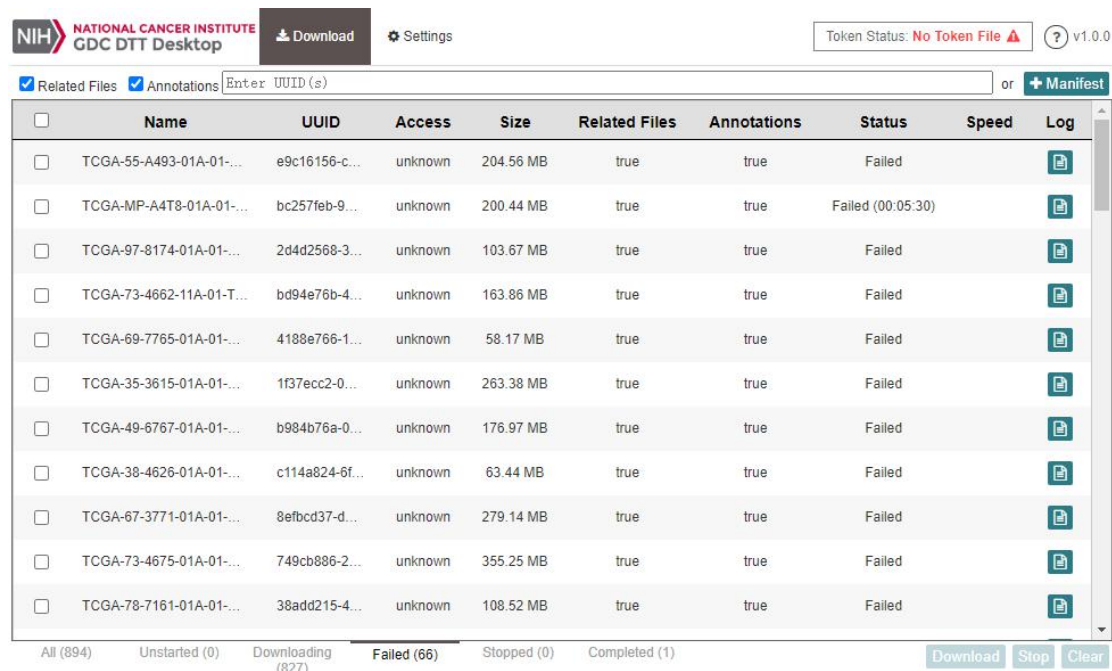


Explanation of code items

After a lot of attempts, unfortunately, I didn't realize the project in the end.

My initial idea was very simple. Just take the digits 0 and 7 as *solid_tissue_normal* and *primary_solid_tumor* and complete them step by step according to the code in the paper. I first tried to find out the structure of the original data, so I downloaded GDC client and GDC data transfer tool and planned to download WSIs data according to the manifest file. After spending a lot of time, I found that almost all the UUIDs provided by the manifest are invalid. Only one can be downloaded. I have to manually search for items to download, but the time cost is unacceptable.



The screenshot shows the NIH National Cancer Institute GDC DTT Desktop interface. At the top, there's a header with the NIH logo, 'NATIONAL CANCER INSTITUTE GDC DTT Desktop', a 'Download' button, and a 'Settings' icon. On the right, a 'Token Status' box shows 'No Token File' with a red warning triangle and a version 'v1.0.0'. Below the header, there's a search bar with 'Enter UUID(s)' and a '+ Manifest' button. The main area is a table with columns: Name, UUID, Access, Size, Related Files, Annotations, Status, Speed, and Log. The table lists 10 items, all with a status of 'Failed'. At the bottom, there's a summary bar showing counts for 'All (894)', 'Unstarted (0)', 'Downloading (827)', 'Failed (66)', 'Stopped (0)', and 'Completed (1)'. There are also 'Download', 'Stop', and 'Clear' buttons.

<input type="checkbox"/>	Name	UUID	Access	Size	Related Files	Annotations	Status	Speed	Log
<input type="checkbox"/>	TCGA-55-A493-01A-01-...	e9c16156-c...	unknown	204.56 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-MP-A4T8-01A-01-...	bc257feb-9...	unknown	200.44 MB	true	true	Failed (00:05:30)		
<input type="checkbox"/>	TCGA-97-8174-01A-01-...	2d4d2568-3...	unknown	103.67 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-73-4662-11A-01-T...	bd94e76b-4...	unknown	163.86 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-69-7765-01A-01-...	4188e766-1...	unknown	58.17 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-35-3615-01A-01-...	1f37ecc2-0...	unknown	263.38 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-49-6767-01A-01-...	b984b76a-0...	unknown	176.97 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-38-4626-01A-01-...	c114a824-6f...	unknown	63.44 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-67-3771-01A-01-...	8efbcd37-d...	unknown	279.14 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-73-4675-01A-01-...	749cb886-2...	unknown	355.25 MB	true	true	Failed		
<input type="checkbox"/>	TCGA-78-7161-01A-01-...	38add215-4...	unknown	108.52 MB	true	true	Failed		

