Skin Cancer Detection

Kenny Luu – kepluu@ucsc.edu
Sinclair Liang – wliang13@ucsc.edu

Abstract

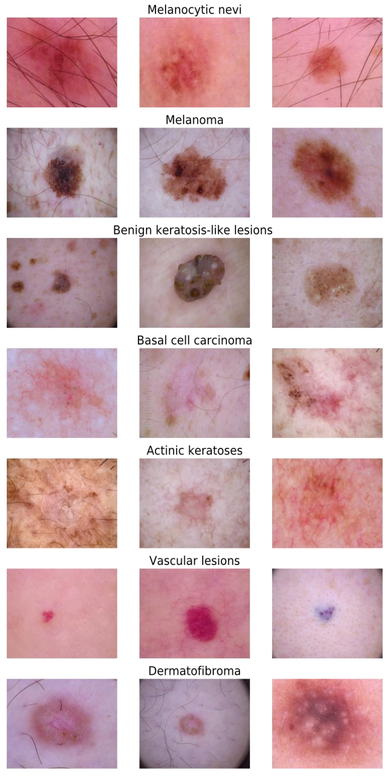
The differentiation and identification of cancerous skin lesions is usually done through histopathology, a microscopic examination of the suspect tissue. What we wanted to achieve was an accurate model which can discern from seven different types of cancerous and non-cancerous skin lesions from an image of the skin lesion. We achieve this through the training of of convolutional neural networks.

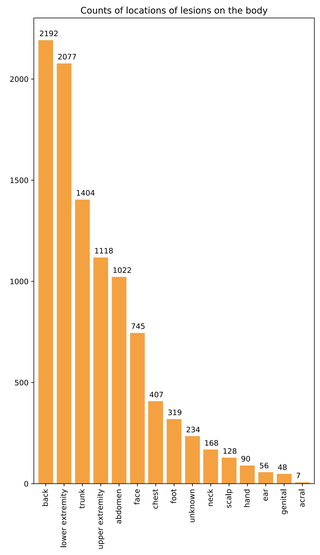
Introduction

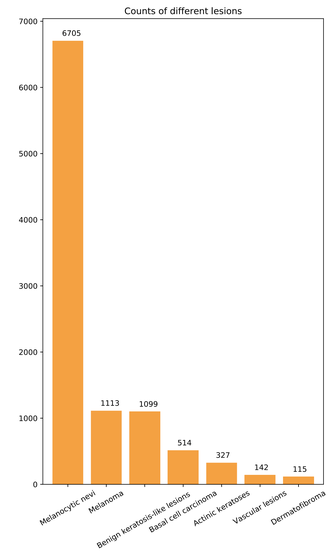
Skin cancer is one of the most common human cancers. They are usually diagnosed by initial clinical screening then followed by a confirmatory examination. By developing a model to automate some parts of the process, it can take some of the burden off from a medical professionals' shoulders. In this dataset there were seven different types of skin lesions as depicted below. The different types of skin lesions in this dataset are either cancerous or pre-cancerous although some are just benign. The dataset from which we trained our model on contains 10,015 images, each of which represented one of seven different types of skin lesions.

The types of lesions

Melanocytic Nevi (nv)- A usually noncancerous lesion of pigment producing skin cells commonly called birth marks or moles
Melanoma (mel) - One of the most serious and deadly types of skin cancer
Benign Keratosis-like Lesions (bkl) - A harmless growth on the skin
Basal Cell Carcinoma (bcc) - A type of cancer that arises due to sun exposure, this type of cancer is not very deadly
Actinic Keratosis (akiec) - A scaly growth caused by damage from exposure to the sun, the most common type of precancerous skin lesion
Vascular Lesions (vasc) - Comprise of all skin diseases that originate from or affect blood, including malignant or benign tumors
Dermatofibroma (df) - A benign growth of the skin







Methodology

Cleaning the Data

This dataset contains meta information about the images which included: age, sex, localization of the lesion, and the method of determination of the type of lesion. The images in this dataset were 600 pixels by 450 pixels, which needed to be compressed to a size of 120 pixels by 90 pixels. The data was quite clean when we unpacked it, there were few missing fields and unknown data. The missing fields in the age category were replaced by the mean of all the ages, and the unknown fields in the gender field were replaced by a random (50/50) assignment of Male and Female.

