Birkbeck, University of London

Pneumonia Detection from Chest X-Ray Images Draft

Author:
Baran Buluttekin

Supervisor:
Dr. George Magoulas

A thesis submitted in fulfillment of the requirements for the degree of MSc Data Science

in the

Department of Computer Science

March 11, 2019

Abstract

This is placeholder text. To add more information type it after this line.

Declaration

I hearby declare this file a text.

```
    import pymongo as pm
    import json
    # line comment
    with open("DSTA/Lab/mongo.json") as f:
    url = json.load(f)
    """ Comment 1"""
    "Comment 2"

    c = pm.MongoClient(url["url"])

    def Myfunc(x):
        print(x)

    print(c.admin)
```

and text goes on.

Contents

1 Introduction

Pneumonia is swelling (inflammation) of the tissue in one or both lungs. It is usually formed at the end of breathing tubes of the lungs and cause these tubes to inflame and fill up with fluid. In the UK, pneumonia effects around 8 in 1000 adults each year [2]. Global economic cost of pneumonia has estimated at $17 billion annually [1]. Currently detecting pneumonia cases heavily relies on chest X-ray image examination which requires expert radiologists to diagnose. Building intelligent system to diagnose the pneumonia can help health care services to increase efficiency and reduce costs and could help increase early diagnoses in countries with inadequate access to health care.

1.1 Related Work

There are number of research has been published about chest X-ray pathology related detection. Most prevalent ones are the CheXNet [3] and ChestX-ray8 [4], both of these research carried out by training on same dataset ChestX-ray8 [4]. ChestX-ray8 comprises of approximately 115,000 frontal view chest X-ray images labelled by extracting information from the accompanied radiologists notes with using variety of different NLP (Natural language processing) techniques.

2 Dataset

Choosing and processing dataset have a crucial importance on success of the any machine learning task. There are several dataset available online that relate to chest X-Ray images. Given the large number of choices for selecting the dataset there are few criteria important to check while deciding the final dataset.

2.1 General Guidelines While Deciding on the Dataset

In this section I have highlighted my reasons for deciding on the dataset of choice for this research project. Main points for decision are:

1. **Reproducibility:** Dataset of choice must allow reader to reproduce the work in order to assess all the points discussed in the report. That would require dataset to be public.

2. **Labelled:** Dataset must contain labels of patients state. Such as being diagnosed with pneumonia or not.

3. **License:** Dataset should have a license that permits for research.

I will be evaluating dataset available while considering general guidelines outlined above.

2.2 ChestX-ray8

This dataset

References

[1] Girish B. Nair and Michael S. Niederman. "Community-Acquired Pneumonia: An Unfinished Battle". In: *Medical Clinics of North America* 95.6 (2011). Pulmonary Diseases, pp. 1143–1161. ISSN: 0025-7125. DOI: https://doi.org/10.1016/j.mcna.2011.08.007. URL: http://www.sciencedirect.com/science/article/pii/S0025712511000927.

[2] *Pneumonia*. URL: https://www.nhs.uk/conditions/pneumonia.

[3] Pranav Rajpurkar et al. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning". In: *CoRR* abs/1711.05225 (2017).

[4] Xiaosong Wang et al. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases". In: *CoRR* abs/1705.02315 (2017). arXiv: 1705.02315. URL: http://arxiv.org/abs/1705.02315.