MACHINE LEARNING ENGINEER CAPSTONE PROPOSAL

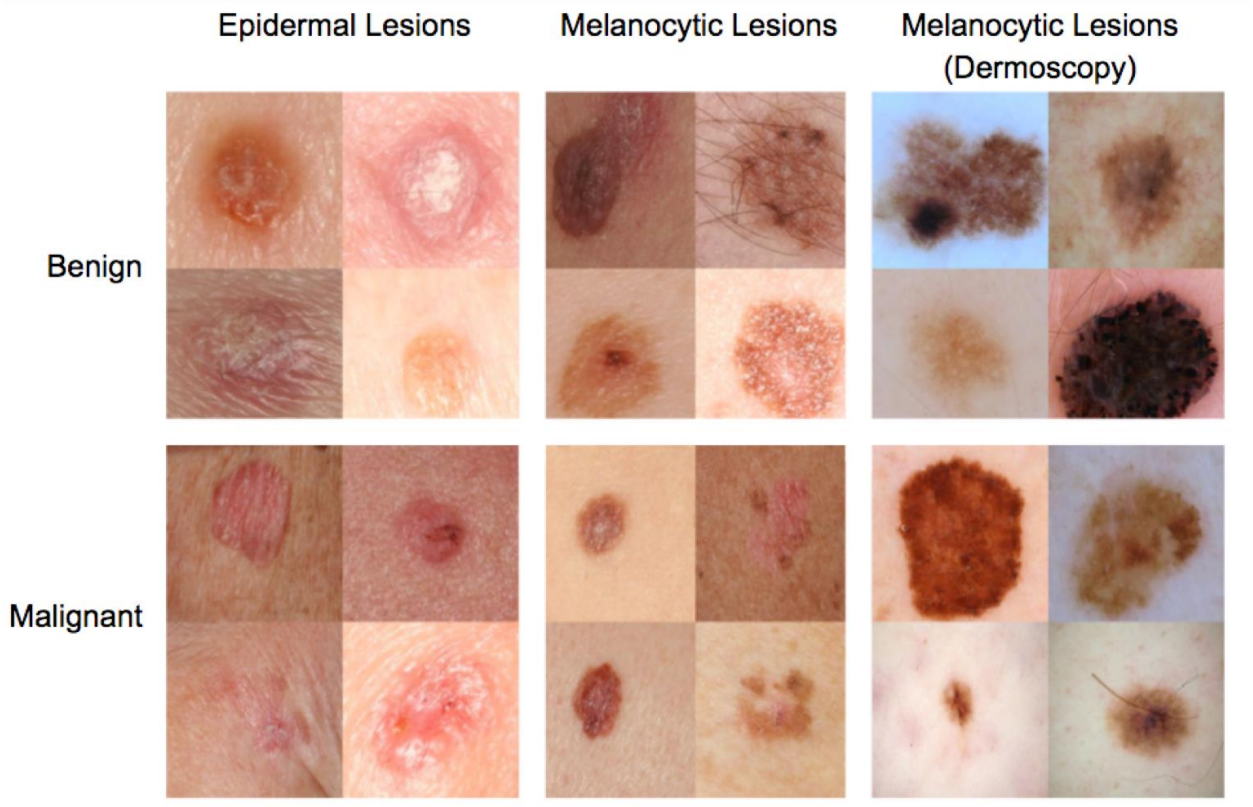
29-11-2020

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 training, validation and test data are received from the 2017 ISIC Challenge on Skin Lesion Analysis Towards Melanoma Detection and 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 representing images from one of the three image classes: melanomas, nevus and seborrheic keratoses. There are 2000, 150 and 600 images on the training, validation and test set respectively. The distribution of classes in each dataset can be seen on the following figures:

