MACHINE LEARNING ENGINEER CAPSTONE PROPOSAL

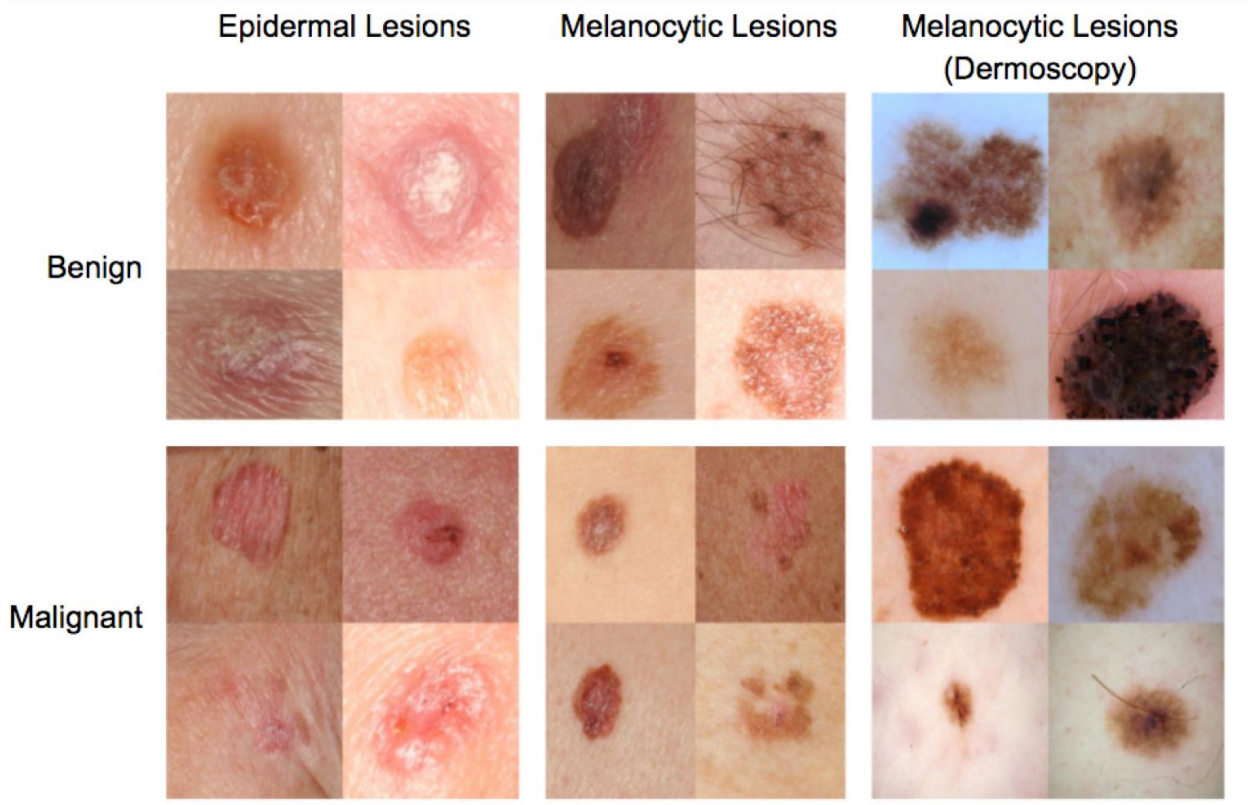
29-11-2020

1. Domain Background

Skin cancer is the most common cancer in the world. In the US there are 5.4 million new cases of skin cancer every year. Different types of skin cancer can be found: Carcinomas, Melanomas (black cancer), etc. Survival chances of patients at the stage IV of the type of cancer is roughly 20%. Thus, early detection is essential for preventing dying from the skin cancer.

2. Problem Statement

Classifying melanomas from clinical images of skin conditions is very hard problem. For example, looking at the following images, it is very difficult to determine if a lesion is benign (above) or malignant (below).



In this project an algorithm will be designed to diagnose melanoma from two types of benign lesions (nevi and seborrheic keratoses).

3. Datasets and Inputs

The data can be downloaded from the below links.

- Training data: https://s3-us-west-1.amazonaws.com/udacity-dlnfd/datasets/skin-cancer/train.zip
- Validation data: https://s3-us-west-1.amazonaws.com/udacity-dlnfd/datasets/skin-cancer/valid.zip

- Test data: https://s3-us-west-1.amazonaws.com/udacity-dlnfd/datasets/skin-cancer/test.zip

Each dataset contains three sub-folders (melanoma/, nevus/, seborrheic_keratosis/ representing images from one of the three image classes.

