**REPORT ON IMAGE CLASSIFICATION WITH DEEP LEARNING OF PNEUMONIA AND NORMAL CHEST X-RAYS**

**PROBLEM STATEMENT**

Tibabu hospital has observed an increase in Pneumonia infections. The hospital is understaffed, and cannot keep up with the high volume of patients. They approached Avengers LTD for a solution to help with the diagnosis. So to improve diagnosis of pneumonia using x-rays, tasked Avengers to come up with a solution.

Our task is to come up with a classification algorithm that will determine if a patient has pneumonia or not based on chest X-ray Image

**Metrics of Success**

The algorithm will be considered a success once it makes predictions with an accuracy of 90-96% and Recall above 95%

**Data Understanding**

The medical dataset contains a set of x-ray images of pediatric patients. The images are of patients who have pneumonia and those who don't.

Chest X-ray images were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou, China. All chest X-ray imaging were 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.

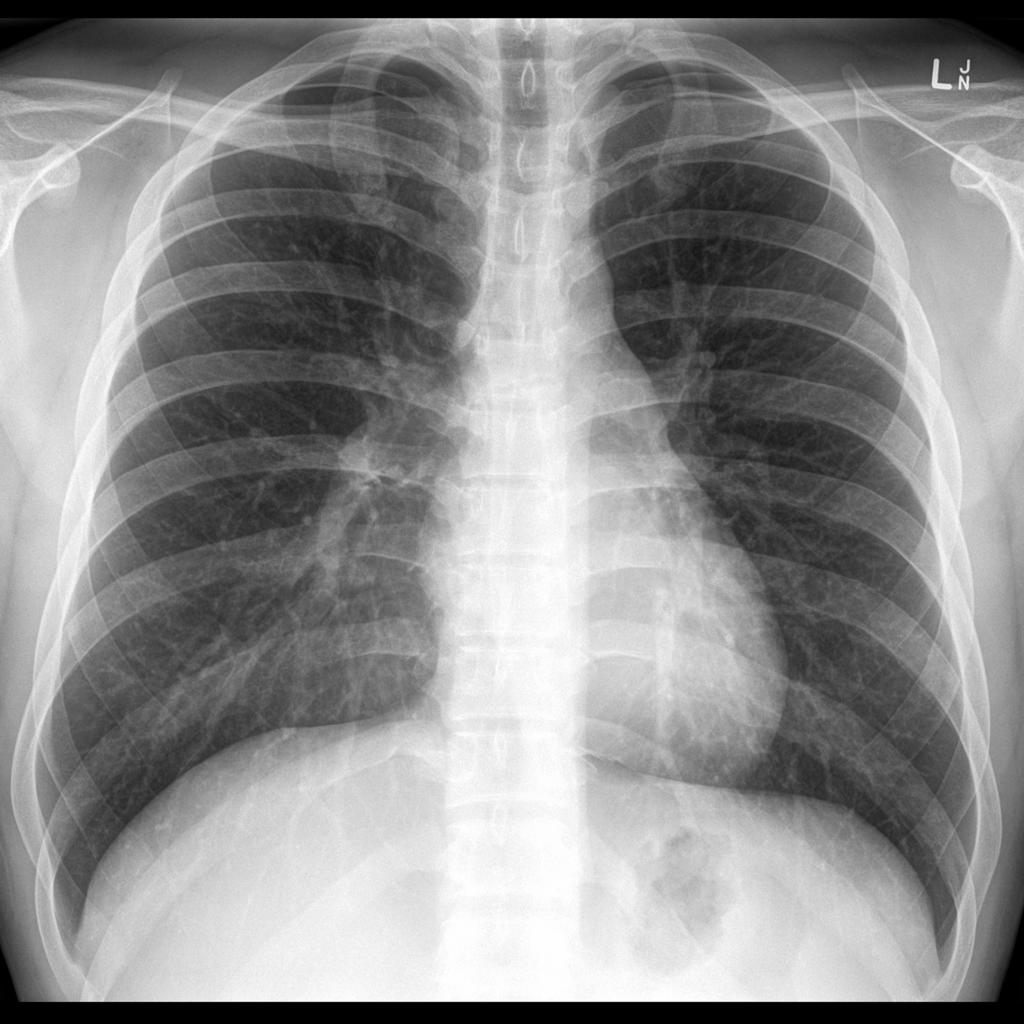
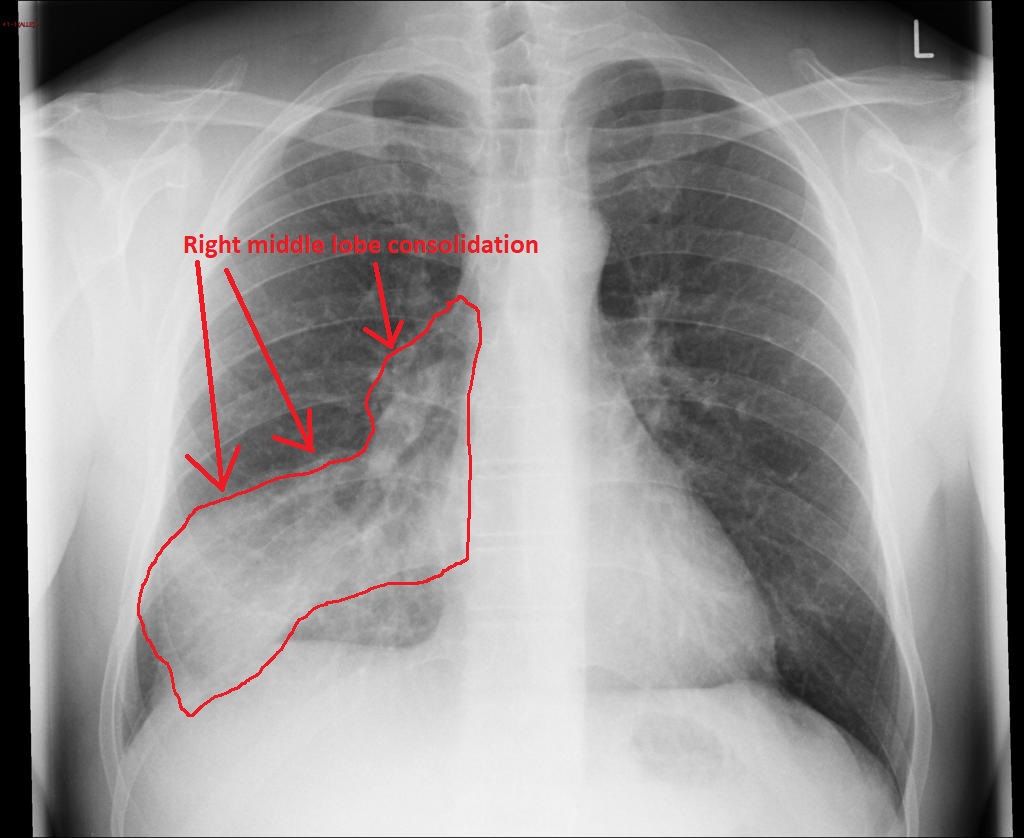
The final data is in two classes, Normal and Pneumonia. This dataset was downloaded from [Kaggle](https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia). The data was further divided into three folders of train, test and validation, each with two classes; normal and pneumonia.