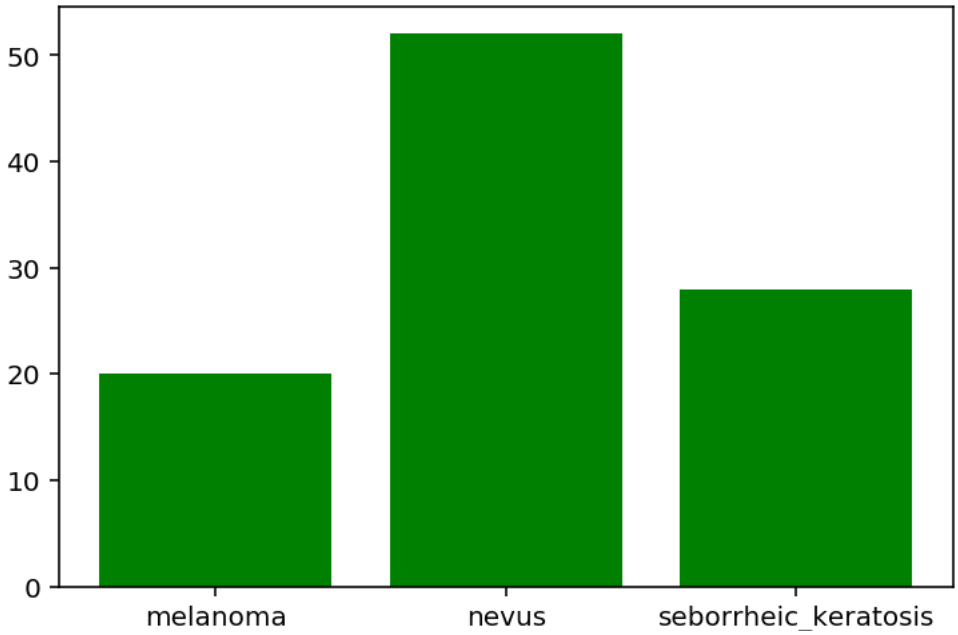


Figure 2: Classes distribution on the validation set. In comparison with the training set there are more seborrheic keratosis cases than melanoma cases.

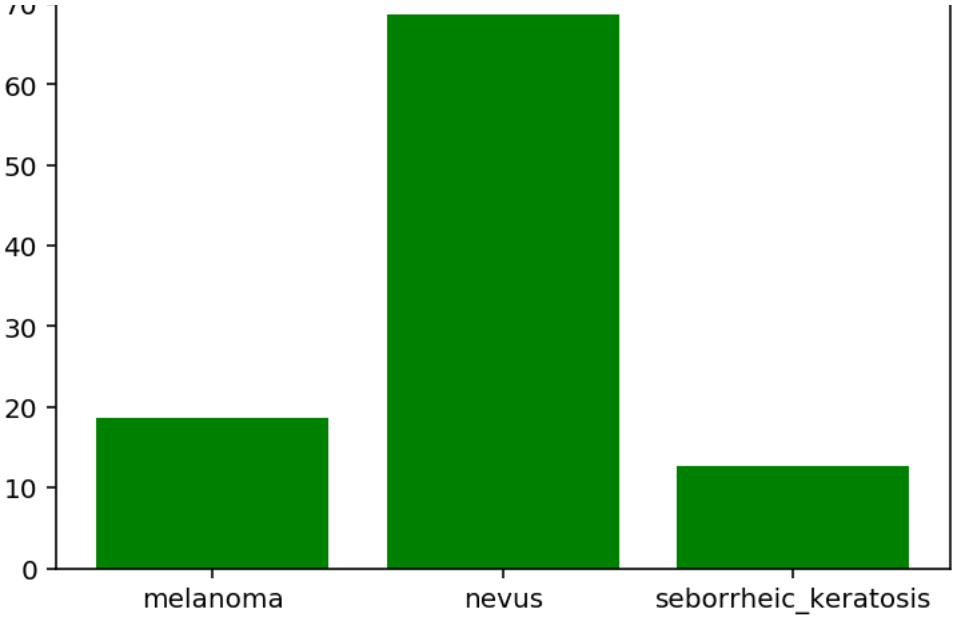


Figure 1: Classes distribution on the training set, most of the cases are nevus (68%), followed by melanomas and then seborrheic_keratosis

