

# Overview of MuTILs output

---

## > Where are the MuTILs inference results?

From any of the cooper lab servers, go to the mounted directory:

```
cd /mnt/cooperlab/cooperlabgpu/MUTILS_OUTPUT_08252021/
```

From there, go to the dataset you want to use:

```
|_ CPSII_40X/ ← start with this dataset
|_ CPS3_40X/
|_ plco_breast/
|_ TCGA_BRCA/
|_ DissertationAndItsAnalysis_08252021/ ← ignore this
|_ LOGS/ ← ignore this
|_ z_Meh/ ← ignore this
```

## > How are the outputs aggregated?

Each slide has different types of outputs at various levels of “aggregation”/granularity. There are features for each of the following:

- *Individual nuclei*: Each row is a feature vector for one nucleus
- *Individual regions*: Each row is a feature vector for a single contiguous tissue region
- *ROIs*: Each row is a feature vector for a rectangular area, summarizing its nuclear and collagen contents. Most ROI-level features are attained by taking the mean and variance of nuclear features within that ROI.
- *Slides*: Each row represents one slide. The same patient may have multiple slides/rows.

Note how various aggregation operations are happening:

```
Individual nuclei in ROI → ROI nuclear feature summary → Slides → Patients
                        ROI collagen features           → Slides → Patients
Individual regions in slide → Slides → Patients
```

I recommend starting by working with the ROI-level features by concatenating the nuclear feature summary with the collagen feature summary for the same ROI. This is a more tractable problem than working with individual nuclei or regions.

*Note: you may need to do some imputation or other measures to account for missing features. When aggregating the features, I used a combination of non-NAN mean and imputation, but feel free to devise a strategy that makes sense for your own method.*

## > How and where are the outputs organized?

You can find all levels of aggregation in the dataset folder. The python file `ctme/regions/MuTILsWSIRunner.py` is used to:

- Take a WSI image and run the MuTILs deep learning model
- Get a whole-slide level semantic segmentation mask at 20x for regions and cells
- Get the *individual* nuclear features for nuclei in a slide

The outputs are organized as follows:

```
|_ CPSII_40X/
|_ LOGS/ ← ignore this
|_ perSlideResults/ ← Output from ctme/regions/MuTILsWSIRunner.py
| |
| | |_ 000020-B18/ ← Results for a single slide
| | |
| | | |_ 000020-B18.json ← Metadata including slide info, TILs scores, runtime errors
| | | |_ 000020-B18_RoiLocs.csv ← ROI locations, scores, trained model fold assigned
| | | |_ 000020-B18_RoiLocs.png ← Visualization of ROI locations on slide
| | | |_ 000020-B18.tif ← WSI mask at 20x. Channels: 0=regions, 1=nuclei, 2=nuclei edges
| | | |_ nucleiMeta/ ← Metadata about nuclei in each ROI: location + classification prob.
| | | |
| | | | |_ 000020-B18_roi-0_left-9040_top-30964_right-11300_bottom-33224.csv
| | | | . → Each row is a single nucleus in this ROI
| | | | .
| | | |
| | | |_ nucleiProps/ ← Other metadata about nuclei in each ROI: shape, texture, etc.
| | | |
| | | | |_ 000020-B18_roi-0_left-9040_top-30964_right-11300_bottom-33224.csv
| | | | . → Each row is a single nucleus in this ROI
| | | | .
| | | |
| | | |_ roiMeta/ ← Metadata about each ROI
| | | |
| | | | |_ 000020-B18_roi-0_left-9040_top-30964_right-11300_bottom-33224.json
| | | | . → ROI location, TILs score, summary of nuclei, summary of regions, etc.
| | | | .
| | | |
| | | |_ roiMasks/ ← Ignore this, it's a temp folder used during inference
| |
| |
| |
```

The output above was then used as the input for another python file, `ctme/regions/SlideFeatureExtractor.py`. This python class:

- Aggregates the nuclear features for each ROI
- Extracts collagen features for each ROI
- Extracts *individual* region features for each slide
- Aggregates ROI-level features and region features for each slide

The results are organized as follows:

```
|_ CPSII_40X/
|_ LOGS/ ← ignore this
|_ perSlideResults/ ← Output from ctme/regions/MuTILsWSIRunner.py
|_ cTMEfeats/ ← Output from ctme/integrative_analysis/SlideFeatureExtractor.py
|
|_ LOGS/ ← ignore this
|_ perSlideROIsummaries/ ← ROI-level aggregated nuclear features
| |
| | 000020-B18.csv ← Each row is a single ROI in this slide
| | .
| | .
|_ perSlideCollagenFeatures/ ← ROI-level collagen features
| |
| | 000020-B18.csv ← Each row is a single ROI in this slide
| | .
| | .
|_ perSlideRegionFeatures/ ← Individual region features in each slide
| |
| | 000020-B18.csv ← Each row is a single region in this slide
| | .
| | .
|_ perDatasetSlideSummaries/ ← Aggregated ROI-level and region features per slide
|_ Raw/ ← ignore this
|_ RoiFeatureSummary_Means.csv ← Each row is a slide
|_ RoiFeatureSummary_Stds.csv ← Each row is a slide
|_ GlobalRoiBasedFeatures.csv ← Each row is a slide
|_ CollagenFeatureSummary_Means.csv ← Each row is a slide
|_ CollagenFeatureSummary_Stds.csv ← Each row is a slide
|_ RegionFeatureSummary_Means.csv ← Each row is a slide
|_ RegionFeatureSummary_Stds.csv ← Each row is a slide
```

You should start working with the folders/files outlined in **bold**.

Let me know if you have any questions.

Best of luck!!