Skin cancer self-diagnosis using mobile device dermatoscopic attachments

Chafik Taiebennefs, Tchafik@Stanford.edu
Stanford University, Winter 2020

Motivation

- Skin cancer is the most common form of cancer in the United States, with the annual cost of care exceeding $8 billion. With early detection, the 5-year survival rate of the deadliest form, melanoma, can be up to 99%; however, delayed diagnosis causes the survival rate to dramatically decrease to 23%.
- With the availability of affordable mobile phone skin magnifier attachments, people living in areas with limited access to healthcare services can leverage AI to have access to convenient skin cancer assessment, while reducing unnecessary and burdensome trips to healthcare providers.

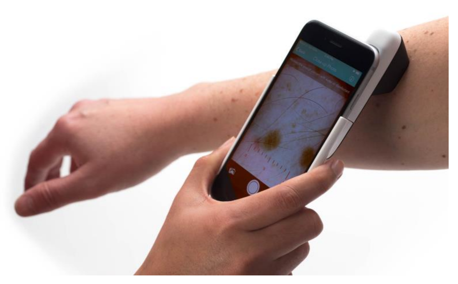

Figure 1: mobile phone skin magnifier attachment

Data

Original Data Set: The HAM1000 dataset is a large collection of multi-source dermatoscopic images of common skin lesions. The data set consists of 10,015 JPEG images which were made public through the International Skin Imaging Collaboration (ISIC) archive.
The images labels are stored in a CSV file and classified into 7 different disease categories: *Actinic keratosis(akiec); Basal cell carcinoma(bcc); Benign keratosis(bkl); Dermatofibroma(df); Melanoma(mel); Melanocytic nevus(nv); Vascular lesion(vasc).*

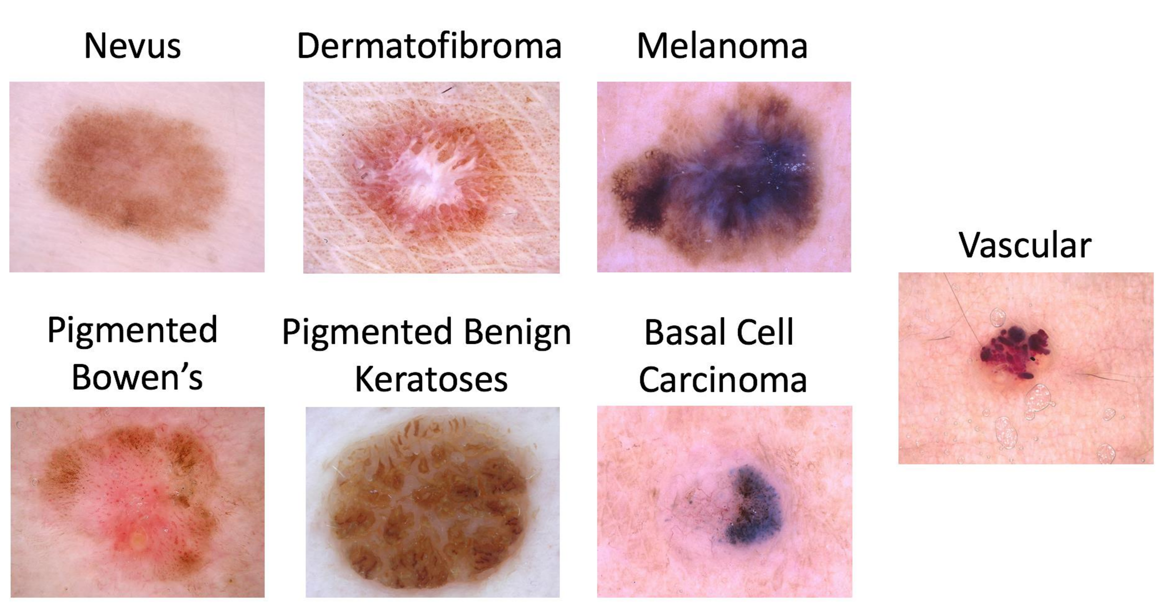

Figure 2: Sample HAM1000 data set images

Original data distribution: The HAM1000 dataset is heavily unbalanced with ~ 70% of the HAM1000 data set images belonging to the Melanocytic Nevus (NV) class.

