***AI-based Skin Disease Diagnosis Model***

This task will build an AI diagnosis model with skin disease prediction for ``buruli ulcer'', ``leprosy'', and ``scabies'', “Mycetoma” and “Yaws”.

First of all, the skin images are resized into the same size 224x224x3, which are sent to the VGG-16 model pretrained on ImageNet dataset to extract the sample features from raw skin images. That is, each image is represented as a 4096-dimensional feature vector for VGG-16. To adapt to the skin diagnosis task, we design the classifier with two-layer FC layers with activation function as Relu from 4096->256->c (where c=5 is the target disease number). The cross-entropy loss is used to define the cost function. For the training stage, we adopt stochastic gradient descent (SGD) with momentum of 0.9 as optimizer to update whole network parameters (i.e., ResNet50 and classifier parameters) and set the batch size to be 10. We implement experiments on PyTorch platform with one GPU (NVIDIA Titan V).

Table 1 shows the data summary of 5 diseases, where each patient may have multiple images recorded.

Table 1. Data Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Yaws | Leprosy | Scabies | Mycetom | BU |
| Patients | 149 | 38 | 107 | 12 | 200 |
| Images | 373 | 131 | 389 | 32 | 784 |

**RESULTS**

We adopt the following metrics to evaluate our model as: top-1 accuracy (%) and Matthew’s correlation coefficient (MCC, 0~1) [ref], which is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all the four confusion matrix categories (true positives, false negatives, true negatives, and false positives).

Given the dataset, we split into training and test set based on different patients into different portions (k%). That is, we use k% patients to train the model and evaluate the rest of the patients. There is no patient overlap between training and test set.

Task 1: Diagnosis Results

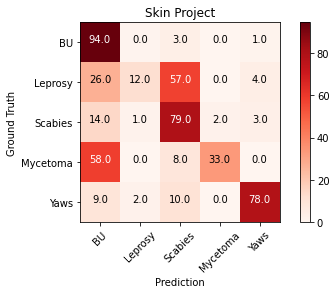
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| k | 30 | 40 | 50 | 60 | 70 |
| Top-1 Acc | 77.66% | 79.00% | 80.33% | 81.22% | 81.67% |
| MCC [R1] | 0.662 | 0.686 | 0.700 | 0.715 | 0.7233 |

***Confusion Matrix***

To further understand our AI diagnosis model for each disease, we report the confusion matrix for 40 and 50 percent’s patients to train our AI and rest to be the evaluation. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa. Each diagonal element denotes the correct diagnosis per disease. From the results, we can notice the BU, Scabies and Yaws show good performance as they have more samples, while other two have very low performance.



40 % Training, 60 % Test

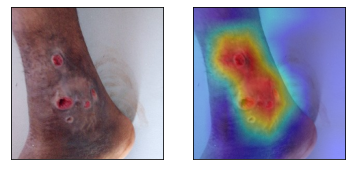
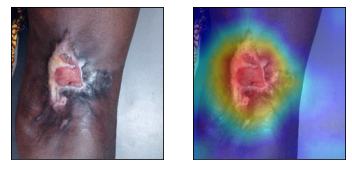


50 % Training, 50 % Test

***Grad-CAM Interpretation***

This section is to explore the interpretable technique to understand how AI makes the diagnosis decision. That is, the AI diagnosis model will also provide visual patterns (disease-inducing diagnosis cues) to make AI diagnosis behavior more trustable and reliable.

***Correct Samples:***

***Incorrect Samples:***