

DATA COLLECTION

The dataset of this work has been collected from Kaggle repository, which contains Chest X-Ray scans of Covid-19 affected, normal and pneumonia. This collected dataset is not meant to claim the diagnostic ability of any Deep Learning model but to research about various possible ways of efficiently detecting Coronavirus infections using computer vision techniques. The collected dataset consists of 6432 total chest X-ray images. This data set is further divided into training (i.e., 5467) and validation (i.e., 965) set of normal, covid, and pneumonia. In the training set, 1345 is normal, 490 are covid, and 3632 is pneumonia. In the validation phase, 238 samples of a normal case, 86 covid, and 641 of pneumonia were considered for this analysis. At the time of the drafting of this paper, we had 576 PA (Posteroanterior) View scans of Covid-19 affected patients. The scans were scaled down 128×128 to aid the fast training of our model. The PA view scans were deemed to be consistent with our covid dataset. Table 1 displays the data distribution for training and testing the data. Our previous section described that Deep Learning models performance better with larger datasets. In our coverage of Medical Image Analysis, we looked at Transfusion from Raghu et al. which shows that out-of-domain data like ImageNet has little benefit for medical imaging tasks. An issue with constructing these datasets is privacy. Imagining the role of Deep Learning in precision, tailor-made medicine and diagnostics, we would expect performance to improve by looking a massive collection of patients' EHRs, genomes, blood testing results, family history, etc. However, most patients would not feel comfortable revealing such intimate data to a potentially hackable centralized database. We note that data privacy is especially problematic when the patient can be identified from the data. This is not much of an issue with chest radiographs alone, but can be problematic with metadata associated with it, or stored in electronic health records.