Data set 1 - Chest X -Ray Images (Pneumonia)

Image count - 5863 jpeg
With 2:1 where 2 is normal and 1 is patient with pneumonia
It is pre divided in test, train and val
Colour mode is grayscale because x-ray images
Pneumonia class includes 2 subtypes - bacterial and viral
The age range is pediatric focused

Challenges

The validation set size was extremely small
Metadata is limited as no age gender or clinical history is mentioned
Its binary classification so it may miss clinical complexity
Potential label noise

Summary

This dataset represents a real-world screening scenario where automated systems could assist radiologists in prioritizing pneumonia cases, especially in resource-limited settings.

Data set 2 - Skin Cancer MNIST

Image count - 10015
The image type is dermatoscopic skin lesion images
There are 7 classes, 7 lesion types with icd10 codes
Csv provided had 70% train , 15% validation and 15% test
Colour mode is rgb

Challenges

There is class imbalance, melanoma had very few samples as compared to benign nevi
Many images contain hair, ruler marks, air bubbles etc
Images are collected from different institutions so equipment are varied
Some classes are visually similar

Summary

This dataset enables computer-aided diagnosis (CAD) systems for early melanoma detection, which is crucial for survival rates.