# **Data Quality Report**

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

This Data Quality Report summarizes potential issues in the chest X-ray dataset and provides resolution strategies.

**Data Quality Issues :**

| Data Source | Data Quality Issues | Severity | Resolution Plan |
| --- | --- | --- | --- |
| Chest X-Ray Dataset (Image Folders) | Class imbalance (unequal number of COVID-19, Normal, and Bacterial images) | High | Apply data augmentation and possibly use class weights during training. |
| Chest X-Ray Dataset (Image Folders) | Variations in image quality (contrast, brightness, noise) | Moderate | Normalize pixel values (rescale 0–1), apply augmentation, and use pretrained CNN (VGG16) to improve robustness. |
| Chest X-Ray Dataset (Image Folders) | Different image dimensions across raw dataset | Low | Resize all images to fixed size (64×64) for uniform model input. |
| Chest X-Ray Dataset (Image Folders) | Limited dataset size may affect generalization | High | Use transfer learning (VGG16 with ImageNet weights), data augmentation, and validation split to reduce overfitting. |