Duke + DataScience

Mobile Classification of Pigmented Skin Lesions

Adrian Lopez, Meredith Brown, Brian Kim, Amy King, Felix Ogbeide, Kinjal Sanghvi, and Matt Engelhard

Motivation & Objectives

More people are diagnosed with skin cancer in the U.S. each year than all other cancers combined. One in five Americans develop skin cancer by age 70, and the estimated annual cost of treating skin cancer in the U.S. is 8.1 billion dollars [1]. Skin lesions, defined as a region of skin with abnormal appearance or growth, can be cancerous or precancerous. Prompt detection and removal of such lesions is critical to prevent further growth or spread. Our project aims to provide a solution which automates the diagnosis of skin lesions to aid in accurate, early detection of malignant (i.e. cancerous) skin lesions that does not require direct access to physicians. Our objective is to implement the MobileNetV2 CNN [2] in the detection of the seven classes of pigmented skin lesions in the HAM10000 data set. After training this model, we will use it in a mobile application.

Data & Methods

Human Against Machine with 10000 Training Images (HAM10000)

bkl: benign keratosis

bcc: basal cell carcinoma

nv: melanocytic nevi

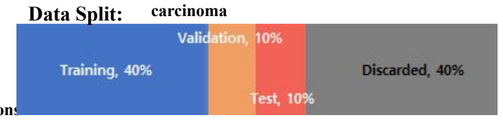
akiec: actinic keratosis and intraepithelial carcinoma

df: dermatofibroma

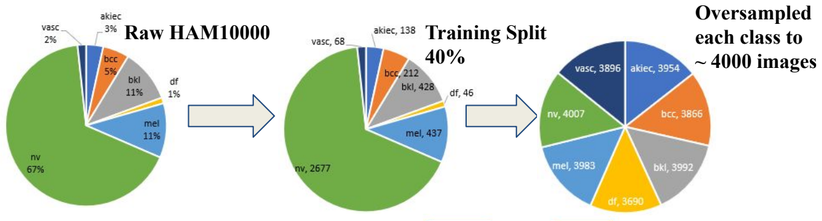
mel: melanoma

vasc: vascular lesions

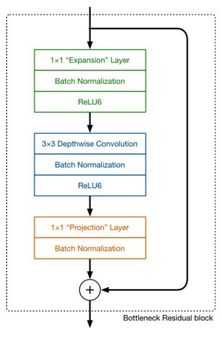
Data Split:

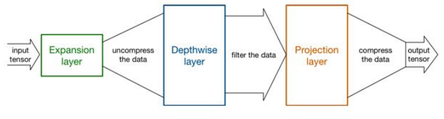


Data Augmentation: Rotation, Zoom, Flip, Brightness Adjustment



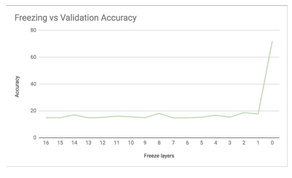
MobileNetV2:





MobileNetV2 is a second generation network architecture that incorporates novel "bottleneck" residual blocks which has shown to best optimize speed, memory usage, and battery consumption making it the best choice for mobile implementation.

Training: We experimented with data augmentation, class weight regularization, dropout, freezing and unfreezing blocks, optimizer, optimizer learning rate, batch size, number of epochs.

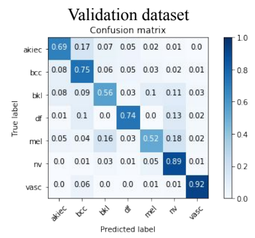


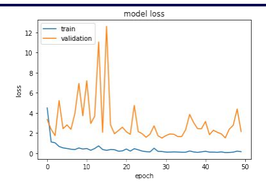
Results

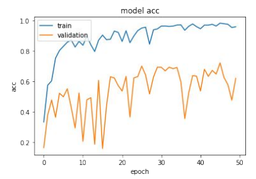
Training Results:

Hyperparameters:

- Optimizer: Adam, learning rate = 1e-4
- Batch size: 42, Epochs: 50
- All layers trainable
- Loss function: Categorical cross-entropy
- Accuracy metric: Categorical accuracy