4. Solution

Convolutional neural networks (CNNs) and transfer learning has been proved to be very efficient for medical images classification problem in general and skin cancer in particular. Indeed, low-level features learned from early layers of a pretrained CNN model can detect simple features like edges, colors, blobs, etc. For specific features to detect skin cancer we then need to finetune the last layers for the CNN model on our skin cancer dataset. s

5. Benchmark Model

In this project different pretrained CNN models in Pytorch (ResNet, AlexNet, GoogLeNet, etc.) will be fine-tuned for classifying the three classes of skin cancer. The performance of the models can be compared to the top scores (from the ISIC competition).

Rank	User	Title	Organization	Documentation	Date	Score
1	RECOD Titans	release (rc36xtrm) "alea jacta est"	RECOD Titans / UNICAMP		Wed, 1 Mar 2017, 11:42:07 pm	0.874
2	Lei Bi	EResNet (single scale w/o attributes)	USYD-BMIT		Wed, 1 Mar 2017, 8:04:42 pm	0.870
3	Kazuhisa Matsunaga	ResNet ensemble with normalized image	Casio and Shinshu University joint team		Wed, 1 Mar 2017, 11:18:03 pm	0.868
4	monty python	gpm-LSSSD	Multimedia Processing Group - Universidad Carlos III de Madrid		Wed, 1 Mar 2017, 12:57:35 pm	0.856
5	T D	Last Minute Submission!!!!	University of Guelph - MLRG		Wed, 1 Mar 2017, 11:55:50 pm	0.836
6	Xulei Yang	multi-task deep learning model for skin lesion segmentation and classification-3	Institute of High Performance Computing + National Skin Center, Singapore		Tue, 28 Feb 2017, 6:34:10 pm	0.830
7	Rafael Sousa	Araguaia Medical Vision Lab - GooglAlexNet	Universidade Federal de Mato Grosso		Wed, 1 Mar 2017, 3:26:22 pm	0.805
8	x j	finalv_L2C1_trir	CVI		Wed, 1 Mar 2017, 11:17:56 am	0.804
9	Cristina Vasconcelos	comb	icuff		Tue, 28 Feb 2017, 1:11:21 am	0.791
10	C V	all	icuff		Tue, 28 Feb 2017, 1:06:44 am	0.789
11	Euijoon Ahn	DeepAhn	USYD-BMIT		Wed, 1 Mar 2017, 10:30:13 am	0.786
12	Balázs Harangi	Ensemble of deep convolutional neural networks	University of Debrecen		Wed, 1 Mar 2017, 8:25:16 pm	0.783
13	Matt Berseth	Final Classification Submission	NLPLOGIX / WISEEYE.AI		Tue, 28 Feb 2017, 6:32:47 am	0.782
14	INESC TECNALIA	Final	INESC TEC Porto / TECNALIA		Wed, 1 Mar 2017, 7:05:40 pm	0.765
15	Dylan Shen	task3_final_RQ	Computer Vision Institute, Shenzhen University		Wed, 1 Mar 2017, 9:20:22 pm	0.759
16	Vic Lee	task3_final_Alice	Computer Vision Institute, Shenzhen University		Wed, 1 Mar 2017, 9:11:31 pm	0.757
17	Masih Mahbod	Skin Lesion Classification Using Hybrid Deep Neural Networks	IPA		Wed, 1 Mar 2017, 12:51:43 pm	0.715
18	Dennis Murphree	Transfer Learning from Inception	Dennis Murphree		Wed, 1 Mar 2017, 11:06:33 pm	0.684
19	Hao Chang	MYBrainAI	Yale		Wed, 1 Mar 2017, 11:53:55 pm	0.636
20	Jaisakthi S.M.	Lesion Classification	SSNMLRG		Wed, 1 Mar 2017, 9:25:02 pm	0.623
21	Wenhao Zhang	testPhase	CSMedical		Wed, 1 Mar 2017, 7:08:07 pm	0.500
22	Wiselin Jiji	Dr Jiji P2 Test	Dr Sivanthi Aditanar College of Engineering		Thu, 2 Mar 2017, 12:46:52 am	0.495
23	Yanzhi Song	submit of yanzhi	song		Wed, 1 Mar 2017, 8:05:13 am	0.475

6. Evaluation Metrics

The model will be evaluated by calculating the area under the receiver operating characteristic curve (ROC AUC) for Melanoma and Melanocytic Classification.

7. Project Design

The project will follow the above steps:

1. Data loading and pre-processing
2. Data visualization
3. Data Augmentation
4. Define and train CNNs
5. Evaluating the networks
6. Conclusions