Building the Model

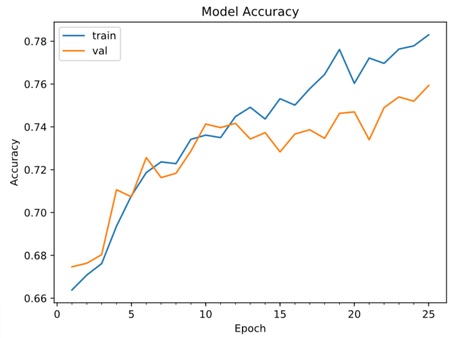
Our model was built using Keras with the TensorFlow backend, it contains 9 layers. We will be using a 3 modules stacked right on top of each other, each module contains a convolutional layer and a max pooling layer. Our convolutions operate on 3x3 window, with ReLu activations and our max pooling layers operate on 2x2 windows. The first convolutional layer produce 64 filters, the second one produce 128, and the third and final one will also produce 128 filters. At the end of the 3 convolutional layers we had a dropout layer which dropped 10% of the nodes and weights. We then take that output and flatten it into a 1-D tensor then add a fully-connected Dense layer with 128 nodes at the end, and then added another dropout layer which dropped 30% of the nodes and weights. Finally the last layer was a fully-connected with 7 outputs with a SoftMax activation function. To prevent overfitting we preprocessed the images in Keras so that they would be randomly rotated, horizontally flipped or zoomed into.

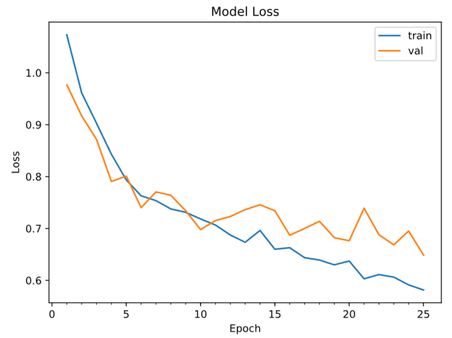
Training the Model

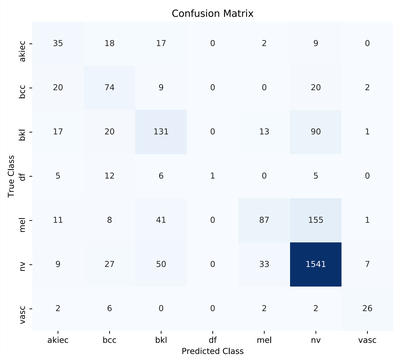
The model was trained using 6000 images in the training set and 1500 images in the validation set. The rest of the images were then separated and used for the testing set. We trained the model for 25 epochs, with each epoch being 60 steps, with batches of 100 images.

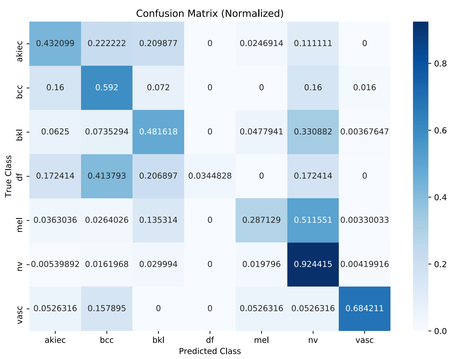
Results

Our model was able to achieve an accuracy of **75.35%** on data it has never seen before. Looking at the confusion matrices we can see that our model very rarely predicts that a skin lesion is a Dermatofibroma as there is so little data on it. Although our confusion matrices indicate good results it is apparent that our model tended to mislabel Melanomas as Melanocytic Nevi which is not a desired outcome as Melanomas are cancerous while Melanocytic Nevi are not.









References

Dataset:
Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse

Baskin Engineering
UC SANTA CRUZ