Machine Learning Engineer Nanodegree

Capstone Proposal

Felix Gruener

(felixgruener@hotmail.com)

September 28th, 2018

Proposal

Abstract

The project proposal is to create a Webservice based on an ML model which accurately predicts Skin Cancer. In order to do that , the User needs to take a picture of his Skin with his Cell Phone or any Device that own a digital Camera. The picture will be uploaded by any browser by a HTTP Request and our Webservice will predict by our pretrained Resnet Model plus own customized Model on Top if the Picture contains Skin Cancer or not .

Keywords: Supervised learning, scikit learn, keras, tensorflow.

Domain Background

The medical sector is one of the most promising field in machine learning where already much development has taken place and is currently used in real world application. Skin Cancer is Number 1 Cancer with highest decease rate for humans. Finding a Skin Cancer in Phase 0 for humans gives a probability of nearly 90% surviving the skin cancer. This Survival Probability already sinks to 10-15% in Phase 4

So my personal motivation is to be of direct use for saving Life with machine learning skills.

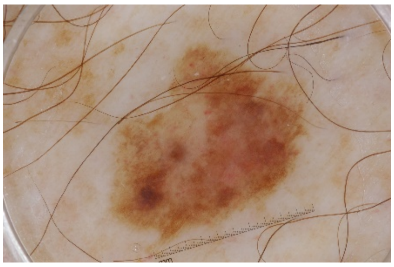
Problem Statement

Skin cancer is the most common cancer in the world. In the U.S., we find 5.4 million new cases of skin cancer every year. They come in different types. Some are called Carcinomas, some are called Melanomas. Melanomas are the ones that typically kill people, it's called the black cancer. Twenty percent of Americans will eventually get skin cancer. In most cases, benign cancer. Pre-cancer also cause a Actinic Keratosis,affects 58 million Americans and many more in the world. We have, in this country,76,000 Melanomas each year and 10,000 deaths.To put this in perspective,traffic accidents in this country are 40,000, so,you are about four times as likely to die in a traffic collision than a skin cancer.The cost for United States per year is 8.1 billion in treatment and diagnostics. This is a classification problem the App will be classifying 3 classes. One Class is Melanoma which is bad and cancerous, the other two classes are good and mean no cancer: Nevus and Seborrheic Keratosis. The Features are categorical such as dermoscopic features for extraction

Datasets and Inputs

The dataset contains training, validation, and test images in the `data/` folder, at `data/train/`, `data/valid/`, and `data/test/`, respectively. Each folder should contain three sub-folders (`melanoma/`, `nevus/`, `seborrheic_keratosis/`), each containing representative images from one of the three image classes. Challenge provides training data (~2000 images) for participants to engage in all 3 components of lesion image analysis. A separate public validation dataset (~150 images) and blind held-out test dataset (~600 images) will be provided.

Sample Data:



Melanoma

Nevus

The data and objective are pulled from the [2017 ISIC Challenge on Skin Lesion Analysis Towards Melanoma Detection]. As part of the challenge, participants were tasked to design an algorithm to diagnose skin lesion images as one of three different skin diseases (melanoma, nevus, or seborrheic keratosis

Solution Statement

I use the training and validation data to train a model that can distinguish between the three different image classes `melanoma/`, `nevus/`, `seborrheic_keratosis`. (*After training, I used the test images to gauge the performance of my model.*)

Some of the algorithms that were successful in this competition, please read **this article** that discusses some of the best approaches. So i used ResNet 2 as pretrained Model for Transfer Learning which is available in Keras Library 2.2.2.

First importing the Datasets then loading the images into tensors. There will be no Augmentation because since I use pretrained ResNet2 image transformation wont help much here. Then I load the pretrained ResNet2 and model my Network on Top of it using the Feature of ResNet2 as Input. Training the model and validating will give us the success rate.Since this is a Classification Problem a CNN like ResNet2 is proven by 2017 ISIC Challenge on Skin Lesion to work very good.

Benchmark Model

The algorithm can be ranked according to three separate categories and scale of Probability. It will prove that the success rate is higher than average dermatologists success rate to classify malignant skin tumor.

Category 1: ROC AUC for Melanoma Classification

In the first category, gauges the ability of the CNN to distinguish between malignant melanoma and the benign skin lesions (`nevus/`, `seborrheic_keratosis`) by calculating the area under the receiver operating characteristic curve (ROC AUC) corresponding to this binary classification task.

Category 2: ROC AUC for Melanocytic Classification

In the second category, the ability of the CNN will be tested to distinguish between melanocytic and keratinocytic skin lesions by calculating the area under the receiver operating characteristic curve (ROC AUC) corresponding to this binary classification task.

