Chest X-Ray Images (Pneumonia)

Pegah Khazaei

November 2020

Supervisor: Dr. Hossein Hajiabolhassan

Teaching Assistant: Mr. Khodayari

Course: Deep Learning

Shahid Beheshti University

Department of Mathematics and Computer Science

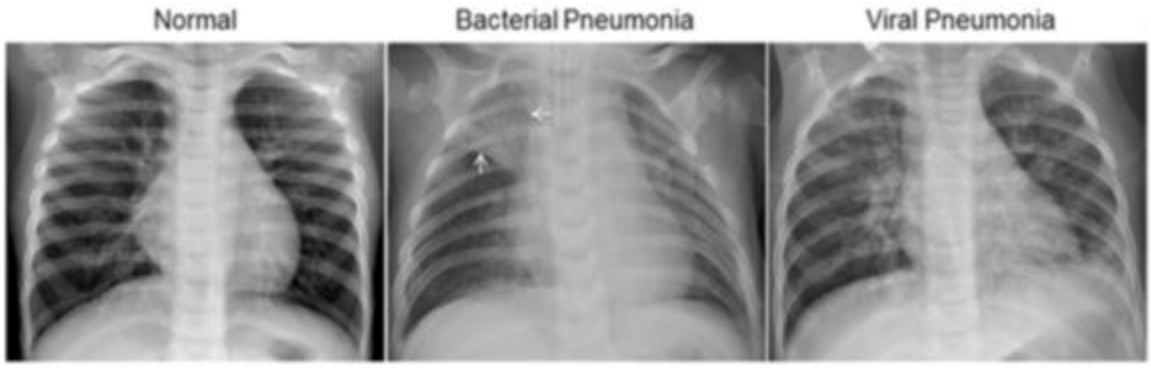
Introduction

According to the World Health Organization (WHO), pneumonia kills about 2 million children under 5 years old every year and is consistently estimated as the single leading cause of childhood mortality, killing more children than HIV/AIDS, malaria, and measles combined. The WHO reports that nearly all cases (95%) of new-onset childhood clinical pneumonia occur in developing countries, particularly in Southeast Asia and Africa. Bacterial and viral pathogens are the two leading causes of pneumonia but require very different forms of management. Bacterial pneumonia requires urgent referral for immediate antibiotic treatment, while viral pneumonia is treated with supportive care. Therefore, accurate and timely diagnosis is imperative. One key element of diagnosis is radiographic data, since chest X-rays are routinely obtained as standard of care and can help differentiate between different types of pneumonia. Pneumonia is diagnosed in many ways, one common way of confirmation is through chest X-rays. Chest X-rays are the best tests, and most accurate, to determine if one has pneumonia. While it is crucial, detecting pneumonia can sometimes be a difficult task. Pneumonia often vaguely shows up in X-rays and can also get mixed in with other diseases present in that area.

Description of the dataset

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal). Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care. For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

The dataset consists of only very few samples and that too unbalanced. The aim of this kernel is to develop a robust deep learning model from scratch on this limited amount of data. We all know that deep learning models are data hungry but if you know how things work, you can build good models even with a limited amount of data. Machine learning has a phenomenal range of applications, including in health and diagnostics. This tutorial will explain the complete pipeline from loading data to predicting results, and it will explain how to build an X-ray image classification model from scratch to predict whether an X-ray scan shows presence of pneumonia. This is especially useful during these current times as COVID-19 is known to cause pneumonia.



The normal chest X-ray (left panel) depicts clear lungs without any areas of abnormal opacification in the image. Bacterial pneumonia (middle) typically exhibits a focal lobar consolidation, in this case in the right upper lobe (white arrows), whereas viral pneumonia (right) manifests with a more diffuse “interstitial” pattern in both lungs.

Post processing of the dataset

we are going to apply CNN on X-Ray images of chest. Dataset is imbalanced (approx. 1:3), images may have different site and can have one or 3 color channels. At start, we will preprocess our data in very simple and intuitive way (load, resize, convert to grayscale, create labels). Model is using grayscale images as for me it did not make too much sense to use X-Ray images as colorful images. We will try to maximize the accuracy in recognizing (above 90%) whether a person is healthy or has pneumonia. To do this, create a pair of neural networks based on the MLP and CNN classes.

References

https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

http://www.cell.com/cell/fulltext/S0092-8674(18)30154-5