**UNDERSTANDING CONTEXT**

Pneumonia is an acute disease of the lungs that can be caused by bacteria, viruses and fungi. When an individual is infected with pneumonia, the alveoli (air sacs) are filled with fluid and pus causing painful breathing, difficulties in breathing and limits oxygen intake (reduced oxygen saturation).

In 2019, pneumonia accounted for 14% of all deaths among children under 5 years, [according to WHO](https://www.who.int/news-room/fact-sheets/detail/pneumonia). It is prevalent in children and the elderly.

Pneumonia is diagnosed based on a patient’s clinical symptoms, imaging and lab work. The preferred imaging is a plain chest radiograph (X-ray). On a chest X-ray, pneumonia is seen as opacities (white areas on the lung fields) which represent lung consolidation. While some consider X-rays to be the gold standard of diagnosis, they are prone to misinterpretation depending on the quality of the image, experience of healthcare professionals etc. Accurate diagnosis of the disease is crucial in reducing disease burden, hospitalizations and mortality.



Chest Xray with pneumonia Normal Chest X-ray

**Data Validity**

Original data is from [Mendeley](https://data.mendeley.com/datasets/rscbjbr9sj/3) , it contains validated chest X-ray images of patients split into training and test sets. Images are labeled as (disease)-(randomized patient ID)-(image number by this patient)

**Data Uniformity**

All the images are stored in jpeg format.

**Data Consistency**

Data has no duplicates

**DATA PREPARATION**

Importation of necessary packages to use in the development of the classification algorithm

Images of the chest x-ray were imported and reshaped to (156\* 156) pixels.

Data was normalized to take values between 0 and 255 to make sure that each pixel value is between 0 and 1

Generated labels that correctly identify the images as one with pneumonia and one that is normal.

Matplotlib was used to visualize the distribution of the data and observed that the majority(3875) of the chest x-ray images had pneumonia, and 1341 normal chest X-rays. Obvious class imbalance

The data is set to proceed to the modeling phase.

**MODELING**

Modelling involved the following steps:

1. A convolutional baseline model was created
2. Data was augmented to deal with class imbalance
3. Iterative convolutional models were created using the augmented data. In this step, regularization techniques were also applied to deal with overfitting.
4. Finally transfer learning was done using VGG19’

**RECOMMENDATIONS**

The sensitivity of detecting pneumonia by emergency medicine specialists and radiologists[according to this paper](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6377225/#:~:text=Sensitivity%20of%20plain%20chest%20radiography,radiation%20(9%2C%2010).) is 83%, therefore:

1. Given that the model has 98% sensitivity, interpretation of chest x-rays using this Convolutional Algorithm might help in improving the diagnostic accuracy of pneumonia.
2. The algorithm is fast hence will increase efficiency considering the current understaffing in the Hospital.

**CONCLUSIONS**

Re-evaluate the model to improve the recall score on unseen data.