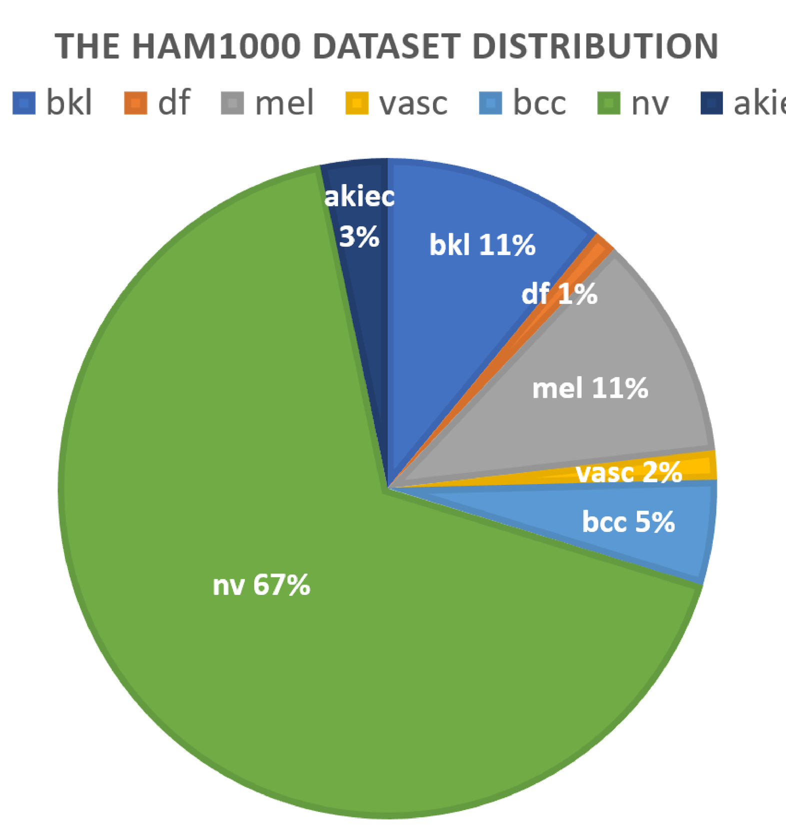

Figure 3: HAM1000 data distribution

Data processing:
- Normalized and resized images to 224 x 224 x 3 dimensions
- Shuffled images then stored them in a Numpy array on disk for faster data loading
- Split data set: 90% training set; 10% dev set
- Converted labels to stacked transposes of one-hot vectors.

Discussion

- The baseline Resnet50 model achieved an MCA performance ~ 70%. That's is to be expected as it is easy for the model to achieve close to 70% accuracy by simply predicting all images to be of the dominant class (70% images belong to the Melanocytic Nevus(NV) class).
- SqueezeNet model performance does not improve regardless of whatever architecture and hyperparameter values I tried. It seems the model is too small/simple to fit the data set. It is also overwhelmed by the unbalanced data.
- Using weighted loss function caused the most significant performance improvement.
- MobileNetv2 is a highly performant model that beat Resnet50 with 1/10th of the model size.
- Unbalanced data has a significant impact on the models training-to-dev performance gap.

Models & Approach

Approach: Due to the small size of my data set, my strategy was to leverage transfer learning using CNN models pretrained on the ImageNet data set.
Models: I experimented with three different CNN architectures:

- **ResNet50**

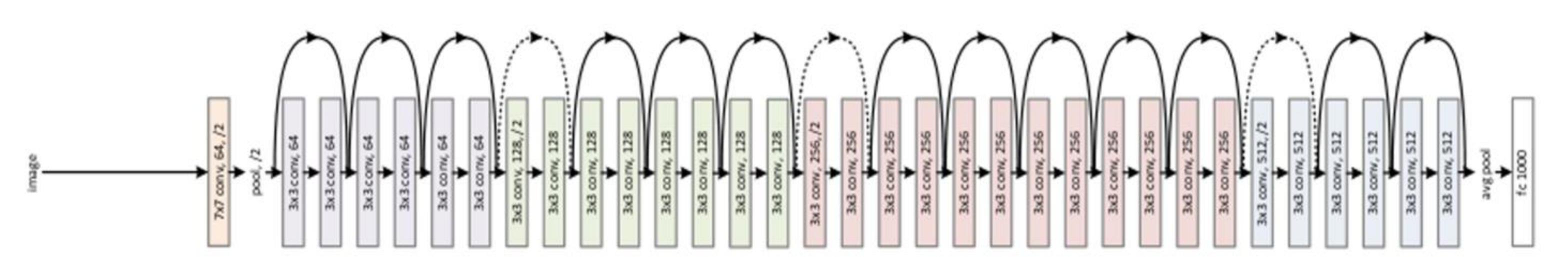

Figure 4: ResNet50 Architecture

- **MovileNetv2**

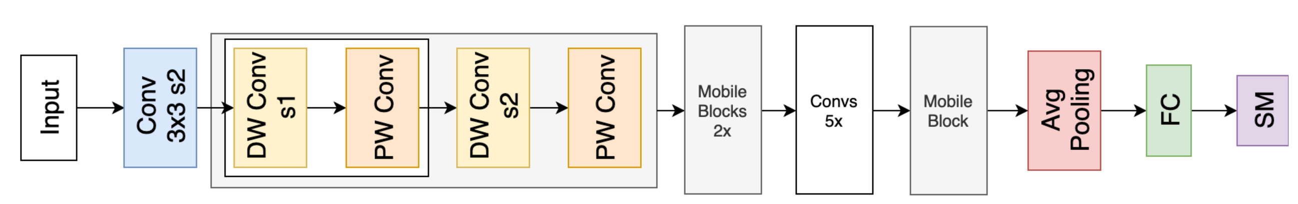

Figure 5: MobileNet Architecture

- **SqueezeNet**

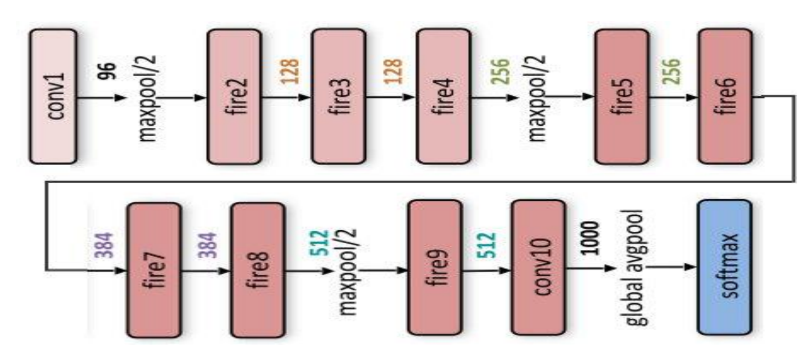

Figure 6: SqueezeNet Architecture

Transfer learning configuration:
- Transfer and freeze all layers and weights except the last fully connected (FC) layer and the Softmax layer
- Replace the removed FC layer with different FC architectures
- Replace the original softmax layer (1000 ImageNet classes) with a new Softmax layer (7 classes)
- Only train the weights of the new FC layers
- Use ReLU activation in the fully connected (FC) layers

Evaluation metric selection:
Given the unbalanced nature of my data set, I chose to use the balanced Multi-Class Accuracy (MCA), i.e. balanced recall, as a model evaluation metric

Loss function selection:
To alleviate the impact of the unbalanced data, I used a weighted loss function that multiplies the cross-entropy loss function by the frequency of classes. The impact of using the new loss function was significant: *the dev set MCA of my best performing Mobilenetv2 model jumped from 0.59 to 0.72.*

$$J = (\sum_{j=1}^{b} w_i Y_{ij}).(\frac{1}{b}\sum_{j=1}^{b}\sum_{i=0}^{n-1} Y_{ij}\log\hat{Y}_{ij})$$

Performance tuning parameters:
- Multiclass cross entropic loss function vs. multiclass weighted loss function
- Different FC layer architectures: 80, 160+80,, 240+160,
- Batch normalization
- Mini batch sizes: 16,32,64, 128
- # of epochs: 20, 30, 50, 100
- Learning rates: 2E-01 --- 2E-05
- Optimizer: Used ADAM only

Results

Mean Human Reader (Dermatologists & general practitioners)	Mean - Human Reader Expert (more than 10 years of experience)	Mean - Top 3 ML Models (2018 ISIC challenge)	Mean – ML Models (2018 ISIC challenge)	My Best Performing Model (MobileNetv2)
0.60	0.63	0.86	0.66	**0.724**

Sample Experiments:

Model Configurations	ResNet50 (size~100Mb)		MobelNetv2 (size~10Mb)		SqueezeNet (Size ~4Mb)	
	Training MCA	*Dev MCA*	*Training MCA*	*Dev MCA*	*Training MCA*	*Dev MCA*
Baseline: cross entropy loss, no batch norm, 120FC+softmax, 20 epochs, bs=16,lr=2e-05	0.694	0.453	___	___	___	___
Weighted loss, no batch norm, 120FC+softmax, 20 epochs, bs=16, lr=2e-05	0.673	0.492	___	___	___	___
Weighted loss, batch norm, 120FC+softmax, 20 epochs, bs=16, lr=2e-05	0.657	0.491	___	___	___	___
Weighted loss, batch norm, 120FC+Softmax , 100 epochs, bs=16, lr=2e-05	0.977	0.648	0.632	0.476	0.142	0.142
Weighted loss, batch norm, 240FC+80FC+Softmax, 100 epochs, bs=16, lr=2e-05	0.835	0.532	0.93	0.724	0.142	0.142

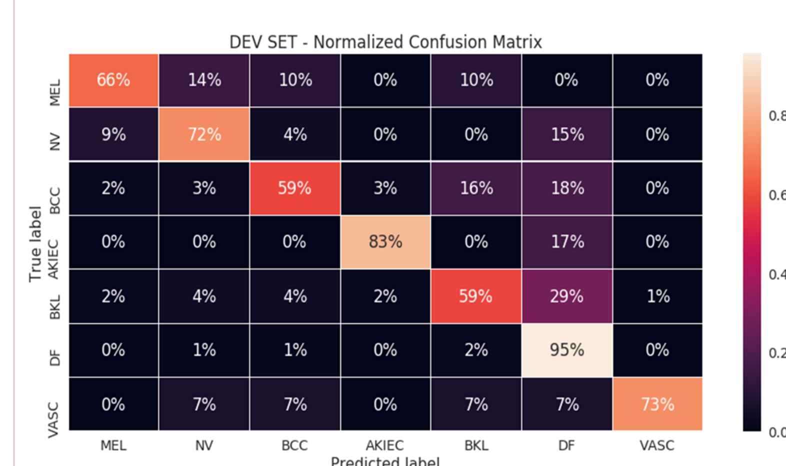

Figure 7: MobileNetv2 Confusion Matrix

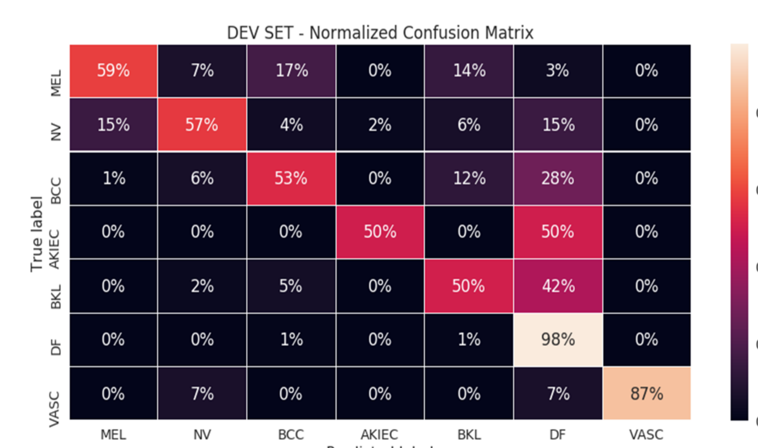

Figure 8: ResNet50 Confusion Matrix

Future Work

- Try to close the variance between my best model's training MCA (93%) and dev MCA(72%) using different regularization techniques (i.e. L1/L2 regularization) and data augmentation techniques (e.g. mirroring, cropping, etc.)
- Evaluate other mobile optimized architectures: i.e. the EfficientNet CNN architecture
- Collect additional data sets to help model generalize better
- Build a mobile app to test the model usability in the real world

References

- Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC); https://arxiv.org/abs/1902.03368
- The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6091241/
- Pruning Convolutional Neural Networks for Resource Efficient Inference; https://arxiv.org/abs/1611.06440
- Visual inspection for diagnosing cutaneous melanoma in adults; https://www.cochrane.org/CD013194/SKIN_how-accurate-visual-inspection-skin-lesions-naked-eye-diagnosis-melanoma-adults)
- Consumer acceptance of patient-performed mobile teledermoscopy for the early detection of melanoma; https://onlinelibrary.wiley.com/doi/full/10.1111/bjd.14630#bjd14630-fig-0001
- Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification; https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(19)30333-X/fulltext
- Tensornets: High level network definitions with pre-trained weights in TensorFlow; https://github.com/taehoonlee/tensornets