Category 3: Mean ROC AUC

In the third category, I will take the average of the ROC AUC values from the first two categories.

Evaluation Metrics

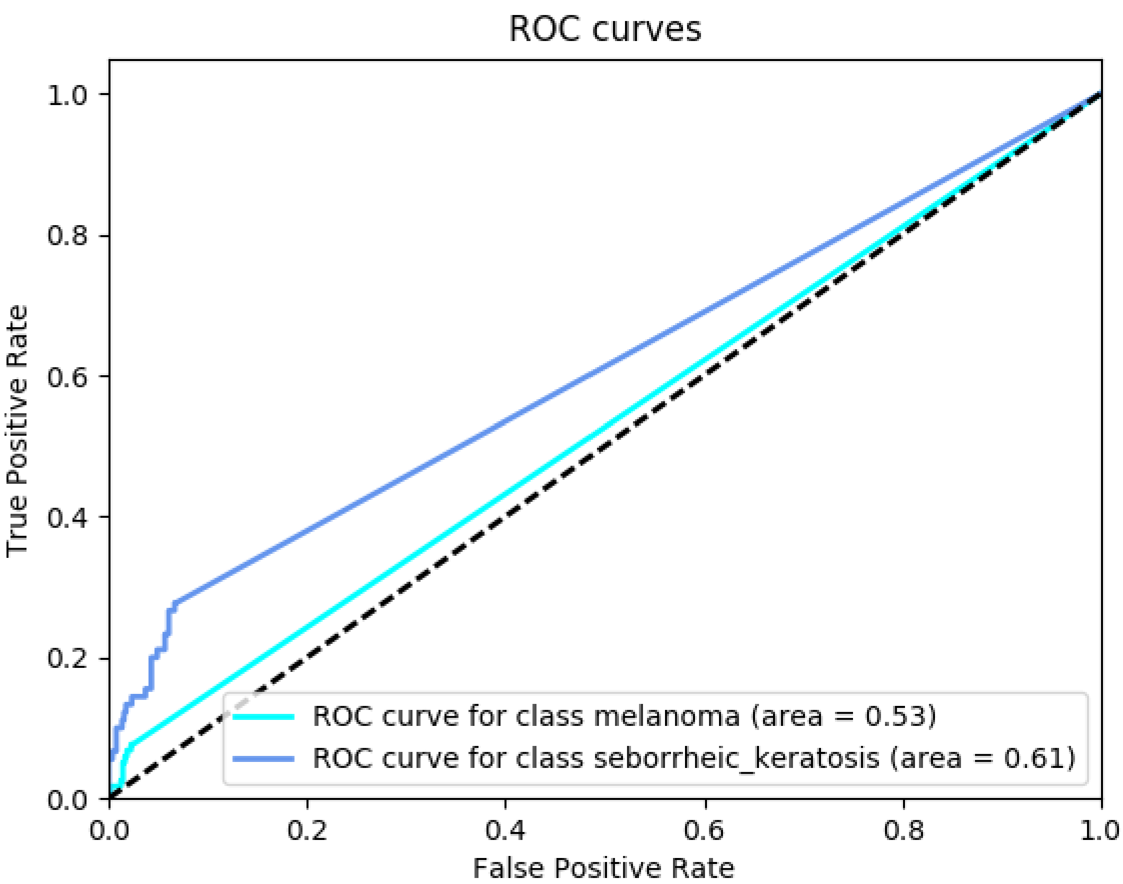
Once I have trained your model, I create a CSV file to store test predictions. My file should have exactly 600 rows, each corresponding to a different test image, plus a header row. I included an example submission file (sample_submission.csv) in the repository. The benchmark model and the model work on the same dataset.

