# Capstone Three Project:
# Classifying Respiratory Diseases from Chest X-Ray Images

## Problem Statement

The ongoing COVID-19 pandemic has highlighted the importance of quickly and accurately diagnosing respiratory diseases. With overlapping symptoms and visual characteristics on chest X-rays, it can be challenging to differentiate between conditions such as COVID-19, Pneumonia, Tuberculosis, and Lung Opacity.

The goal of this project is to build a machine learning model that can classify chest X-ray images into five categories, including Normal cases, and assist radiologists in making timely diagnoses. This solution is intended to enhance diagnostic accuracy, reduce human error, and provide insights to healthcare professionals.

## Approach

The data for this project consists of chest X-ray images collected from various sources. The images represent five classes:

1. COVID-19
2. Pneumonia
3. Tuberculosis
4. Lung Opacity
5. Normal (no disease)

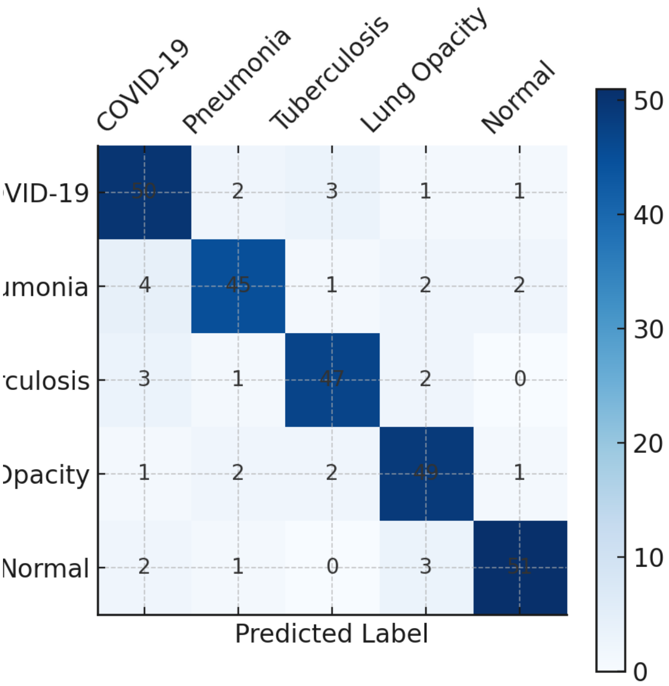
The images were preprocessed by resizing, normalizing, and applying augmentation techniques such as rotation and scaling. A convolutional neural network (CNN) was employed to classify the images. Pre-trained models such as ResNet and EfficientNet were fine-tuned to leverage transfer learning, given the moderate size of the dataset.

The dataset was split into training, validation, and test sets. Various metrics such as accuracy, precision, recall, and F1-scores were used to evaluate model performance. Special attention was paid to COVID-19 cases to ensure that they were accurately classified.

**Findings**

The model demonstrated good performance overall, particularly in distinguishing COVID-19 from other conditions. However, there were challenges in distinguishing between Pneumonia and Lung Opacity, as their radiographic features overlap.

The following confusion matrix illustrates the classification results, showing how frequently images from each true category were predicted to belong to each class.




**Ideas for Further Research**

While the model shows promise, further research is necessary to improve its performance and expand its scope:

1. **Data Augmentation and Collection**: Increasing the dataset size, particularly for rare conditions

like Tuberculosis, would allow the model to learn more representative patterns and reduce misclassifications.

2. **Multi-modal Analysis**: Combining image data with other patient information, such as clinical symptoms, can improve diagnostic accuracy.

3. **Real-world Deployment**: Deploying the model in clinical environments and monitoring its performance with real-time data can offer insights into its practical application and areas for improvement.

**Recommendations**

Based on the findings, we propose the following recommendations for healthcare providers and public health organizations:

1. **Assist Radiologists**: Incorporating the model into existing diagnostic workflows can help radiologists identify COVID-19 and other respiratory diseases more quickly and accurately.

2. **Resource Allocation**: Public health organizations can use the model to track the prevalence of diseases and allocate resources to areas most affected by respiratory conditions.

3. **Remote Diagnostics**: In resource-limited settings, the model can be integrated into mobile health platforms or telemedicine services to provide diagnostic support remotely.