

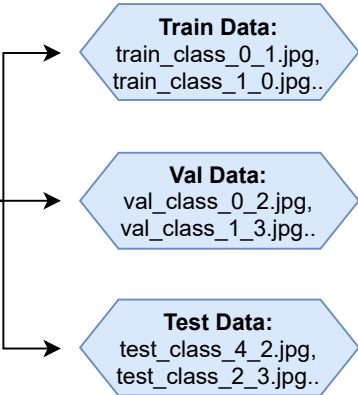
<b>1.</b> <b>Data Acquisition</b> Data is downloaded and unzipped in cwd	<b>2.</b> <b>Clean Data</b> Reduce to 28790 images. 8434, 8069,578, 3903, 7806	<b>3.</b> <b>Data Preprocessing: Resize Images</b> All images are resized to 256 by 256 pixels	<b>4.</b> <b>Data Preprocessing: Data Augmentation</b> Augment Data for Class 2 and 3 to balance the dataset
---	---	---	---

Download **galaxy-zoo-the-galaxy-challenge.zip**  
File size: 1.9 GB



**Pegasus**  
Downloads the data from an online repository (1.9 GB)

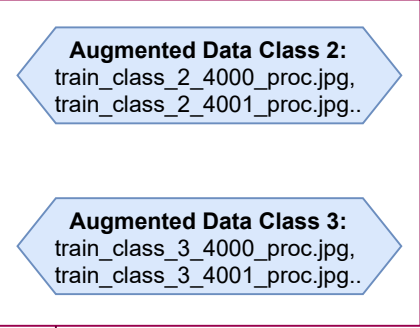
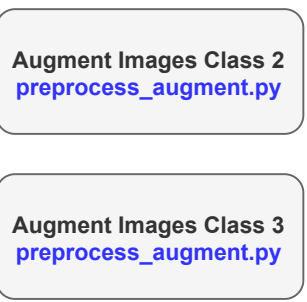
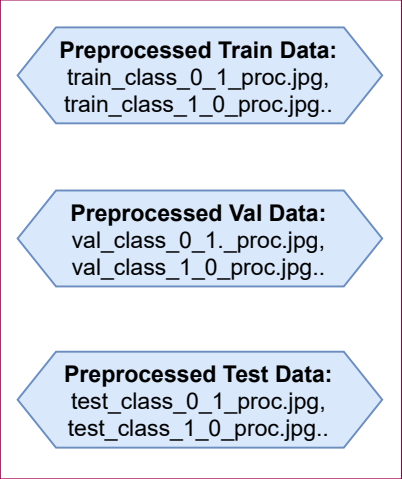
The images are assigned to classes based on their properties. Subset of the images is used as a dataset going forward. Total 28,793 images each between 8 kB to 16 kB. **Total: 386.0 MB**



**Pegasus**  
launches multiple data preprocessing jobs in parallel.



The data preprocessing jobs can be parallized based on number of available nodes. **Total Input: 386.0 MB**



**Decaf**  
Decaf potentially could transfers *in-memory* the preprocessed images to the HPO job..

**Decaf**  
Decaf potentially could transfers *in-memory* the preprocessed images to the HPO job..

<b>5.</b> <b>Hyperparameter Optimization</b> Tune model to find the best setting (Optimizer, Activation)	<b>6.</b> <b>Train Model</b> Use the best optimization parameters to train model	<b>7.</b> <b>Test Model</b> classification metrics, confusion matrix, visualize what network learned
---	---	---

**Pegasus / Decaf**

