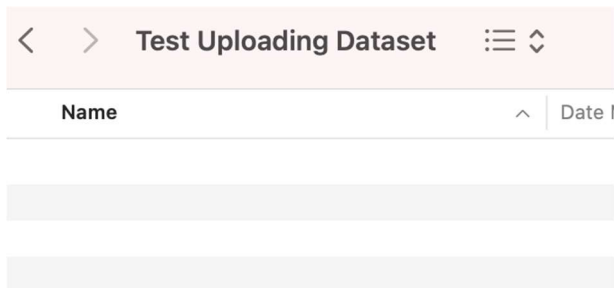


Guideline for Dataset Preparation

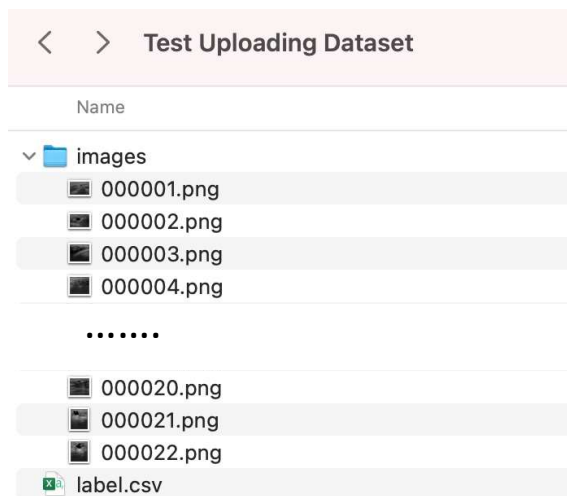
The MIDA labeling system was developed to crop images, draw tumor boundaries, and label the BI-RADS descriptors for B-mode breast ultrasound images. You can upload your dataset, label the images, and export all annotations. The dataset you upload should follow a specific organization. Please follow the following steps to prepare your dataset.

1. Create an empty folder



All the following steps will take place under this empty folder.

2. Put all ultrasound images under a subfolder: 'images', and a *label.csv* file under the root folder

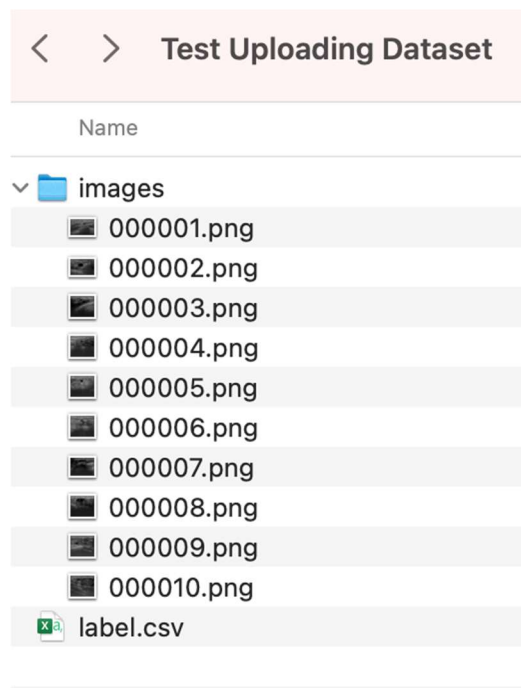


2.1 Prepare images

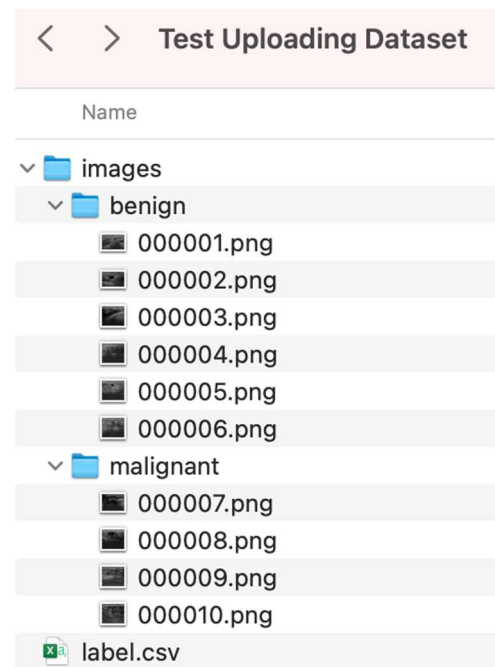
Images should be saved in .jpg or .png format. Other formats may not be compatible. It's acceptable to 1) keep all images under a single folder or 2) use subfolders to organize images.

There is no specific rule applied to the names of images and folders.

For example:



1) single folder



2) subfolders

Step 2.2: Prepare a label.csv file

The label.csv file records tumor labels and diagnosis information of all images on the dataset. At least three columns are required as below.

- 1) case_id: a positive integer number that denotes the identification of a case
- 2) image: a relative path to an image
- 3) tumor_type: one of the following choices: N(normal), B(benign), M(malignant), O(others), and U(unknown)

| | A | B | C |
|---|---------|-------------------|------------|
| 1 | case_id | image | tumor_type |
| 2 | 1 | images/000001.png | B |
| 3 | 8 | images/000010.png | M |
| 4 | 8 | images/000011.png | M |
| 5 | 9 | images/000012.png | M |
| 6 | 10 | images/000013.png | M |

The above 3 columns must be the first 3 columns in the label.csv file, and their names and order should not be changed.

It is allowed to add additional columns to include extra information; and there is no requirement for the names and order of the extra columns. The extra columns must be placed after the first three essential columns. The extra columns should be named with letters, numbers, and spaces; and other special symbols should be avoided.

| | A | B | essential columns C | D | extra columns E | F |
|---|---------|-------------------|---------------------|----------------|-----------------|------------------|
| 1 | case_id | image | tumor_type | sub tumor type | author | extra annotation |
| 2 | 1 | images/000001.png | B | NONE | author A | annotation |
| 3 | 8 | images/000010.png | M | NONE | author B | annotation |
| 4 | 8 | images/000011.png | M | NONE | author C | annotation |

3. Zip your dataset

The last step is to select the *images* folder and *label.csv* file and zip them together. '.zip' format is required, and other formats such as '.rar', '.7z' are not supported. Please do not zip the parent folder. The *label.csv* and *images* folder should be top-level in the zipped file.

