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 18, 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 [3]. 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 lung diseases related detection. Most prevalent ones are the CheXNet [4] and ChestX-ray8 [5], both of these research carried out by training on same dataset ChestX-ray8 [5]. ChestX-ray8 comprises of approximately 100,000 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 from the openi[2] database. ChestX-ray8 authored by researchers from National Institute of Health (NIH) and published at 2018. Most profound effect of this paper is the creation of the ChestX-ray8 dataset which has become one of the widely used dataset in computer vision research related to lung diseases. More detailed information about the dataset can be found in dataset section of this proposal.

CheXNet is another related article authored by researchers from Stanford University ML group. Prediction of lung diseases achieved by 121 layer convolutional neural network and designed to predict 14 pathologies in the ChestX-ray8 dataset. One of the major importance of this paper is the setting the setting benchmark for human level detection for chest X-ray images. One of the most fundamental difference of the X-ray related disease prediction is the definition of human level accuracy. Due to the nature of required expertise in X-ray images leaves general public out of the scope when it comes to human level performance of these pathologies. Anyone who have not been trained in radiology will not be able to detect any lung diseases in the Chest X-ray images. For example the image below is sample of two chest X-ray images almost indistinguishable to general audience.

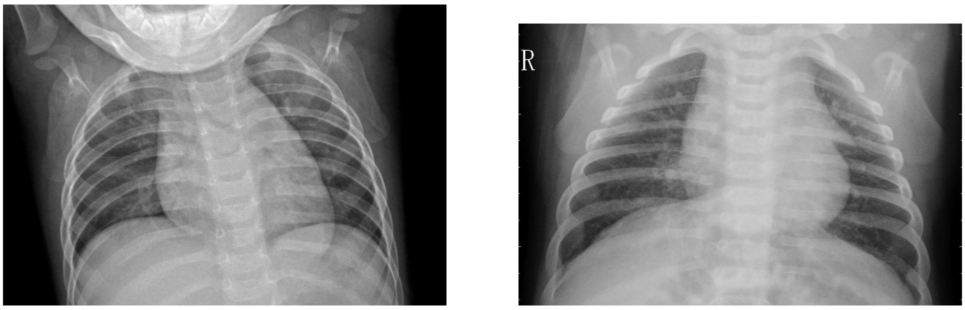

Figure 1: Two sample X-ray Chest images.

Given this challenge authors of the CheXNet conduct a test to establish benchmark for radiologists. They have collected 420 frontal chest X-rays and asked practicing radiologists in Stanford University to label them for all 14 pathologies. Radiologists selected with different range of experience, and had 4, 7, 25 and 28 years of experience. X-ray images presented to radiologists without any patient information or any symptoms experienced by patients and their diagnoses predictions measured based on underlying state of the X-ray patients. Fallowing is the table showing the summary statistic of this test for the 4 radiologists participated to test on F1 score, which is harmonic average of precision and recall.[4]:

	F1 Score (95% CI)
Radiologists 1	0.383 (0.309, 0.453)
Radiologists 2	0.356 (0.282, 0.428)
Radiologists 2	0.365 (0.291, 0.435)
Radiologists 4	0.442 (0.390, 0.492)
Radiologists Avg	0.387 (0.330, 0.442)

Table 1: Radiologist prediction performances from CheXNet.

Importance of this test is that it gives us a rough estimate for human level accuracy benchmark to assess the model performance for new detection models.

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 research.

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

2.2 OpenI Database

2.3 ChestX-ray8

This dataset created part of the ChestX-ray8[5] paper. Original source of this dataset is OpenI[2] medical database. OpenI database is a hospital-size knowledge database that contains X-ray images along with radiologists notes for reference. Radiologists notes often in format of guidance for doctors rather than diagnoses and was not fit for training large scale neural network algorithm that was planned to train in the ChestX-ray8 article.

2.4 Mendeley Research Group

3 Neural Networks

3.1 Convolutional Neural Networks (CNN’s)

3.2 Prominent Computer Vision Architectures

3.3 AlexNet

3.4 VGGNet

3.5 Inception

3.6 YOLO

4 Project Aims and Objectives

5 Tools and Techniques

6 Project Plan

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] *Open Access Biomedical Image Search Engine*. `https://openi.nlm.nih.gov/`. Accessed: 2019-03-12.

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

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

[5] 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`.