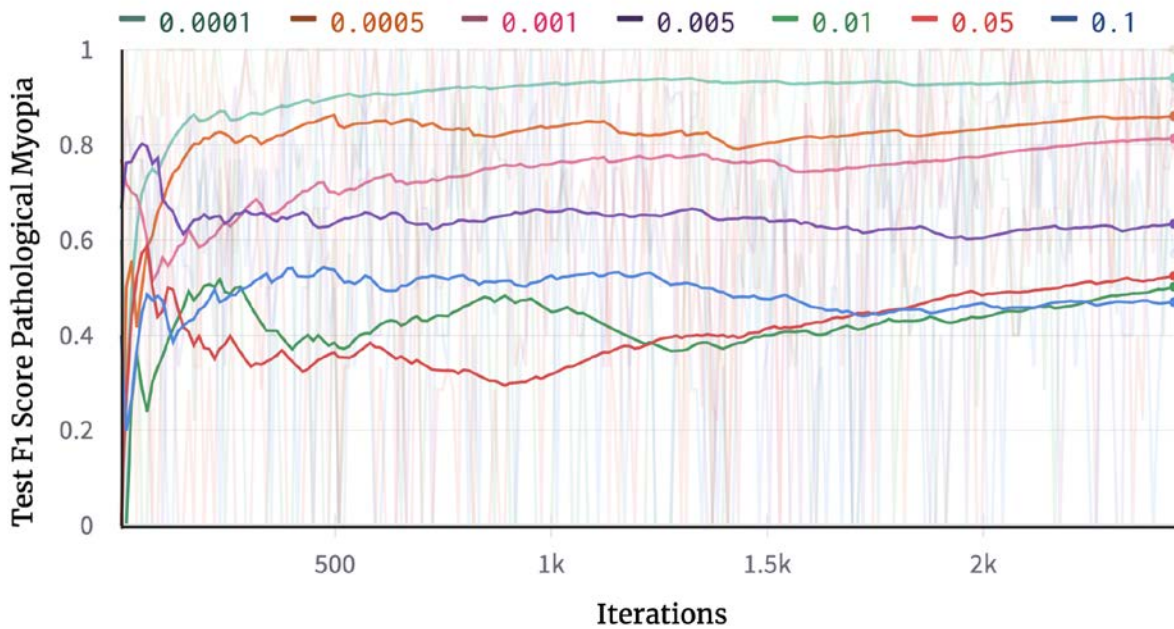
Figure S3. ResNet 18 performance on the seven learning rates for pathological myopia. The ResNet 18 model was trained individually on pathological myopia for seven F1 scores with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data. The best F1 score after the testing process was completed was by a learning rate of 0.0001 with a F1-score of 94.0%.
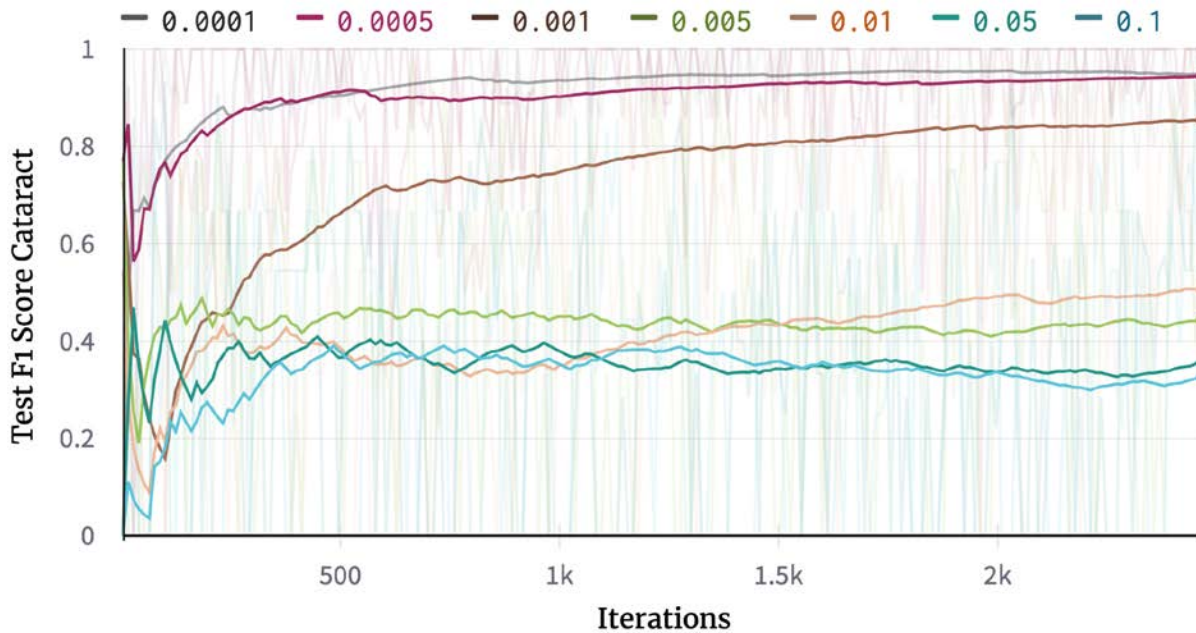

Figure S4. ResNet 18 performance on the seven learning rates for cataract. The ResNet 18 model was trained individually on cataract for seven f1-scores with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data. The best F1 score after the testing process was completed was by a learning rate of 0.0001 with a F1-score of 94.8%.
