AI for Healthcare: Task 3

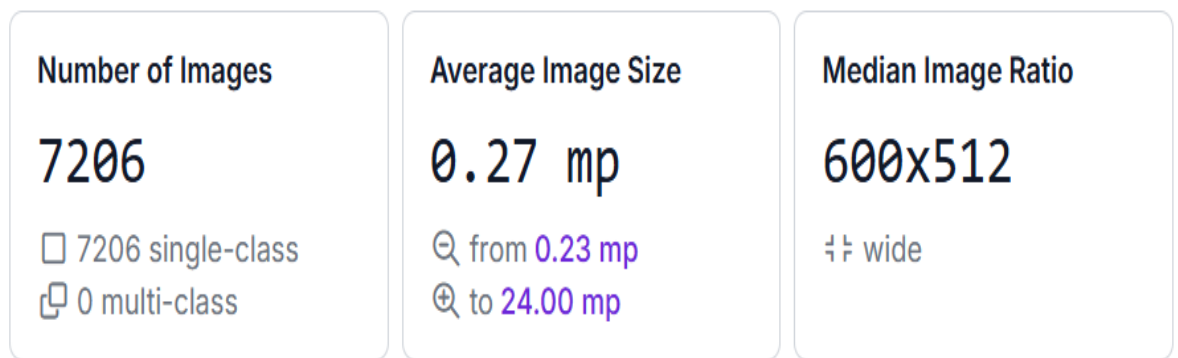## Analysis of CXR Dataset: Skin

- *Types of Images:*
  medical skin images showing different types of skin conditions or lesions. 2D colour photos of skin areas with different diseases and conditions.

