AIML Online Capstone - Pneumonia Detection Challenge

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 your 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 time points and correlation with clinical symptoms and history are helpful in making the diagnosis.

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 reading high volumes of images every shift.

Pneumonia Detection

Now to detection 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 bounding box.

Assist physicians to make better clinical decisions or even replace human judgement in certain functional areas of healthcare (eg, 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.

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

Pre-Processing, Data Visualization, EDA
- Exploring the given Data files, classes and images of different classes.
- Dealing with missing values
- Visualization of different classes
- Analysis from the visualization of different classes.

Model Building
- Building a pneumonia detection model starting from basic CNN and then improving upon it.
- Train the model
- To deal with large training time, save the weights so that you can use them when training the model for the second time without starting from scratch.

Test the Model, Fine-tuning and Repeat

- Test the model and report as per evaluation metrics
- Try different models
- Set different hyper parameters, by trying different optimizers, loss functions, epochs, learning rate, batch size, checkpointing, early stopping etc. for these models to fine-tune them
- Report evaluation metrics for these models along with your observation on how changing different hyper parameters leads to change in the final evaluation metric.

Project Objectives

The objective of the project is,
- 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, checkpointing, early stopping etc.
- Read different research papers of given domain to obtain the knowledge of advanced models for the given problem.

Project submissions and Evaluation Criteria

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

Submit the project as given below.
- Report with Problem Statement, Related Work, Your Approach and comparison of results with other models written in PDF.
- Github link where the model is hosted.

Evaluation Criteria: You must receive a minimum of 60% on each milestone to complete the project.

>60 % Points = Complete
>80 % Points = Excellent

Reference

Acknowledgment for the datasets.

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

Project Support

You can clarify your queries related to project during the weekly Review sessions.
If you are interested in writing a research paper for the same then please refer to the sample paper shared.

Happy Learning!