Two models for lesion classification

E. Sargent, esargent at hmc dot edu

December, 2019

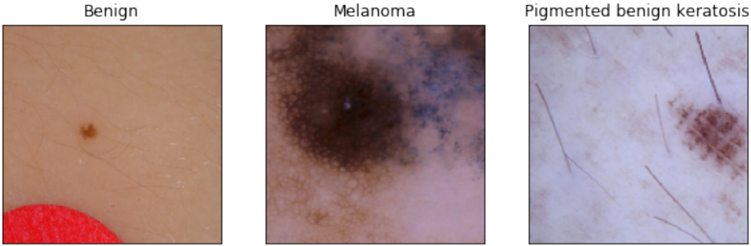
Outline

1. Introduction
   - ISIC Archive
   - Tools used
2. Binary Classifier
   - Preprocessing
   - Evaluation
3. Multiclassifier
   - Preprocessing
   - Evaluation

Outline

ISIC Archive

- 23k skin lesion images with associated clinical metadata.
- Most images have both a label of 'malignant' or 'benign,' as well as a subclassification with a specific diagnosis.
- The most common diagnoses are nevus (i.e. benign, n=18566), melanoma (n=2169), and pigmented benign keratosis (n=1099).

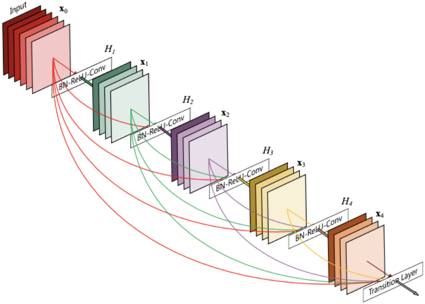


Outline

1. Introduction
   - ISIC Archive
   - Tools used
2. Binary Classifier
   - Preprocessing
   - Evaluation
3. Multiclassifier
   - Preprocessing
   - Evaluation

Tools used

- Pretrained Densenet-161 neural networks were fine tuned on the ISIC archive using PyTorch, in both the binary and multiclass case.
- All models were trained on Paperspace VM Cloud instance on a single GPU.
- Densenet is a neural network architecture introduced in [1] which lends itself to transfer learning with limited computational resources.



Outline

Preprocessing

- The image shapes are standardized by the procedure above and then loaded into large PyTorch Tensors.
- Initially, we employed a weighted cross-entropy loss function - misclassifications in the positive class were weighted by some $\gamma > 1$.
- Eventually changed up batch formation during preprocessing with `WeightedRandomSampler`
  - images were selected during batch formation with probability inversely proportional to their class frequency.

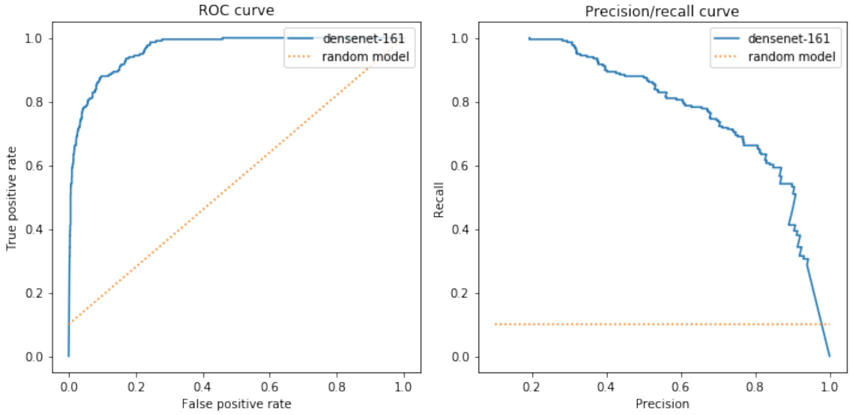
Outline

Evaluation

A batch size of 64 was used, the learning rate was tuned with the Adam optimizer, and the neural network was trained for 100 epochs.

Metric	**Value**
Accuracy	95.07%
True negatives	88.47%
False positives	1.568%
False negatives	3.366%
True positives	6.593%
F1 score	.7277
Average precision score:	.7863
Area under ROC curve	.9616

Evaluation (cont.)



Outline

Preprocessing

- The multiclassifier is trained on 9 lesion diagnosis classes.
- Using PyTorch's `ImageFolder` utility, we compose the following image augmentations pipeline.
  - Training
    1. `RandomResizedCrop((300, 300),`
    2. `RandomHorizontalFlip(),`
    3. `ToTensor()`
  - Validation
    1. `Resize((300,300)),`
    2. `ToTensor()`

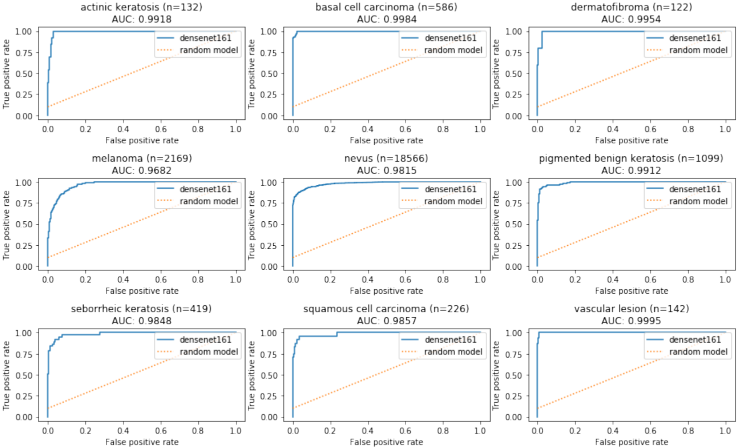
Preprocessing (cont.)

- Batches collated using both image augmentations and `WeightedRandomSampler`.
- This time, `WeightedRandomSampler` uses log inverse class frequency as weights.
    - $w_i = \log \frac{n}{n_i}$.
- We again use the Adam optimizer, the cross-entropy loss function, and train on a single GPU for 100 epochs.

Outline

Evaluation

The classifier has a balanced multiclass accuracy of 73.40%. The balanced multiclass accuracy is the arithmetic mean of the true-positive-rates for each class.



For Further Reading I

- G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger. "Densely Connected Convolutional Networks." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017. https://arxiv.org/pdf/1608.06993.pdf
- ISIC Archive, n.d,. https://www.isic-archive.com
- K. Hajian-Tilaki. "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation." Caspian J Intern Med 2013. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/pdf/cjim-4-627.pdf
- A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun. "Dermatologist-level classification of skin cancer with deep neural networks." January 2017. https://www.nature.com/articles/nature21056