# DATASET OVERVIEW

## Type of Images

Clinical skin images used for classification of dermatological conditions. Coloured pictures.
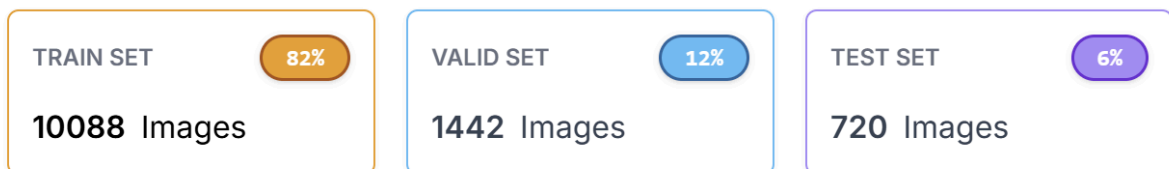
## Dataset Structure

Number of images  - 7206
Avg image size - 0.27mp
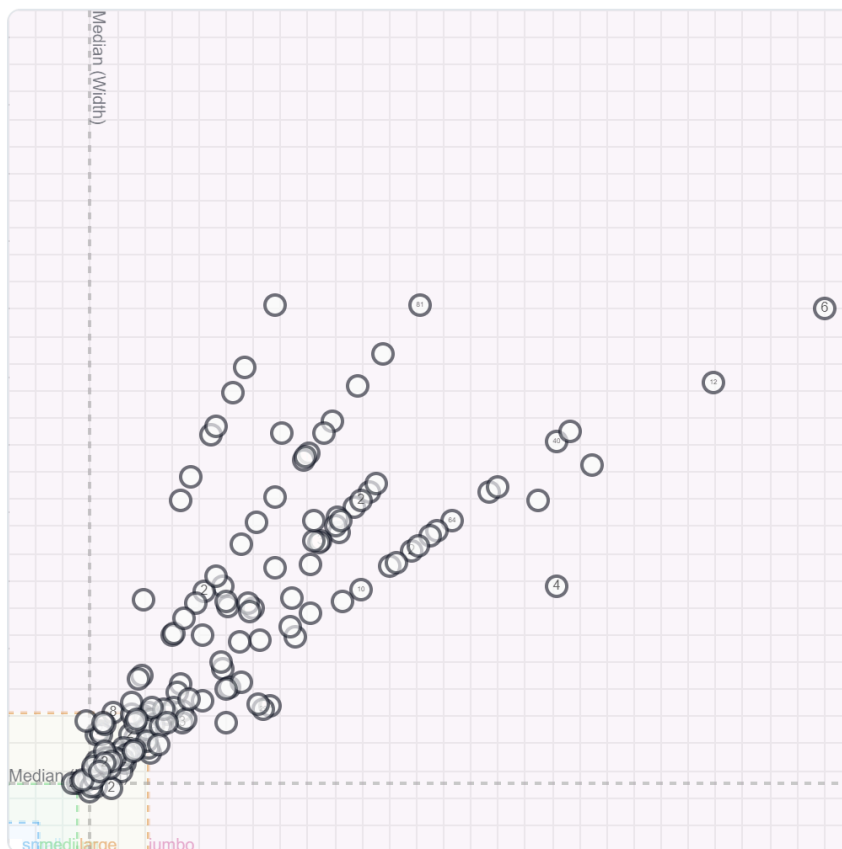There is a split  into 2 classes - melanoma and basal cell carcinoma
It is also split into Training set, Validation set, Test set

### Dataset Split
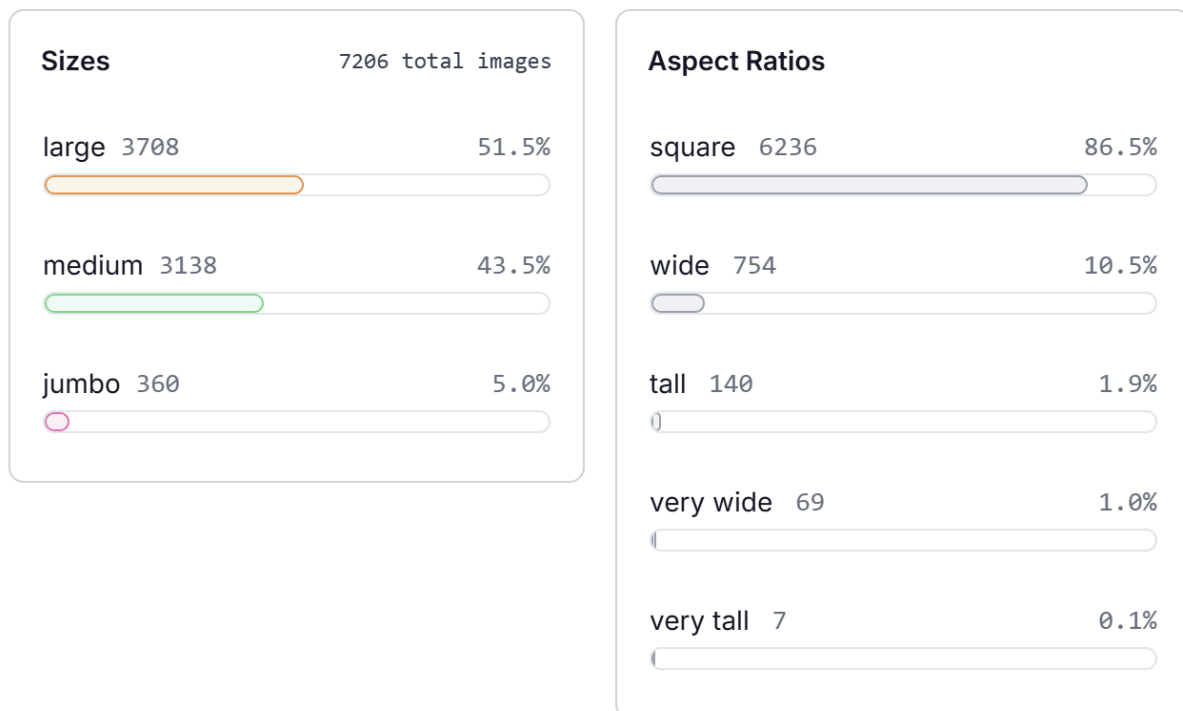
| TRAIN SET | 82% | VALID SET | 12% | TEST SET | 6% |
|---|---|---|---|---|---|
| **10088** Images | | **1442** Images | | **720** Images | |

Dimension insights

The dashed lines represent the **median width** (600 px) and **median height** (512

| Sizes | 7206 total images |
|---|---|
| large 3708 | 51.5% |
| medium 3138 | 43.5% |
| jumbo 360 | 5.0% |

| Aspect Ratios | |
|---|---|
| square 6236 | 86.5% |
| wide 754 | 10.5% |
| tall 140 | 1.9% |
| very wide 69 | 1.0% |
| very tall 7 | 0.1% |

## Classes

The dataset contains multiple skin condition categories, including visually similar lesion types.

## Challenges

Class imbalance (some categories have fewer samples)
Variations in lighting, resolution, and skin tone
High visual similarity between certain classes