Tim Crnkovic

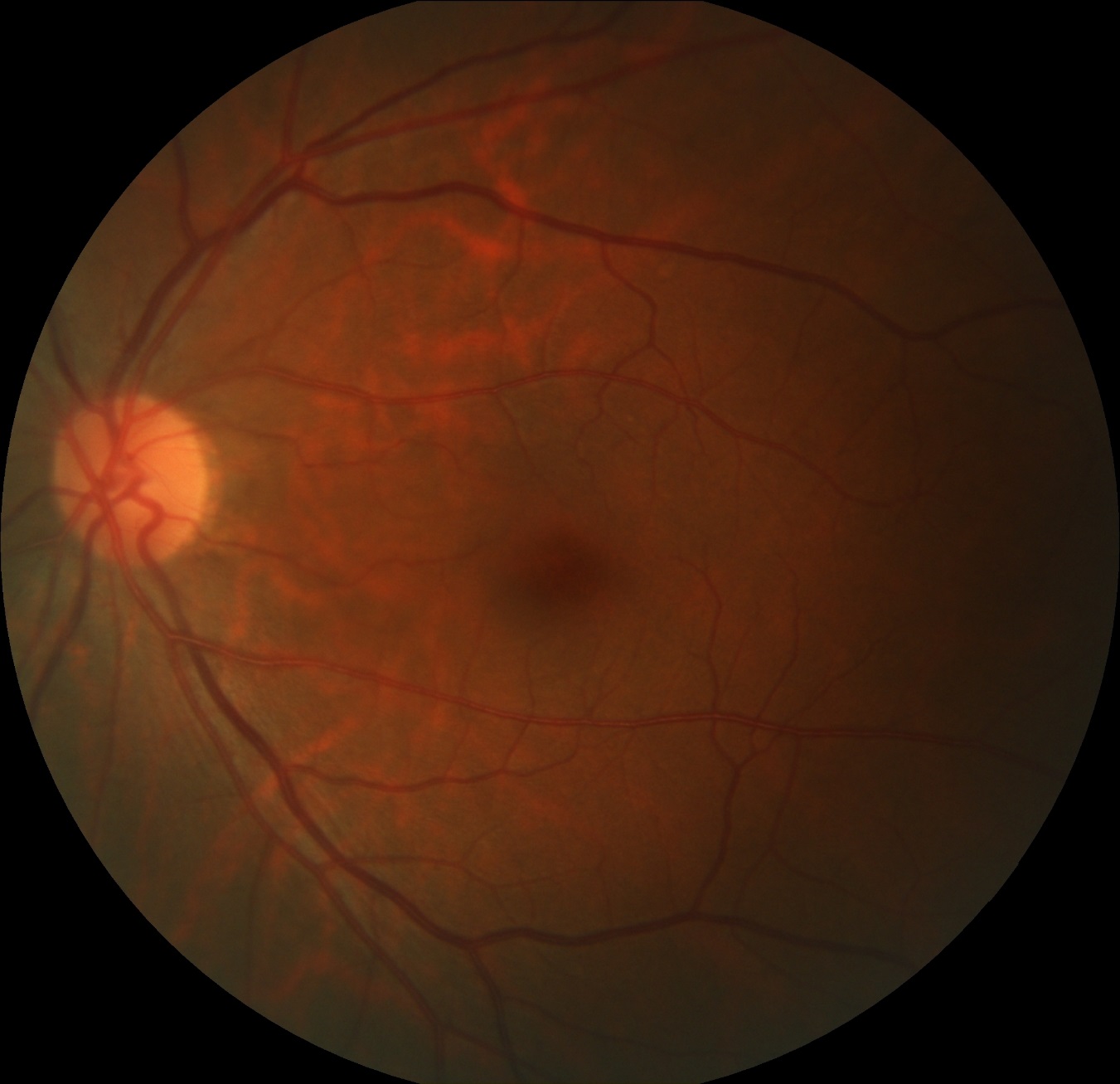
Classification of Ophthalmologic Fundus images for Age-Related Macular Degeneration using a Convolutional Neural Network

