# RADIOLOGY GUIDE: IMAGE CLASSIFIER FOR PNEUMONIA DETECTION DATA REPORT.

## Business Understanding:

## Research Question:

1. Which deep learning model architecture achieves the best performance in terms of accuracy, sensitivity, specificity in detecting pneumonia from chest X-ray images?
2. How can data augmentation techniques be employed to improve the generalizability and robustness of the deep learning model for pneumonia detection on unseen data?
3. Can transfer learning from pretrained models such as ResNet50v2 improve pneumonia detection performance?
4. How do pneumonia detection models perform compared to human radiologists, and what are the implications for clinical practice?

## Problem Statement:

* Pneumonia is a leading cause of respiratory illness to humanity. It poses a significant threat to vulnerable populations, particularly young children and older adults. Therefore, early and accurate diagnosis is essential for successful treatment and improved patient outcomes. Current methods for diagnosing pneumonia, primarily relying on chest X-ray interpretation by radiologists, have limitations that include time constraints that lead to delays in treatment, potential for subjectivity leading to misdiagnosis, and high dependence on specialized expertise not always readily available. A recent article by the National Library of Medicine reveals that the average accuracy of human radiologists in detecting pneumonia cases is about 60%, [**Link here**](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8506182/)
* To curb the increasing fatalities from Pneumonia, the need for faster, more objective, and accessible diagnostic tools for pneumonia detection is rapidly growing, especially with the continuous advancement in technology.
* With the above in mind, we explore the potential of deep learning to address these challenges by developing an automated system for pneumonia identification from chest X-ray images. Deep learning has revolutionized medical image analysis due to its ability to learn complex patterns from large datasets. The ability to extract intricate patterns from vast datasets, have transformed the field of medical image analysis by offering a unique opportunity to automate such analyses. We shall therefore harness this approach and capabilities to automate pneumonia detection with an objective to improve efficiency, accuracy, and accessibility of diagnosis.

## Objectives:

* Main Objective:
* Specific Objectives:

## Data Understanding:

* Data Source
* Data Description

## Data Preparation:

* Loading the data
* Cleaning the data

## Feature Engineering:

## External Data Source Validation:

## Exploratory Data Analysis (EDA):

## Modelling:

## Conclusions:

## Recommendation:

## Future Improvement Ideas:

## Deployment:

**## Conclusions and Recommendations:**

For our next steps we would be able to augment our image data to create more images to train our model on. Given more time we may, through more tweaking of the hyper parameters, be able to find a more successful model. Our final step is to get the model and the portable X-Ray machines into the hands of people who need them.

This analysis leads to the following conclusions:

The Convolutional Neural Network (CNN) model performs the best in image classification; specifically, when the data has been augmented.

The model is 40.3% accurate when testing and classifying chest X-ray images for pneumonia.

**## Limitations**

The performance of CNN models is greatly improved with the use of more data (images) in the training process. Although the available data used for this project is limited, an attempt to generate more training data by creating augmented images from the given image collection may provide sufficient to achieve improved results.

In an effort to reduce training times, only a limited number of models were attempted. Running additional models with greater levels of complexity & a greater number of hyperparameters adjusted, may have eventually yielded a more precise model.

**### Recommendations**

While the results from the models are positive, there can be further improvements to get the most accurate diagnoses using deep learning.

Accuracy:

Optimization algorithms can be employed to refine the existing deep learning models

Deployment:

Focus on creating a system that functions effectively with limited computational resources.

Validation:

Validating the developed model through clinical trials is needed to ensure its reliability for diagnoses.

Outline Approach

Data Acquisition and Preprocessing - This involves use of representative data that contains both healthy and pneumonia-infected cases. We annotate the images with labels indicating the presence or absence of pneumonia. We also enhance image quality and consistency by resizing, normalization/standardizations of values to a common range, facilitating better training for the deep learning model and reduction of noise to improve the model's ability to identify the true underlying structures in the X-ray.

Model Development - This involves selection of a suitable deep learning architecture such as Convolutional Neural Network(CNN) that has had good traction in medical image analysis. We then train the model on the preprocessed dataset, splitting it into training, vaidation, and tests sets. We also address class imbalance, if any, by data augmentation to artificially expand the training data then monitor the training process by adjusting parameters for optimal performance.

Model Evaluation - In the model evaluation phase, we assess the performance of the trained models using key metrics such as accuracy and confusion matrices. Through the interpretation of these metrics, we gain insights into the models' efficacy in distinguishing between normal and pneumonia cases in X-ray images. Additionally, visualization techniques are employed to dissect the models' predictions, providing valuable insights into their strengths and weaknesses. This comprehensive analysis allows for informed decisions regarding model refinement and optimization, ultimately ensuring the development of robust classifiers for accurate binary image classification.

Interpretation and Deployment - We explore techniques for explaining the model's decision-making process for instance, saliency maps, class activation maps to gain insights into how it differentiates between healthy and pneumonia-infected X-rays. If the model achieves satisfactory performance, consider deploying it in a controlled clinical environment for further validation and potential real-world use cases.