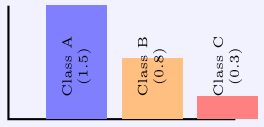
# Medical Imaging: Exploratory Data Analysis & Preprocessing

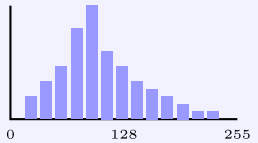
## Exploratory Data Analysis

### Plot Class Distribution

Detect class imbalance
(60% - 33% - 7%)




### Visualize Pixel Intensity




Examine distribution
before normalization

### Inspect Sample Images

Sample Image

Check for
artifacts,
alignment,
orientation
Look for anomalies and
track patient relationships

Workflow Direction

## Preprocessing Pipeline

### Image Normalization

$$I_{norm} = \frac{I - \mu}{\sigma}$$

Apply consistently across
train/validation/test sets

### Valid Augmentations

| Rotation | Crop | Brightness |
|---|---|---|
| ±15° | 90% | ±10% |

Use only medically valid
transformations

### Apply Augmentations

| Training | Validation | Test |
|---|---|---|
| | | |

Only apply augmentations
to training data

| Training (70%) |
|---|
| Validation (15%) |
| Test (15%) |

Split at patient level
not image level

## Documentation for Reproducibility

### Save Code

- Preprocessing scripts
- Configuration files
- Environment details

Version control everything

### Document Parameters

- Normalization method
- Augmentation settings
- Random seeds

Record all decisions

### Document Process

- Processing order
- Patient-image mapping
- Exclusion criteria

Enable full reproducibility

## Summary: Do's and Don'ts

| | |
|---|---|
| Plot class distribution to detect imbalance | Don't apply random left-right flips unless medically valid |
| Visualize pixel intensity histograms before normalization | Don't normalize before inspecting image characteristics |
| Apply medically-valid augmentations (rotation, crop, brightness) | Don't augment validation or test datasets |
| Normalize images consistently across train/val/test | Don't lose track of image-patient relationships |