Udacity Machine Learning Engineer Nanodegree Capstone Proposal

Ronald Mulumba

July 18, 2019

Identify Pneumothorax Disease in Chest X-rays [1]

Domain Background

A pneumothorax is an abnormal collection of air in the pleural space between the lung and the chest wall [2]. This air pushes on the outside of the lung making it collapse. A pneumothorax can be a complete lung collapse or a collapse of only a portion of the lung [3]. The symptoms of a pneumothorax typically include sudden onset of sharp, one-sided chest pain and shortness of breath. A pneumothorax can be caused by a blunt or penetrating chest injury, certain medical procedures and damage from underlying lung disease. It may also occur for no obvious reason [2][3].

A pneumothorax is generally diagnosed using a chest X-ray [4]. However, they can sometimes be difficult to confirm from chest X-rays[1].

Problem Statement

The problem is a segmentation problem that stems from the SIIM-ACR Pneumothorax Segmentation - Identify Pneumothorax disease in chest x-rays Kaggle competition [1] where it is defined as follows:
"In this competition, you'll develop a model to classify (and if present, segment) pneumothorax from a set of chest radiographic images. If successful, you could aid in the early recognition of pneumothoraces and save lives."

Datasets and Inputs

The datatest, just like the problem stems from the SIIM-ACR Pneumothorax Segmentation - Identify Pneumothorax disease in chest x-rays Kaggle competition [1]. The data is supplied via the Cloud Healthcare API and it contains images in DICOM [5] format and annotations in the form of image IDs and run-length-encoded (RLE) masks. Some of the images contain instances of pneumothorax (collapsed lung), which are indicated by encoded binary masks in the annotations. Images without pneumothorax have a mask value of -1.

The dataset is appropriate to the problem since it contains all the data needed to classify and segment a pneumothoraces in the images.

Solution Statement

The solution will try to predict the existence of pneumothorax in a test image using a convolutional neural network and indicate the extent of the image using binary masks and encode them using RLE.

Benchmark Model

Evaluation Metrics

Project Design

Abbreviations

SIIM - Society for Imaging Informatics in Medicine (SIIM)

ACR - American College of Radiology

DICOM - Digital Imaging and Communication in Medicine

API - Application Programming Interface

RLE - Run-length-encoding

References

[1] SIIM-ACR Pneumothorax Segmentation

Identify Pneumothorax disease in chest x-rays.

https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation

[2] Pneumothorax

https://en.wikipedia.org/wiki/Pneumothorax

[3] Pneumothorax

Symptoms and Causes

https://www.mayoclinic.org/diseases-conditions/pneumothorax/symptoms-causes/syc-2035036 7

[4] Pneumothorax

Diagnosis and Treatment

https://www.mayoclinic.org/diseases-conditions/pneumothorax/diagnosis-treatment/drc-2035 0372

[5] Digital Imaging and Communication in Medicine (DICOM)

https://www.dicomstandard.org/dicomweb/

[6]