Introduction

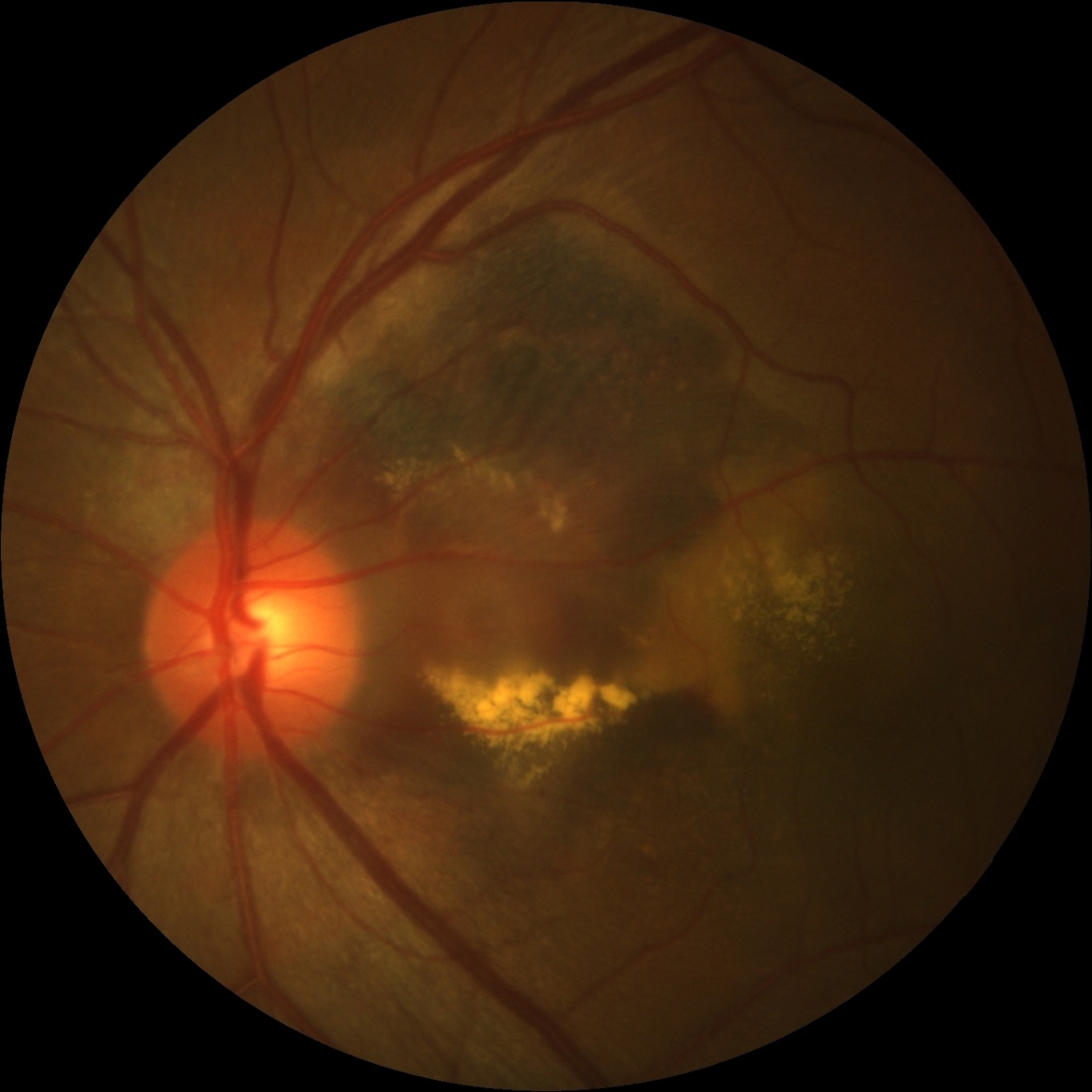
The iChallenge-AMD (<https://amd.grand-challenge.org/Home/> ) is a medical image analysis challenge for Age-Related Macular Degeneration (AMD), a disorder of the macular region of the eye. Eye fundus photography is one of the tools used to diagnose AMD. One of the purposes of this challenges is the classification of fundus imaging as either exhibiting signs of Age-Related Macular Degeneration or not.

