TCGA-WSI classification Documentation

1. Code Path

GitHub: https://github.com/CaixdLab/WSI-Classification-using-DWT

data (uploading): /home/yuanzhe/Storage/tcga_save/

2. Setup

1. Install Anaconda
    1. Download the suitable anaconda version https://www.anaconda.com/products/individual
    2. Execute the installer.
2. Create a Python environment.
    1. Create a conda environment with python = 3.7.13
    2. Activate the created environment.
    3. Check the installed python version.
    4. Conda user guide: https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#activating-an-environment

3. Required Environment

- cycler==0.11.0
- fonttools==4.38.0
- idna==3.4
- imageio==2.27.0
- joblib==1.2.0
- kiwisolver==1.4.4
- matplotlib==3.5.3
- numpy==1.21.6
- packaging==23.0
- pandas==1.3.5
- pillow==9.5.0
- plotly==5.14.1
- pyparsing==3.0.9
- python-dateutil==2.8.2

- pytz==2023.3
- pywavelets==1.3.0
- requests==2.28.2
- scikit-image==0.19.3
- scikit-learn==1.0.2
- scipy==1.7.3
- seaborn==0.12.2
- six==1.16.0
- tenacity==8.2.2
- torch==1.13.1
- torchvision==0.14.1
- tqdm==4.65.0
- urllib3==1.26.15

4. Files Description (updating)

Name	Description
gdc_manifest.2023-05-10_LUSC.txt	LUSC metadata for gdc download
gdc_manifest.2023-05-10_LUAD.txt	LUAD metadata for gdc download
crop_2048_256.py	Crop 2048*2048 tiles to 256*256 tiles
crop_224_nonoverlap.py	Crop 2048*2048 tiles to nonoverlap 224*224 tiles
crop_224_overlap.py	Crop 2048*2048 tiles to overlap 224*224 tiles
Sizedistribution_2022_1125.pdf	Size distribution of WSI (updated)

label.csv	TCGA luad/lusc bag label
crop_single20.py	Generates tiles
wpd_0225.py	wavelet packet transform
wsi_wpd.py	wavelet packet transform
abmil.py	Baseline abmil Ilse, M., Tomczak, J. and Welling, M., 2018, July. Attention-based deep multiple instance learning. In *International conference on machine learning* (pp. 2127-2136). PMLR.
clam.py	Baseline CLAM Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M. and Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, *5*(6), pp.555-570.
MIL_ElasticNet_v3_5.py	multiple instance learning with elasticnet