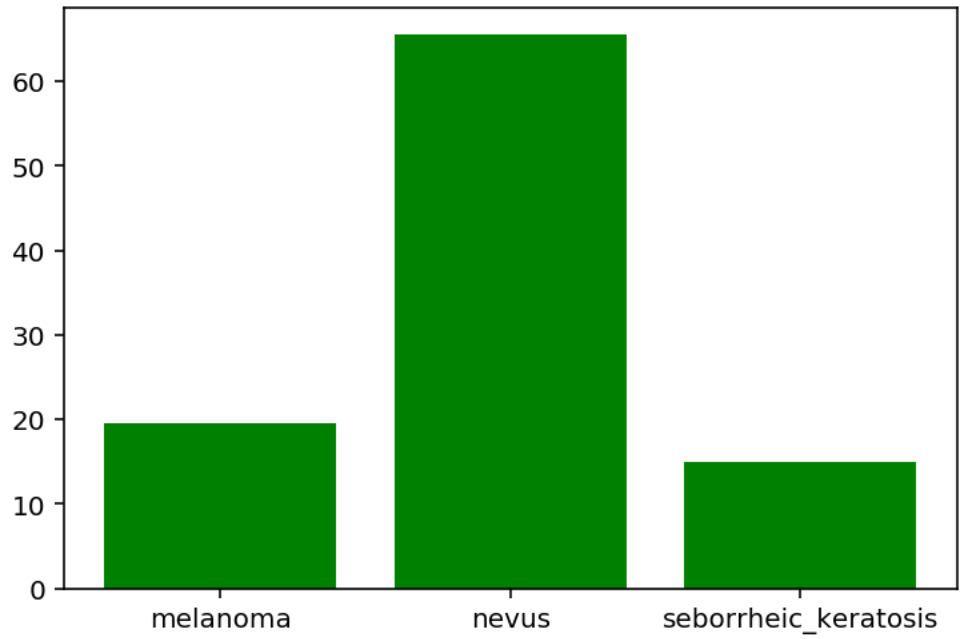


Figure 3: Classes distribution on the test set. The distribution is quite similar to that on the training set.

In general, it seems that we do not really suffer from the classes imbalance challenge, which is very often seen in medical images classification.

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

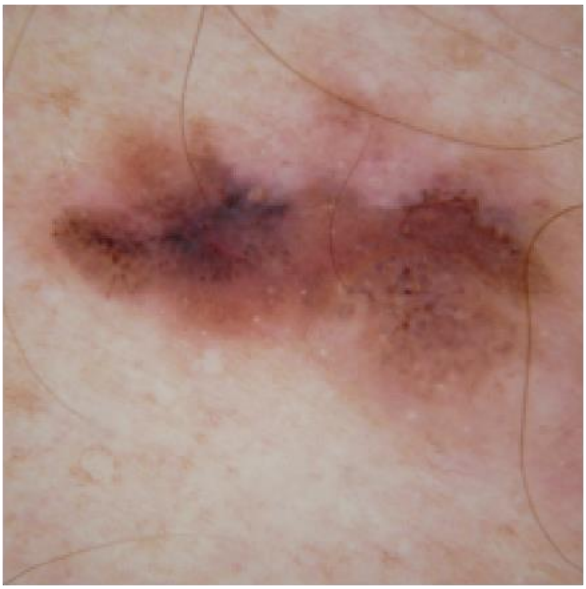
The project will follow the above steps:

1. **Data loading and pre-processing**
   All images are loaded using the Pytorch DataLoader. The images are resized to 255 by 255 and center cropped so that we can easily fed them to a CNN model.

```
# Transform the image size to 255 by 255 and then center crop
transform = transforms.Compose([transforms.Resize(255),
                                transforms.CenterCrop(224),
                                transforms.ToTensor()])

# Load train/valid/test set using ImageFolder from torchvisionc
train_dir = 'data/train'
valid_dir = 'data/valid'
test_dir = 'data/test'
train_set = datasets.ImageFolder(train_dir, transform=transform)
valid_set = datasets.ImageFolder(valid_dir, transform=transform)
test_set = datasets.ImageFolder(test_dir, transform=transform)

# Using Pytorch DataLoader to load the images
train_dataloader = torch.utils.data.DataLoader(train_set, batch_size=32,
shuffle=True)
valid_dataloader = torch.utils.data.DataLoader(valid_set, batch_size=32,
shuffle=True)
test_dataloader = torch.utils.data.DataLoader(test_set, batch_size=32,
shuffle=True)
```

