Capstone Project - Pneumonia Detection

Background

What is Pneumonia?

Pneumonia is an infection in one or both lungs. Bacteria, viruses, and fungi cause it. The infection causes inflammation in the air sacs in the lungs, which are called alveoli.

Pneumonia accounts for over 15% of all deaths of children under 5 years old internationally. In 2017, 920,000 children under the age of 5 died from the disease.

It requires review of a chest radiograph (CXR) by highly trained specialists and confirmation through clinical history, vital signs and laboratory exams. Pneumonia usually manifests as an area or areas of increased opacity on CXR. However, the diagnosis of pneumonia on CXR is complicated because of a number of other conditions in the lungs such as fluid overload (pulmonary edema), bleeding, volume loss (atelectasis or collapse), lung cancer, or post-radiation or surgical changes. Outside of the lungs, fluid in the pleural space (pleural effusion) also appears as increased opacity on CXR. When available, comparison of CXRs of the patient taken at different points in time and correlation with clinical symptoms and history are helpful in making the diagnosis.

Pneumonia Detection

CXRs are the most commonly performed diagnostic imaging study. A number of factors such as positioning of the patient and depth of inspiration can alter the appearance of the CXR, complicating interpretation further. In addition, clinicians are faced with the challenge of reading high volumes of images every shift. To detect Pneumonia, we need to detect inflammation of the lungs. In this project, you're challenged to build an algorithm to detect a visual signal for pneumonia in medical images. Specifically, your algorithm needs to automatically locate lung opacities on chest radiographs.

Business Domain Value

Automating Pneumonia screening in chest radiographs, providing affected area details through a bounding box.
Assist physicians to make better clinical decisions or even replace human judgement in certain functional areas of healthcare (ex. radiology).

Guided by relevant clinical questions, powerful AI techniques can unlock clinically relevant information hidden in the massive amount of data, which in turn can assist clinical decision making.

Project Description

In this capstone project, the goal is to build a pneumonia detection system, to locate the position of inflammation in an image.

Tissues with sparse material, such as lungs which are full of air, do not absorb the X-rays and appear black in the image. Dense tissues such as bones absorb X-rays and appear white in the image.

While we are theoretically detecting "lung opacities", there are lung opacities that are not pneumonia related.

In the data, some of these are labeled "Not Normal No Lung Opacity". This extra third class indicates that while pneumonia was determined not to be present, there was nonetheless some type of abnormality on the image and oftentimes this finding may mimic the appearance of true pneumonia.

Dicom original images: Medical images are stored in a special format called DICOM files (*.dcm). They contain a combination of header metadata as well as underlying raw image arrays for pixel data.

Details about the data and dataset files are given in below link :

https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data

Project Objectives

The objectives of the project are,
- Learn to how to do build an Object Detection Model
- Use transfer learning to fine-tune a model
- Learn to set the optimizers, loss functions, epochs, learning rate, batch size, check-pointing, early stopping etc.
- Read different research papers of given domain to obtain the knowledge of advanced models for the given problem.

Milestone 1: Pre-Processing, Data Visualization and EDA

- To explore the given data files, classes and images of different classes
- To deal with missing values
- To visualize different classes and analyze to better understand how different classes are visually different

Milestone 2: Detection Model Building – initial draft

- To build (set up) a pneumonia detection model starting from basic CNN
- To train the model and test to understand the performance using right evaluation metrics
- To deal with large training time, saving the weights so that you can use them when training the model for the second time without starting from scratch.

Milestone 3: Detection Model Building – Final Draft

- To try different models to identify the best performing one
- To iterate with different hyper parameters, by trying different optimizers, loss functions, epochs, learning rate, batch size, check-pointing, early stopping etc. for these models to fine-tune them
- To report evaluation metrics for these models along with your observation on how changing different hyper parameters leads to change in the final evaluation metrics

Reference

Acknowledgment for the datasets.
https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview/acknowledgements

Project submissions and Evaluation

While we encourage peer collaboration and contribution, plagiarism, copying the code from other sources or peers will defeat the purpose of learning in this program and working on this project. We expect the highest order of ethical behavior.

Submit the project as per the points below:
1. Each milestone should comprise of a detailed presentable report on the steps taken to complete the milestone. The same is referred to as *Project notes*
2. Share the notebook (code file) showing the steps towards completion of the milestone
3. [Optional] GitHub code source link. Ex. - https://github.com/sri-teja/chemical-NER

Final submission:
- Detailed report with the problem statement, related work, your approach, final insights, recommendations and comparison of results with other models.
- Detailed presentation to be used for the final project presentation session. The slides must be neat and must have a high present-ability with a defined structure and flow walking through the problem at hand, the approach and the solution
- Final submission notebook. Please make sure the notebooks are well commented and detailed.
- A 5-min video explaining the final submission, covering on problem, approach and key takeaways (final model performance and other findings)
- [Optional] GitHub code source link for the final submission

Refer to the timeline document in the capstone course to understand the timeline of the submissions of the required documents.

Passing criteria: You must receive a minimum of 60% on each milestone to complete the project.
>60 % Points = Complete
>80 % Points = Excellent

Detailed rubrics are provided in the 'Capstone Project: Guidelines and Grading Criteria' section of the capstone course.

Project Support

You can clarify your queries related to project during the weekly mentor sessions.

Happy Learning!