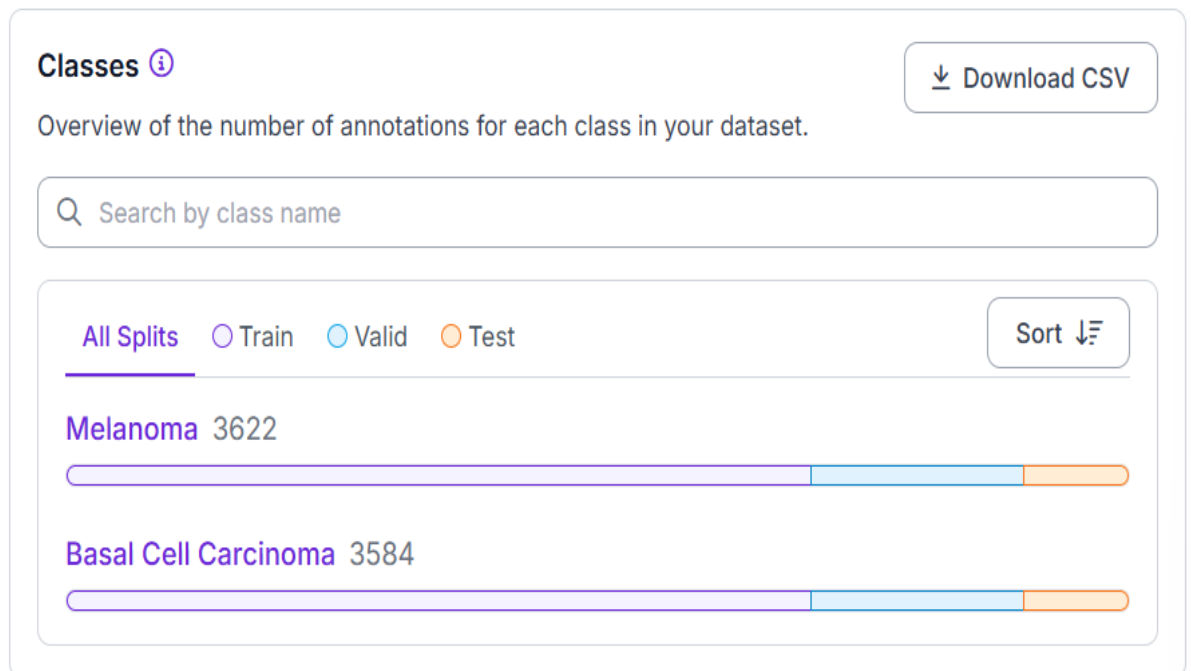- *Number of Samples, Classes and Splits:*
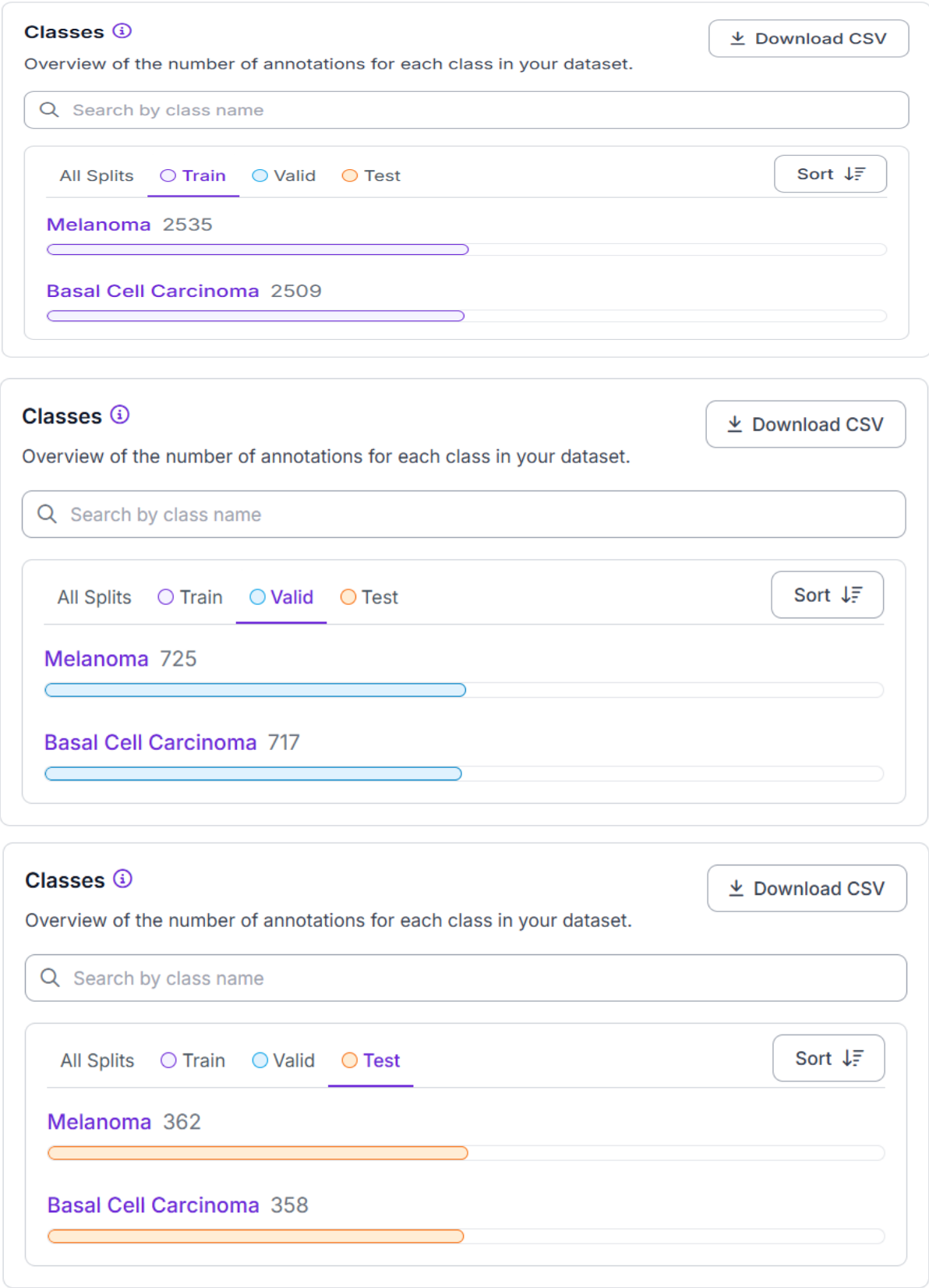 As per the analytics on Roboflow, the following is the data:
  The total number of images in the dataset is 7206.

These are split into two classes as per the image below:

The dataset is split into training, Validation, and testing sets as follows:

Classes ⓘ   ⭳ Download CSV

Overview of the number of annotations for each class in your dataset.

🔍 Search by class name

| All Splits   ○ Train   ○ Valid   ○ Test | Sort ↓⁼ |

Melanoma 2535

Basal Cell Carcinoma 2509

Classes ⓘ   ⭳ Download CSV

Overview of the number of annotations for each class in your dataset.