2. **Data visualization:**
   From a DataLoader object we can have a look at a particular image using the iterator.

```
dataloader = torch.utils.data.DataLoader(train_set, batch_size=32,
shuffle=True)

# Run this to test the data loader
images, labels = next(iter(dataloader))
helper.imshow(images[0], normalize=False)
```



3. **Data Augmentation:**
   We have roughly 2000 images for training a CNN model. It is probably a good idea to do some data augmentation to avoid over-fitting. Some data augmentation algorithms we can think about can be: flipping, rotating, zooming, contrasting, lighting, etc. For example, we can define a transform object as follow:

```
transforms = torchvision.transforms.Compose([
    torchvision.transforms.Resize((224,224)),
    torchvision.transforms.ColorJitter(hue=.05, saturation=.05),
    torchvision.transforms.RandomHorizontalFlip(),
    torchvision.transforms.RandomRotation(20, resample=PIL.Image.BILINEAR)
])
```

4. **Define and train CNNs**

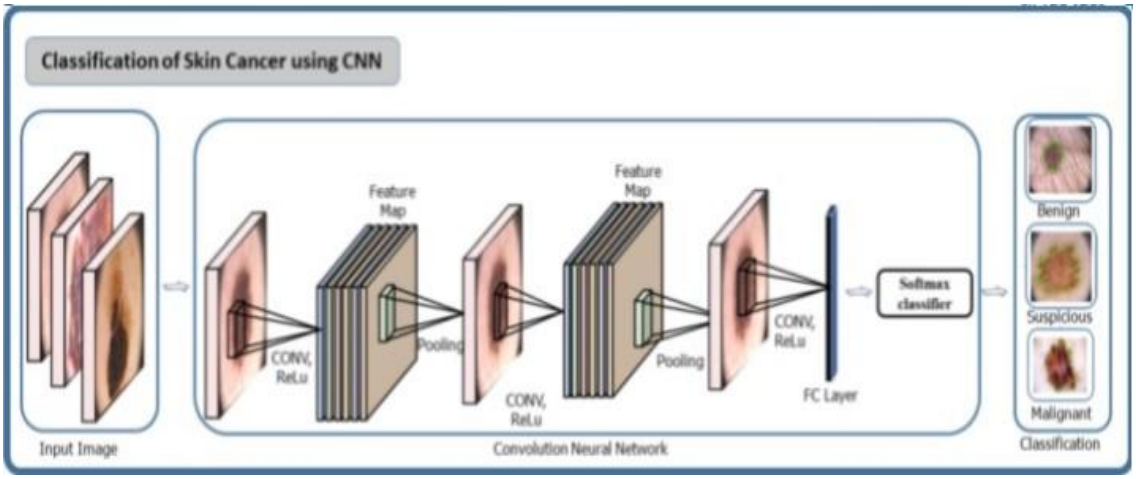


Figure 4 Classification of skin cancer using CNN. Picture from https://www.sciencedirect.com/science/article/pii/S1876034120305633

As described in the solution section different CNN architectures will be trained to classify skin cancer images. However, instead of training the models from scratch we are going to make use of pretrained CNN models in Pytorch (ResNet, AlexNet, GoogLeNet, etc.), which already learned some lower features from earlier layer. Then we will fine-tune the model for classifying the three classes of skin cancer.

5. Evaluating the networks

All models are evaluated using the area under the receiver operating characteristic curve (ROC AUC) for Melanoma and Melanocytic Classification on the validation set. Then the best model will be used on the test set. Then it is interesting to see the final model performance in comparison with the top scores (from the ISIC competition).

6. Conclusions

Based on the performance of different models and the final model, important conclusions can be learned.