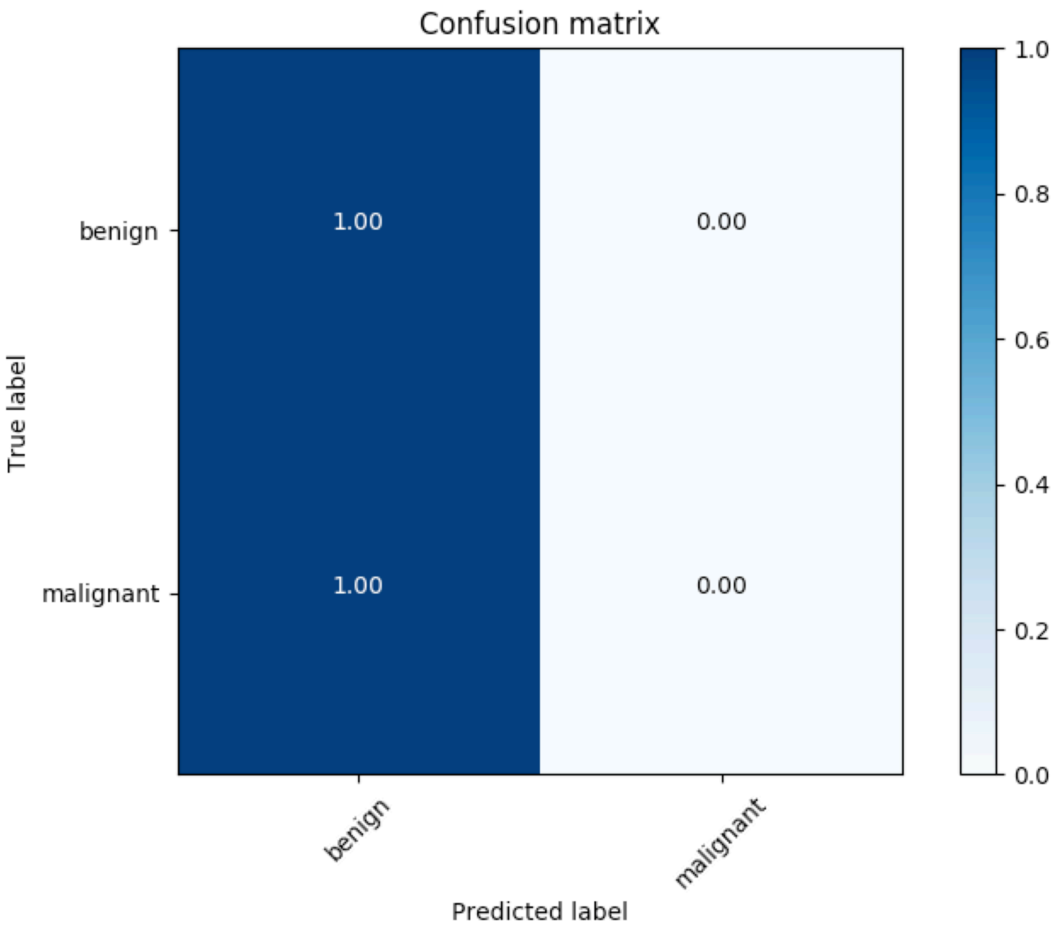
MY file should have exactly 3 columns:

- `Id` - the file names of the test images (in the same order as the sample submission file)
- `task_1` - the model's predicted probability that the image (at the path in Id) depicts melanoma
- `task_2` - the model's predicted probability that the image (at the path in Id) depicts seborrheic keratosis

  Once the CSV file is obtained, I use the get_results.py file to score.

  Corresponding **ROC curves** appear in a pop-up window, along with the **confusion matrix** corresponding to melanoma classification.




Confusion matrix
benign
malignant
True label
1.00
0.00
1.00
0.00
benign
malignant
Predicted label
1.0
0.8
0.6
0.4
0.2
0.0

Project Design

Transfer learning using Inception Resnet V2. My Model on Top of pretrained Resnet.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
global_average_pooling2d_1 ( (None, 1536)              0         
_________________________________________________________________
dropout_1 (Dropout)          (None, 1536)              0         
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              1573888   
_________________________________________________________________
dropout_2 (Dropout)          (None, 1024)              0         
_________________________________________________________________
dense_2 (Dense)              (None, 3)                 3075      
=================================================================
Total params: 1,576,963
Trainable params: 1,576,963
Non-trainable params: 0
```

Image size:
1022 x 767

Using Flask to build REST API on Top. The Request will be a uploaded Image to the Webservice , which will return a probability of skin cancer.

Tools and Libraries used: Python, Jupyter Notebook, pandas, scikit learn,numpy seaborn, matplotlib,tensor flow,Keras, Flask. Other libraries will be added if necessary.

References

A few of the corresponding research papers appear below.

[1] Matsunaga K, Hamada A, Minagawa A, Koga H. **"Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble"**. International Skin Imaging Collaboration (ISIC) 2017 Challenge at the International Symposium on Biomedical Imaging (ISBI).

[2] Daz IG. **"Incorporating the Knowledge of Dermatologists to Convolutional Neural Networks for the Diagnosis of Skin Lesions"**. International Skin Imaging Collaboration (ISIC) 2017 Challenge at the International Symposium on Biomedical Imaging (ISBI). (**github**)

[3] Menegola A, Tavares J, Fornaciali M, Li LT, Avila S, Valle E. **"RECOD Titans at ISIC Challenge 2017"**. International Skin Imaging Collaboration (ISIC) 2017 Challenge at the International Symposium on Biomedical Imaging (ISBI). (**github**)