🔍 Search by class name

| All Splits   ○ Train   ○ Valid   ○ Test | Sort ↓⁼ |

Melanoma 725

Basal Cell Carcinoma 717

Classes ⓘ   ⭳ Download CSV

Overview of the number of annotations for each class in your dataset.

🔍 Search by class name

| All Splits   ○ Train   ○ Valid   ○ Test | Sort ↓⁼ |

Melanoma 362

Basal Cell Carcinoma 358

That is, there are 5044 images in the training split, 1442 in the validation split, and 720 images in the test split.

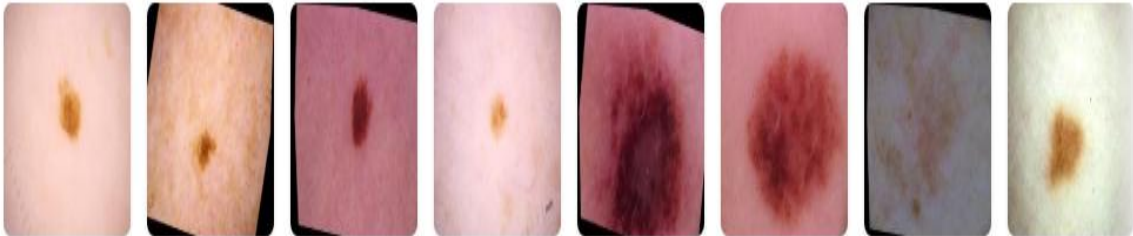But viewing the dataset section shows this thing, which you can see if you download the dataset locally:

v1 **2025-04-12 12:05pm**
Generated on Apr 12, 2025

⬇ Download Dataset

12250 Total Images

View All Images →

Dataset Split

TRAIN SET	82%	VALID SET	12%	TEST SET	6%
10088 Images		**1442** Images		**720** Images	

