

MEDICAL IMAGES FEATURE LEARNING AND ANALYSIS WITH DEEP LEARNING AND RANDOM FOREST

Project proposal of Computational Genomics Spring 2017

Han Den (Andrew Id: hdeng1)

Luyi Ma (Andrew Id: luyim)

Project Idea

- Tuberculosis disease diagnosis can be approached by analyzing clinical symptoms and medical signs. With the help of CT and X-ray images, the process of tuberculosis diagnosis can be quantified. However, before analyzing these images, features for diagnosing tuberculosis must be extracted. Traditionally, it is done manually. In our project, we decide to apply deep learning and random forest methods to approach feature extraction of CT and X-ray images and "diagnose" diseases based on these features. We will use part of our dataset to train our model.

The deep learning network will be worked as a feature extraction layer and the outcoming features will be the characteristics of each pixel in every image. When performing classification, we will use Random Forest method and classify all the test data into two groups. The performance of both feature extraction and classification will be evaluated by the accuracy to classify a test dataset.

- CT and X-ray images of tuberculosis patients are from Belarus tuberculosis portal [1]. It is an open-source database and both CT scan images and chest X-ray images can be found by searching the patient's ID. Clinical records of each patient are also provided. We plan to use these images to construct our data set.

Software Design

- We plan to use Python as our preferred programming language. We will implement algorithms involved in deep learning and random forest in Python. And we also plan to use some deep learning frameworks like Tensorflow to help us in feature extraction. To process data and train the model, we plan to use AWS platform. Finally, we may develop a software to wrap out core scripts implementing our models.

Papers to Read

- **List of papers**

- Deep Learning of Feature Representation with Multiple Instance Learning for Medical Image Analysis [2].
- Random Forest-Based Manifold Learning for Classification of Imaging Data in Dementia [3].

References

- [1] *Open access information resources and the possibility of image-based detection of multiresistant tuberculosis*, <http://obsolete.tuberculosis.by>
- [2] Xu, Yan and Mo, Tao and Feng, Qiwei and Zhong, Peilin and Lai, Maode and Eric, I and Chang, Chao, *Deep learning of feature representation with multiple instance learning for medical image analysis*, Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014: 1626–1630
- [3] Gray, Katherine R and Aljabar, Paul and Heckemann, Rolf A and Hammers, Alexander and Rueckert, Daniel, *Random forest-based manifold learning for classification of imaging data in dementia*, International Workshop on Machine Learning in Medical Imaging, Springer, 2011: 159–166.