

# Skin Cancer Classification Using Deep Learning

Dwarkanath Prabhu

Department of Industrial and Systems Engineering)  
Texas A&M University  
College Station, TX, USA  
pdwarkanath@tamu.edu

Xiaoning Qian

Department of Electrical and Computer Engineering  
Texas A&M University  
College Station, TX, USA  
xqian@tamu.edu

**Abstract—Abstract** - The ISIC Challenge 2018 consists of 3 tasks. This report is aimed at tackling Task 3 - Disease Classification from images of skin lesions. The last few layers of several common models pretrained on ImageNet were retrained for this purpose and the best performing one was selected which turned out to be ResNet50. The class imbalance in the training data was dealt with using a class-weighted loss function and oversampling of low frequency classes.

## I. INTRODUCTION

In the United States, 5 million new cases of skin cancer are diagnosed every year. [1] Of these, melanoma which is the deadliest accounts for over 9000. [2] The diagnosis via visual inspection by patients and dermatologists is accurate only about 60% of the time. [3] Moreover, the shortage of dermatologists per capita has abetted the need for computer-aided methods to detect skin cancer. [4]

The International Skin Imaging Collaboration (ISIC) has aggregated a large amount of publicly accessible dermoscopy images labeled with ground truth data. The ISIC 2018 challenge [5] was divided into 3 tasks - Task1: Lesion Segmentation, Task 2: Lesion Attribute Detection and Task 3: Disease Classification. This report focuses on Task 3 i.e. classification of images into one of 7 possible classes.

## II. DATASET

There are 10,015 images in the labeled training dataset. [6] Some sample images from the dataset and their labels are shown in Fig. 1. The labels are in a CSV file in the form of one-hot vectors. There are no missing labels and all images are classified into one of 7 classes:

- Melanoma
- Melanocytic nevus
- Basal cell carcinoma
- Actinic keratosis / Bowen's disease (intraepithelial carcinoma)
- Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis)
- Dermatofibroma
- Vascular lesion

Evaluation metric for this task is the multi-class accuracy (MCA) i.e. the average of precision for all classes.

$$MCA = \frac{\sum_{i=1}^n P_i}{n} \quad (1)$$

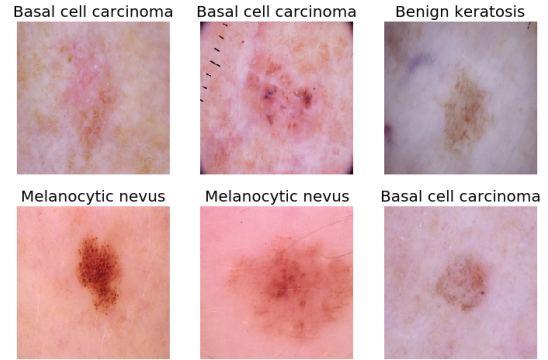


Fig. 1. Sample Images from Task 3 Dataset with Labels

where  $P_i$  is the precision of class  $i$

## III. ARCHITECTURE

Since this is a problem of image classification, a convolutional neural network (CNN) architecture would be most suitable. We used existing models such as VGG19 [7], SqueezeNet [8], Resnet50 [9], Inception [10] with weights pretrained on ImageNet [11] [12]. We found ResNet50 to give the best results among these models. Since ImageNet uses an input size of 224x224 and the images in the dataset are 450x600, the first step was to scale the images down to the required 224x224 size

## IV. IMPROVING PERFORMANCE

Approximately 70% of the images belong to only one class (Melanocytic nevus). Hence, it is trivial to achieve around 70% accuracy by simply predicting all images to be of that class. That is obviously incorrect. In order to improve performance, we try several techniques such as data augmentation, oversampling low-frequency classes, weighted loss etc.

As expected a baseline model with 10% of the labeled dataset randomly kept aside as a validation set achieved only 62.74% MCA while the training MCA was 93.67%. We will attempt to improve this discrepancy in the performance of training and validation set by using some techniques as follows.