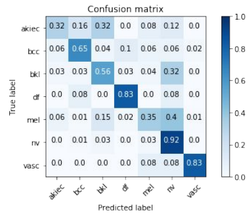


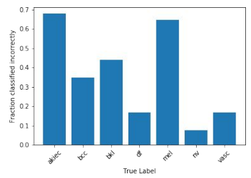
	Training	Validation	Test
Loss	0.06869	1.509	1.087
Accuracy	97.61%	72.20%	77.54%

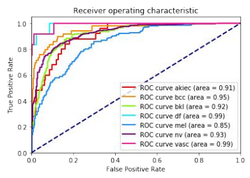
Performance on the Test Dataset: 77.54% Accuracy:

Tends to misclassify:

- melanoma as nevus
- actinic keratosis as benign keratosis



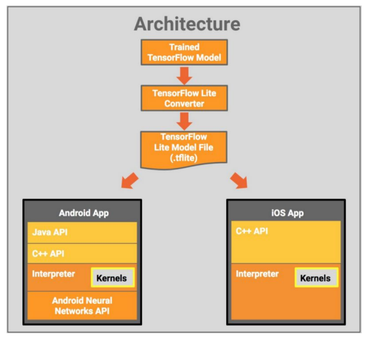




Class	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
akiec	0.32	0.98	0.33	0.98
bcc	0.65	0.99	0.70	0.98
bkl	0.56	0.94	0.55	0.94
df	0.83	0.99	0.43	0.99
mel	0.35	0.97	0.62	0.90
nv	0.92	0.72	0.86	0.83
vasc	0.83	0.99	0.77	0.99

Mobile Application:

The mobile application uses TensorFlow Lite which is an open source deep learning framework for mobile devices. Tensorflow Lite is useful to optimize our model graph and pruning unused graph-nodes to reduce app footprint and improve performance.



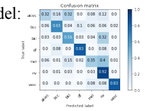
Development: We built the mobile application on the Android platform and included a TensorFlow Lite object which takes the input layer and the output layer from our implementation of the MobileNetV2 CNN.

Conclusions

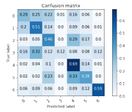
For the HAM10000 dataset, undersampling from the majority nv class and oversampling via data augmentation from minorities classes yielded the best model performance. Using the MobileNet V2, layers were frozen up until the final layer of the model. This reduced the number of parameters, but adversely affected the validation accuracy. By unfreezing all layers, thus training the model on all, the accuracy reached its highest point. Likewise, a higher batch size, a higher number of epochs on which the model could train, and a learning rate of 10e-4 were discovered to be best for increasing validation accuracy.

Compared to the results of other kaggle projects using the HAM10000 dataset [3], our confusion matrix indicates we more accurately classified skin lesions. Using a sequential CNN, the results obtained [3] gave a validation accuracy of 51%. Another kaggle member [4] obtained an accuracy of 72% but both these results were biased by class imbalance, as you can see by the confusion matrix.

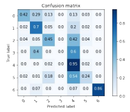
Our model:



[3]



[4]



Limitations of our model are that the training data did not include images of healthy skin. This means the model may be biased towards diagnosing an image as a type of lesion. This issue may be resolved by setting a minimum threshold on the returned softmax probability of the predicted class to be counted as a meaningful classification. Also, the pairs of classes that our model is most often confusing (melanoma vs nevus, actinic keratosis vs benign keratosis) are ones that are important to distinguish from a clinical perspective. Training our model on a dataset with more images from these classes would enable improved cancer detection.

Future Work

Future work includes running this model on a mobile application using TensorFlow Lite. The goal is to deploy the model on cell phones for users to receive real-time evaluations of lesions. Objectives for future work include:

- Ensuring equal accuracy on mobile phones.
- Extending mobile application across other platforms such as IOS.

References

[1]https://www.skincancer.org/skin-cancer-information/skin-cancer-facts
[2]MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381
[3]https://www.kaggle.com/yuningalexliu/dermatology-image-classification/output
[4]https://www.kaggle.com/sid321axn/step-wise-approach-cnn-model-77-0344-accuracy

Acknowledgements: This work was supported by the Duke Plus DS (+DS) Program. We thank Duke Research Computing for the use of the Duke Compute Cluster for high-throughput computation. The Duke Data Commons storage is supported by the National Institutes of Health (1S10OD018164-01). Health Data Science at Duke is supported by the Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002553.

Student Showcase: +DS Projects in Medicine and Health Data Science Internship Program

April 25, 2019, Duke University, Gross Hall