**Part 2 - Notes**

**Explore the label distribution and qualitatively describe the data by plotting some examples for both labels. (1 Pt)**

The dataset is structured into three main folders: train, val and test. Within each of these folders, there are subfolders representing the two categories of X-Ray images: Pneumonia and Normal.

In total, there are 5,863 X-Ray images in JPEG format. Their distribution is shown in table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Train | Validation | Test | Total |
| Normal |  |  |  |  |
| Pneumonia |  |  |  |  |

**Do you see visual differences between healthy and disease samples? (1 Pt)**

The normal chest X-ray typically shows clear lungs without any areas of abnormal opacification. However, in cases of bacterial pneumonia, distinct white areas may appear on the image, indicating focal lobar consolidation. These differences are easily noticeable, providing a clear visual indicator of the presence of disease, as illustrated in Figure 1.

On the other hand, distinguishing between healthy and diseased samples is often more challenging in cases of viral pneumonia. This is because viral pneumonia is often manifested with a more diffuse ‘‘interstitial’’ pattern in both lungs, which can closely resemble the appearance of a healthy chest X-ray. Less pronounced visual differences between healthy and disease samples are depicted in Figure 2.

However, in most cases it is very difficult for someone who is not an expert in the field of medicine to see visual differences between healthy and disease samples.