**Dataset Description**

The dataset for this competition comprises over a thousand high-resolution whole-slide digital pathology images. Each slide depicts a blood clot from a patient that had experienced an acute ischemic stroke.

The slides comprising the training and test sets depict clots with an etiology (that is, origin) known to be either **CE (Cardioembolic)** or **LAA (Large Artery Atherosclerosis)**. We include a set of supplemental slides with a either an unknown etiology or an etiology other than CE or LAA.

Your task is to classify the etiology (CE or LAA) of the slides in the test set for each patient.

**File and Data Field Descriptions**

* **train/** - A folder containing images in the TIFF format to be used as training data.
* **test/** - A folder containing images to be used as test data. The actual test data comprises about 280 images.
* **other/** - A supplemental set of images with a either an unknown etiology or an etiology other than CE or LAA.
* **train.csv** Contains annotations for images in the train/ folder.
  + image\_id - A unique identifier for this instance having the form {patient\_id}\_{image\_num}. Corresponds to the image {image\_id}.tif.
  + center\_id - Identifies the medical center where the slide was obtained.
  + patient\_id - Identifies the patient from whom the slide was obtained.
  + image\_num - Enumerates images of clots obtained from the same patient.
  + label - The etiology of the clot, either CE or LAA. This field is the classification target.
* **test.csv** - Annotations for images in the test/ folder. Has the same fields as train.csv excluding label.
* **other.csv** - Annotations for images in the other/ folder. Has the same fields as train.csv. The center\_id is unavailable for these images however.
  + label - The etiology of the clot, either Unknown or Other.
  + other\_specified - The specific etiology, when known, in case the etiology is labeled as Other.
* **sample\_submission.csv** - A sample submission file in the correct format. See the [**Evaluation**](https://www.kaggle.com/competitions/mayo-clinic-strip-ai/overview/evaluation) page for more details. Note in particular that you should make one prediction per patient\_id, **not** per image\_id.

## File:1158 files, Size:395.36 GB, Type: tif, csv

**DATA LINK:** [**https://www.kaggle.com/competitions/mayo-clinic-strip-ai/data**](https://www.kaggle.com/competitions/mayo-clinic-strip-ai/data)