DeepDerm: Detection of Cancerous Skin Lesions Through Deep Learning

Santosh Murugan, Anna Verwillow | {smurugan, anna19}@stanford.edu

Introduction

- Skin cancer is the most common form of cancer domestically and worldwide, affecting ~20% of Americans by age 70, but is often highly treatable if caught early.
- Large quantities of image data for each of the most common subtypes of both benign and cancerous lesions is available -and in recent years, analysis of such image data has become popular.
- We implement Deep Learning models to perform this task.

Related Work

- The International Skin Imaging Collaboration (ISIC) proposed skin cancer detection challenges in 2016, 2017, and 2018.
  - Number of images available has gone from 900 -> 10k
- Esteva et al., Codella et al., Diaz et al., and Masood et al., represent 4 major benchmarks.
  - Diaz et al. study including dermatologist knowledge
  - Esteva et al. has SOTA results but different task.

Data

- Skin Cancer MNIST dataset, available from Kaggle (2018)
- 10,015 images of skin lesions with seven of the major subcategories (labels)
- One lesion per image, in color (i.e. 3 channels: RGB)
- 10 columns in the dataset, corresponding to the disease class and sub-class, age and sex of patient, localization of the lesion, and the corresponding image name
  - The 'age' column is the only one with NULL / unclean values, to our knowledge.

Features

- We focused on 4 of the 7 available labels:
  - 0 = Melanocytic Nevi (non-cancer)
  - 1 = Melanoma (cancer)
  - 2 = Benign Keratosis-like Lesions (non-cancer)
  - 3 = Basal Cell Carcinoma (cancer)
- Disregarded columns such as 'age' with lots of null values
- The images were resized/reshaped to be:
  - 75 x 100 x 3 for the De Novo model
  - 224 x 224 x 3 for the ResNet50, VGG16, VGG19
  - 299 x 299 x 3 for the InceptionV3 model

Deep Learning Workflow

Data Preprocessing - Before Any Model Selection

- Implemented a data ingestion pipeline which integrates Google Drive (where our data repository resides) with Google CoLaboratory (where our machine learning models are implemented)
- Removed certain columns with troublesome data formats (e.g. NULL values)
- Performed data augmentation techniques to increase the diversity of images in our dataset, using Keras's built-in Image Data Generator module
- Performed a 70:10:20 training-validation-test split of our data to evaluate performance

Training - {De Novo, ResNet50, VGG16/19, InceptionV3}

- Implementation details for each model are presented below:
  - De Novo: Conv2D -> MaxPool -> Flatten -> Dense.
  - ResNet50: 4 output neurons, random weight initialization
  - VGG19: 4 outputs, batch size 512 (computational cost)
  - InceptionV3: 4 output neurons, random weight initialization
- Adam optimizer, 10 epochs, batch size of 32 (except for VGG).

This Led Us To...

Initial and Improved Results Analysis -All Models

- Training and Test errors as well as a metric to understand performance other than average accuracy.
- AUC-ROC (area under the receiver-operating curve) is commonly used in the deep learning community to evaluate model performance, especially in cases where the dataset has some amount of class imbalance.

Table 1: Comparison between Models Run

Algorithm	Training Accuracy	Test Accuracy	AUC
Initial	n/a	0.311	0.597
ResNet	.610	.438	0.541
VGG19	0.592	0.657	0.5
InceptionV3	0.640	0.523	0.443

Results

Initial	0	1	2	3
0	42	31	0	0
1	25	67	0	8
2	139	54	16	6
3	13	3	4	3

ResNet	0	1	2	3
0	27	0	46	0
1	34	0	66	0
2	62	0	153	0
3	10	0	13	0

VGG19	0	1	2	3
0	0	58		
1	0	111		
2				
3				

InceptionV3	0	1	2	3
0	27	0	73	0
1	34	0	100	0
2	62	0	215	0
3	10	0	23	0

Discussion

These models perform multiclass classification where most previously trained models are binary classifiers. While these suffer from both high bias and high variance, the image sourcing and training power required to improve these falls beyond the scope of the resources employed.

Future Work

With substantial additional computational power and time we would:
- train each of these models to a larger number of epochs to correct
- perform a more expansive hyperparameter search
- include additional features and source a larger image set

References

Codella, Noel, et al. "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)." Submitted on 9 Feb 2019 (v1), last revised 29 Mar 2019 (this version, v2). eprint arXiv:1902.03368 [cs.CV]

Tschandl, Philipp, et al. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skinlesions." Scientific Data, Volume 5, Article number: 180161 (2018).