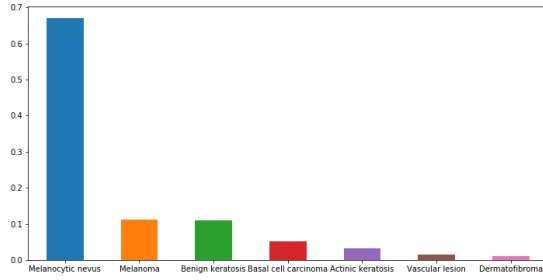


Fig. 2. Distribution of classes in the training set

#### A. Batch Normalization

Since we are using ReLU activation in the fully connected layers and the final output is a softmax i.e. a number between 0 and 1, batch normalization [13] could help speed up training by scaling the output of the fully connected layers appropriately. The performance on the validation set improved slightly to 73.29% but the training MCA went down to 87.77%.

#### B. Data Augmentation - Mirroring

There is still a large difference between the validation and training MCA. Training the model on a larger dataset could help bridge this gap. We can double the dataset by simply **taking mirror images** [14] of the existing dataset while keeping the labels constant. The validation MCA increased to 76.27% while the training MCA was up to 92.85%.

#### C. Weighted Loss

The model still predicts the dominating class more often than it should while ignoring lesser occurring classes. One way to fix this is to penalize the model for predicting the dominating class. [15] This can be **done by multiplying the loss function by the frequency of classes. Thus, a new weighted loss function** can be used to train the model. Training the model with the weighted loss function got a validation MCA of 75.73% and training MCA of 95.25%.

#### D. Oversampling

The presence of classes Dermatofibroma and Vascular lesion is very low in the dataset (~0.01%). We can increase their occurrence by taking random crops of the central part of the image so that the lesion still remains in the image. [14] We took 4 random crops of images belonging to these classes and also their mirror images while keeping the labels constant. These were then added to the dataset from which 90% of the data was randomly selected for training. The validation MCA shot up to 87.47% as a result while training MCA was 98.08%.

### V. RESULTS

The results achieved from training using the techniques listed above are shown in Table 1.

TABLE I  
EFFECT ON MODEL PERFORMANCE

Technique	Training MCA	Validation MCA
Baseline	93.67%	62.74%
Batch Normalization	87.77%	73.29%
Data Augmentation - Mirroring	92.85%	76.27%
Weighted Loss	95.25%	75.73%
Oversampling	98.08%	87.47%

### VI. CONCLUSION AND DISCUSSION

The ResNet architecture with data augmentation is able to perform much better than the baseline but there is still a difference between the training and validation metrics. The best training MCA is over 98% but the best validation MCA is about 87.5%. Since the training and validation sets are randomly selected from the same dataset, the chances of data mismatch are minimal. The training-validation gap may be converged further by training on a larger dataset. Also, there is a possibility of illumination affecting the database which can be corrected using color constancy on the entire dataset.

### VII. ACKNOWLEDGMENTS

The authors would like to thank Texas A&M High Performance Research Computing (HPRC) for providing computational resources. Also, we would like to thank Taehoon Lee who trained several models on the ImageNet dataset and provided an open source implementation in Tensorflow.

### REFERENCES

- [1] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012," *JAMA dermatology*, vol. 151, no. 10, pp. 1081–1086, 2015.
- [2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [3] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The lancet oncology*, vol. 3, no. 3, pp. 159–165, 2002.
- [4] A. B. Kimball and J. S. Resneck Jr, "The us dermatology workforce: a specialty remains in shortage," *Journal of the American Academy of Dermatology*, vol. 59, no. 5, pp. 741–745, 2008.
- [5] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pp. 168–172, IEEE, 2018.
- [6] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *arXiv preprint arXiv:1803.10417*, 2018.
- [7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
- [8] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," *CoRR*, vol. abs/1602.07360, 2016.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.
- [10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014.

- [11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, “Imagenet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [12] T. Lee *et al.*, “Tensornets,” 2017.
- [13] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.
- [14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.
- [15] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.