Labelled fundus images were obtained from the challenge home page link shown above. There were 89 AMD images and 311 Non-AMD images. In order to reduce class imbalance, each AMD image was flipped first on the horizontal axis and then on both axes and all were labelled as AMD. So the total sizes of the data sets are 267 images of AMD and 311 images of Non-AMD.

Examples of images of both classes are shown below.



*Example of a Non-AMD image.*



*Example of an AMD image. Many images in this category are much more subtle than this.*

Methods and Results

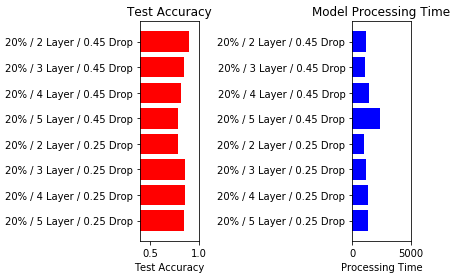
Using TensorFlow and Keras, several configurations and architectures of Convolutional Neural Networks (CNN’s) were tested. The goal was to train the models to accurately classify images as either AMD or Non-AMD. All configurations were tested with scaling of the original images to dimensions of 20% of the average original dimensions, to 25% of the average original dimensions and then to dimensions of 30% of the originals. Leaky ReLU was used for activation layers. Early stopping and dropout regularization was used to avoid overfitting. Class weights were used when the model was fit.

For each image scaling factor mentioned above, four different architectures were tested, one with 5 convolutional layers, one with 4, one with 3 and one with 2. For each architecture and each scaling factor, testing was performed first with 25% dropout regularization after a convolutional layers and then with 45% dropout regularization. In total, 24 models were tested. The results of this testing is presented here and also in three Colab notebooks. The testing with 20% image scaling is presented here: <https://colab.research.google.com/drive/1NOlR_7F93ExpWSkPMSr6v4DbNOxZz1N2> . The testing with 25% image scaling is presented here: <https://colab.research.google.com/drive/10Y6zlXZomwkQVidySBEmwXdGRP-GCRKk> . The testing with 30% image scaling is presented here: <https://colab.research.google.com/drive/1BlI2selQaULEW0iRvSJ_g1DkS9w_KbIv> .

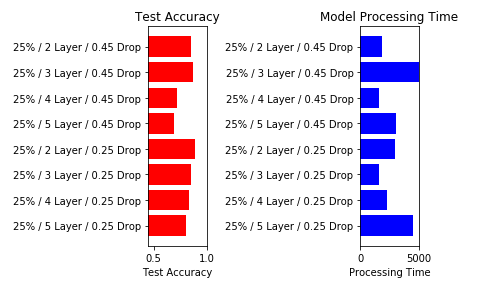
Each model was trained and tested using a 60/20/20 train/validation/test set split. Model performance was measured in accuracy of classification, although cross-entropy loss was also tracked. The time taken to train each model was also recorded. Additionally, an effort was made to determine which types of test images performed best and worst, that is, which types of images were most and least easily classified.

With a few exceptions, the models with 25% dropout outperformed those with 45% dropout. Also, in general, shallower models outperformed deeper ones. The models using the largest images – 30% scaling – slightly outperformed models using smaller images but with considerably longer processing time.

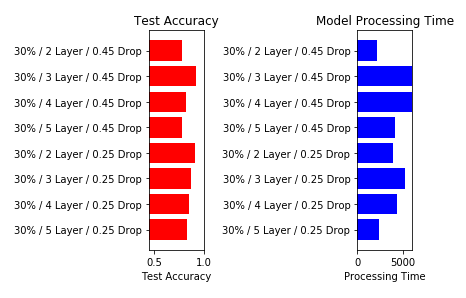
Overall, the best performing model used 30% image scaling, 3 convolutional layers and 45% dropout, which was slightly better than the one using 30% scaling, 2 layers and 25% dropout. Shown below are bar charts of performance and processing time for all models for the three scaling factors.



*Test accuracy and processing times for models using images scaled to 20%*



*Test accuracy and processing times for models using images scaled to 25*



*Test accuracy and processing times for models using images scaled to 30%*