**Data Collection and Preprocessing**

**Project Name**: COVID-19 Chest X-Ray Image Classification

The dataset of chest X-ray images was prepared for deep learning classification to detect COVID-19, Bacterial Pneumonia, and Normal cases using a Convolutional Neural Network (CNN) with transfer learning (VGG16)

**PREPROCESSING STEPS :**

| SECTION | DESCRIPTION |
| --- | --- |
| Data Overview | Collected chest X-ray dataset organized into train and test folders, each containing subfolders for classes (Normal, COVID-19, Bacteria). |
| Image Resizing | Resized all images to 64×64 pixels for uniform input into the CNN model. |
| Rescaling | Applied pixel normalization by rescaling image values from 0–255 → 0–1. |
| Data Augmentation | Used ImageDataGenerator to apply transformations and split 20% of training data into validation automatically. |
| Train / Validation split | Training set = 80% of train folder, Validation set = 20% of train folder (via subset in ImageDataGenerator). |
| Test set | Loaded separately from the test directory without augmentation, only rescaling. |
| Label Encoding | Automatically handled by flow\_from\_directory, mapping class subfolders to categorical one-hot labels. |
| Model Preparation | Used VGG16 pretrained on ImageNet as the base model (without top layers). Added custom layers: Flatten → Dense → Dropout → Output. |
| Visualization | Generated plots comparing training accuracy/loss vs validation accuracy/loss across epochs. |

**Data Preprocessing Code Snapshots :**

| **SECTION** | **CODE** |
| --- | --- |
| Data Overview |  |
| Image Resizing |  |
| Train / Validation split |  |
| Test set |  |
| Model Preparation |  |
| Visualization |  |