Disease Type Prediction(hackerearth deep learning challenge #2)

Yumin Sun suny@in.tum.de
Zhen Yan zhen.yan@tum.de
Yuchang Zhang ga63jof@mytum.de
Xiaojing Li sommerlxj@gmail.com

Project Proposal

We are doing the challenge presented here. Cited from its introduction page: *Chest X-ray exam is one of the most frequent and cost-effective medical imaging examination. However clinical diagnosis of chest X-ray can be challenging, and sometimes believed to be harder than diagnosis via chest CT imaging. To achieve clinically relevant computer-aided detection and diagnosis (CAD) in real world medical sites on all data settings of chest X-rays is still very difficult, if not impossible when only several thousands of images are employed for study. Started in 1953, National Institute of Health - Clinical Centre is one the leading hospitals in US. They are the active partners in medical discovery. Currently, there are around 1600 clinical research studies in progress at NIH centre, USA. With a support staff of around 620 nurses, in 2016, they handled more than 10,400 new patient*

1. Introduction

Disease type diagnosis from X-rays is of low-cost and simple. However, lack of experienced doctors and high miss misdiagnosed rates makes it a challenge. We are trying to solve this problem using deep learning as well as classical machine learning techniques, with over 10,000 labeled data.

1.1. Related Works

- Learning to Read Chest X-Ray Images from 16000+ Examples Using CNN [1]
- Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning [3]
- Dermatologist-level classification of skin cancer with deep neural networks [2]

2. Dataset

We are using the dataset provided by hackerearth. The training data is split into two parts. One with X-ray pictures and disease labels. This other one includes general information of the patients, i.e., gender and age. There are 14 types of different diseases in total.

- Images: The training image data has information for 18577 patients and testing image data has information for 12386 patients. Each row of data has one X-ray image and its disease label. Each image has size 1024*1024*3 with png format.
- Text in CSV format: Same size as image data. Each row of data has 6 rows, i.e., row id, age, gender, view position, image_name, detected disease.

3. Methodology

We are planning to try different pre-trained models combined with our own self-defined layers. Apart from that, we are also planning to use general patient information as additional inputs. For the pre-trained models, we are going to try alexnet, densenet and resnet101, by not freezing the first fews layers of the parameters for feature extraction, then delete weights feeding for the rest layers, then train the weights with our own datasets. The project here could be out first trial.

4. Outcome

The output of our network will be disease labels. We will try to achieve over 70% accuracy on test data, try to reach a score over 0.5 (Currently the best core is about 0.38).

References

[1] Y. Dong, Y. Pan, J. Zhang, and W. Xu. Learning to read chest x-ray images from 16000+ examples using cnn. In *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017 IEEE/ACM International Conference on*, pages 51–57. IEEE, 2017. 1

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017. 1

[3] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016. 1