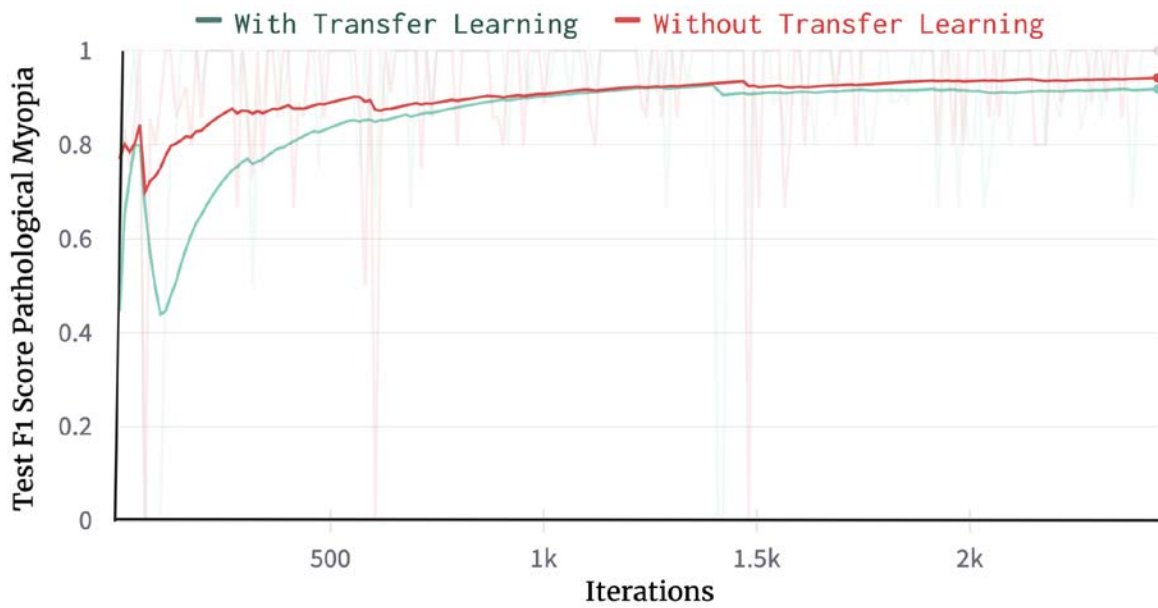
Figure S5. ResNet 18 performance with transfer learning and without transfer learning for pathological myopia. The ResNet 18 model was trained individually on pathological myopia for two F1 scores with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data. Transfer learning reduced the F1-score by 2%.
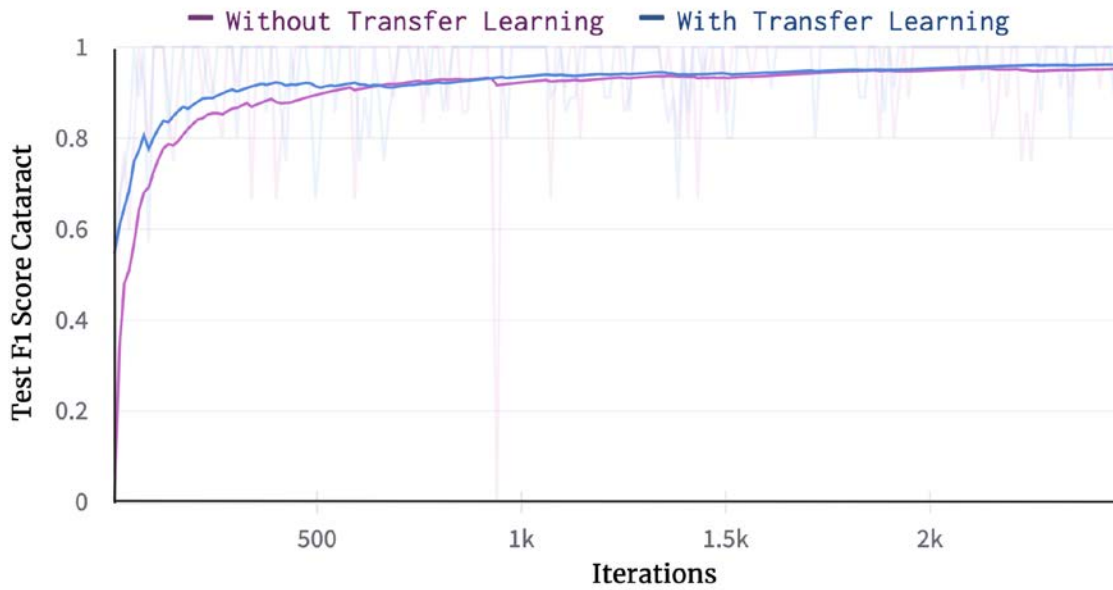

Figure S6. ResNet 18 performance with transfer learning and without transfer learning for cataract. The ResNet 18 model was trained individually on cataract for two F1 scores with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data. Transfer learning increased the F1-score by 2%.