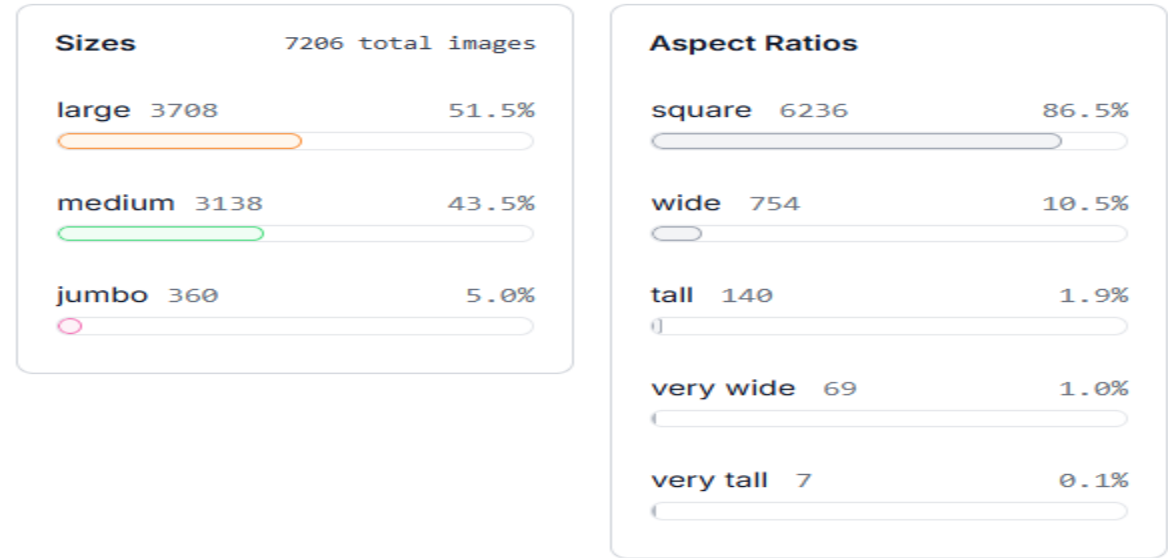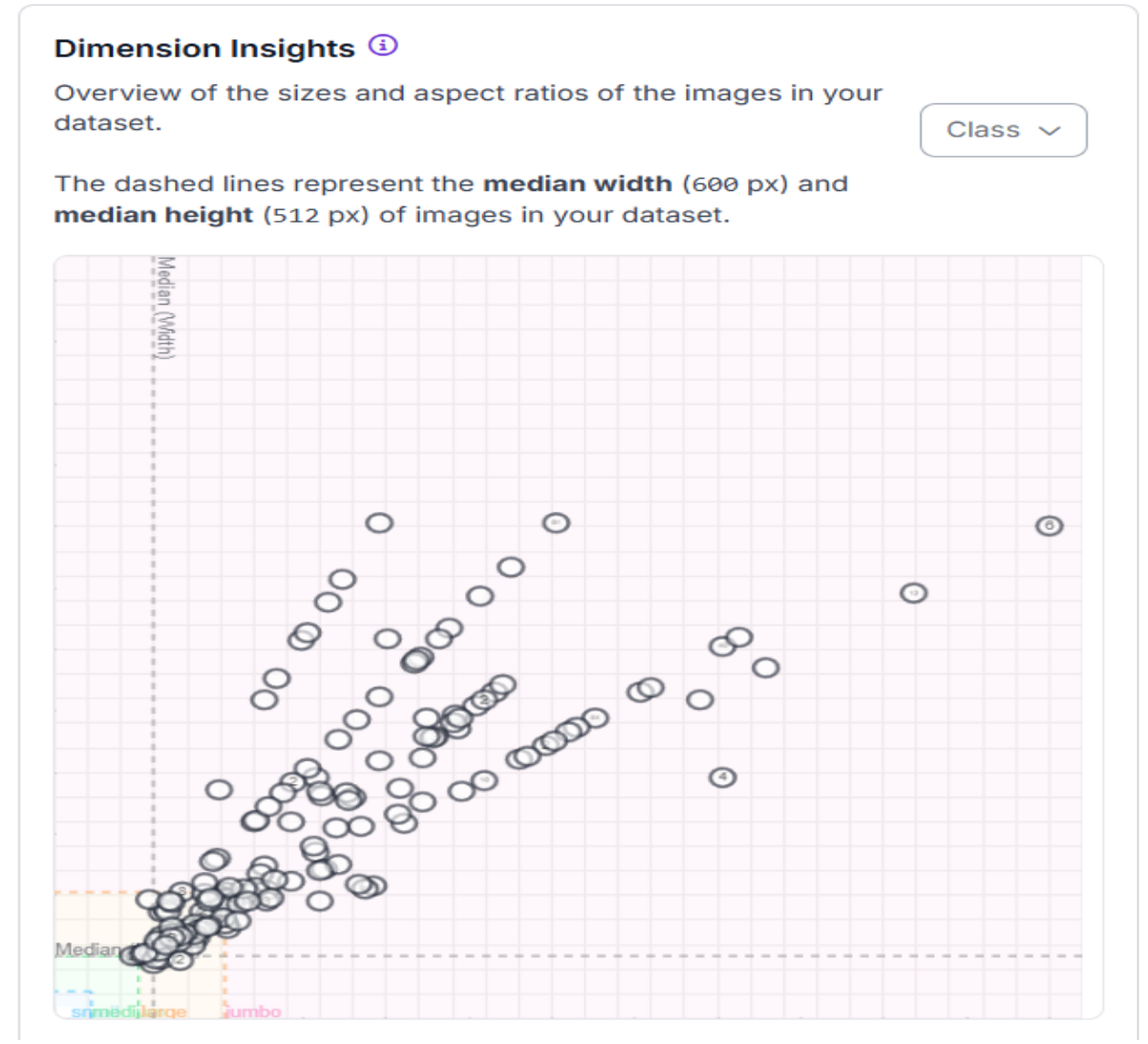**Preprocessing**
Auto-Orient: Applied
Resize: Stretch to 640x640

Augmentations
Outputs per training example: 2
Rotation: Between -15° and +15°
Shear: ±10° Horizontal, ±10° Vertical

- *Dimension insights:*

Dimension Insights ⓘ

Overview of the sizes and aspect ratios of the images in your dataset.

Class ⌄

The dashed lines represent the **median width** (600 px) and **median height** (512 px) of images in your dataset.

Sizes	7206 total images		Aspect Ratios	
large 3708	51.5%		square 6236	86.5%
medium 3138	43.5%		wide 754	10.5%
jumbo 360	5.0%		tall 140	1.9%
			very wide 69	1.0%
			very tall 7	0.1%

According to the split pattern, there are comparable images of each class in each split, and in total, there are also. Hence, there may be a lower chance of class imbalance.