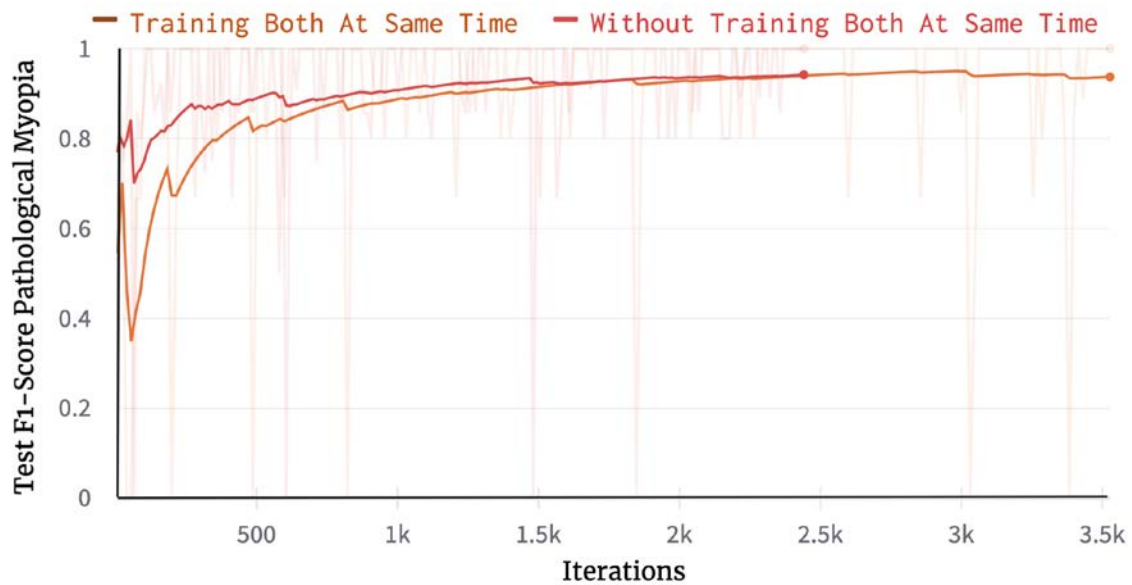
Figure S7. ResNet 18 performance with and without training both cataract and pathological myopia at the same time for pathological myopia. The ResNet 18 model was trained on both pathological myopia and cataract for one F1 score and individually on pathological myopia for another with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data.
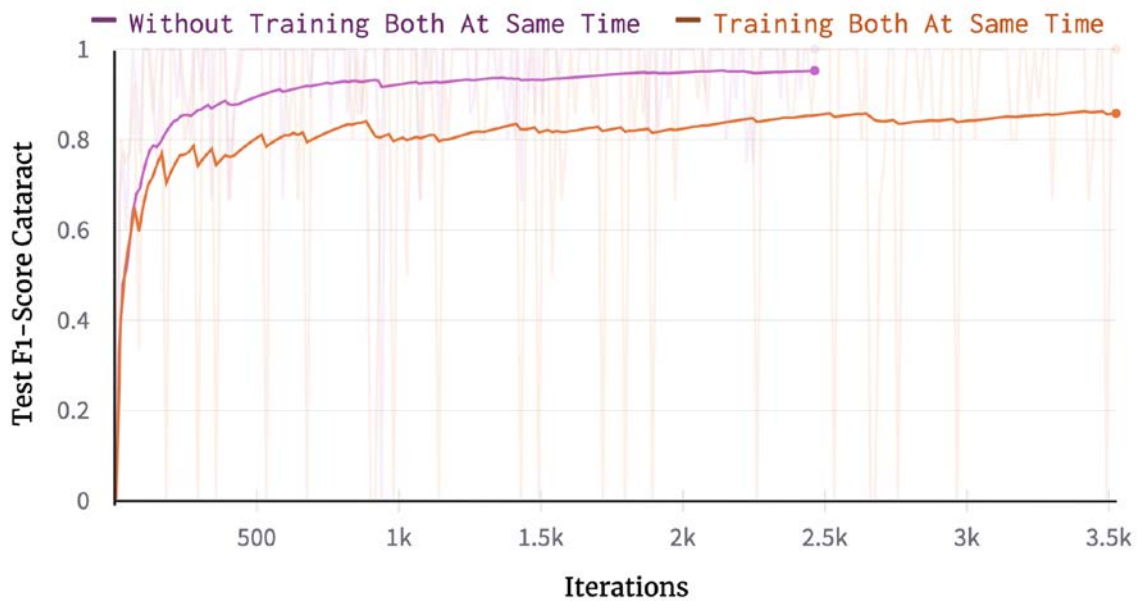

Figure S8. ResNet 18 performance with and without training both cataract and pathological myopia at the same time for cataract. The ResNet 18 model was trained on both pathological myopia and cataract for one F1 score and individually on cataract for another with testing F1 score performance calculated on an independent holdout testing set including 20% of the original data.