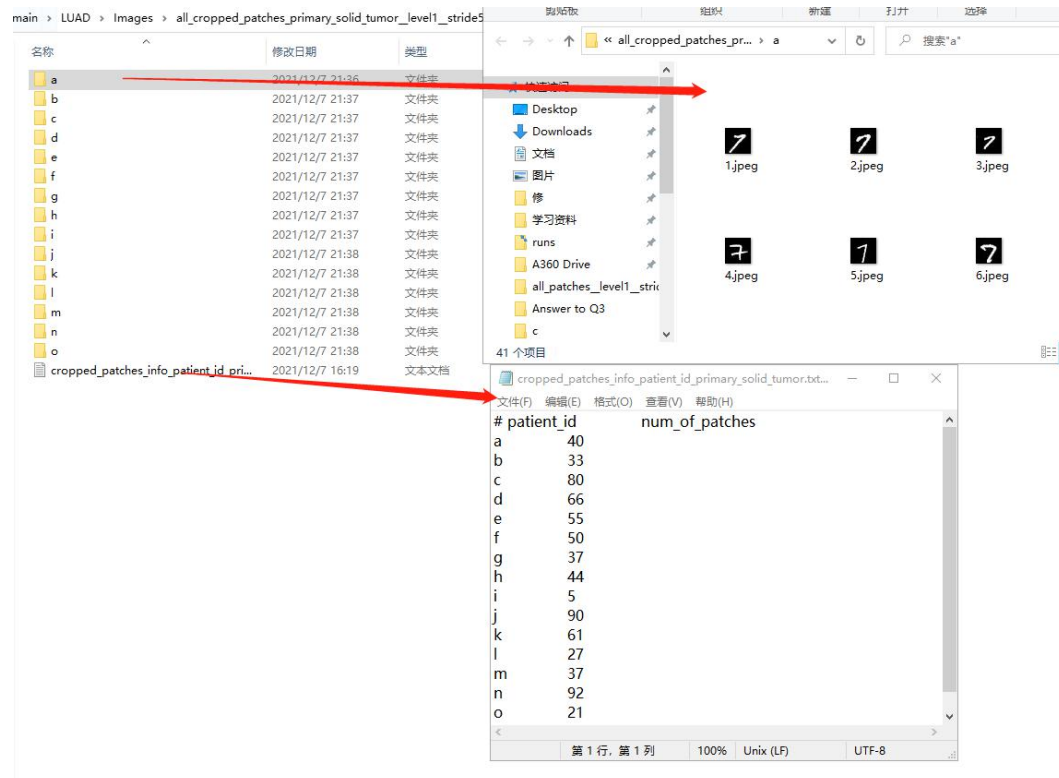
All (894) Unstarted (0) Downloading (827) **Failed (66)** Stopped (0) Completed (1)

Download Stop Clear

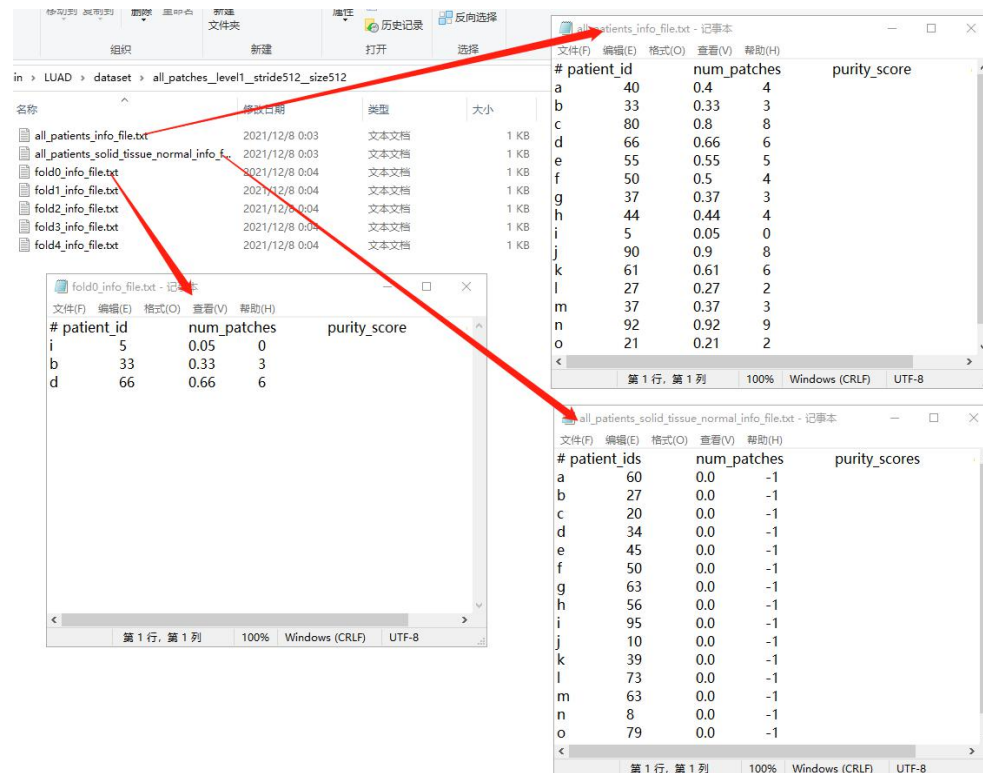
Next, I intend to skip the above steps and directly construct my own dataset according to the information provided by *tissue masks and patch clipping*:

```
all_cropped_patches_primary_solid_tumor_level1_stride512_size512
├─ cropped_patches_info_patient_id_primary_solid_tumor_all.txt
├─ TCGA-05-4244
│   └─ cropped_patches_filelist.txt
│       ├── 0.jpeg
│       ├── 1.jpeg
│       └─ ...
├─ TCGA-05-4420
│   └─ cropped_patches_filelist.txt
│       ├── 0.jpeg
│       ├── 1.jpeg
│       └─ ...
└─ ...
```

I created two folders, constructed 15 samples manually, put the digits 7 in *primary_solid_tumor*, and the digits 0 in *solid_tissue_normal*, and ensured that the number of 7 and 0 add up to 100, and respectively wrote txt files for them:



In *Prepare Dataset* phase, I successfully run the *compare_imaging_and_genomic_data.py* and *generate_5_fold_data.py*, extract the relevant information, and divide the samples into five groups:



But in the training stage, because of the lack of time, I couldn't make the training successful. The following problems occurred when I directly ran the *train.py*:

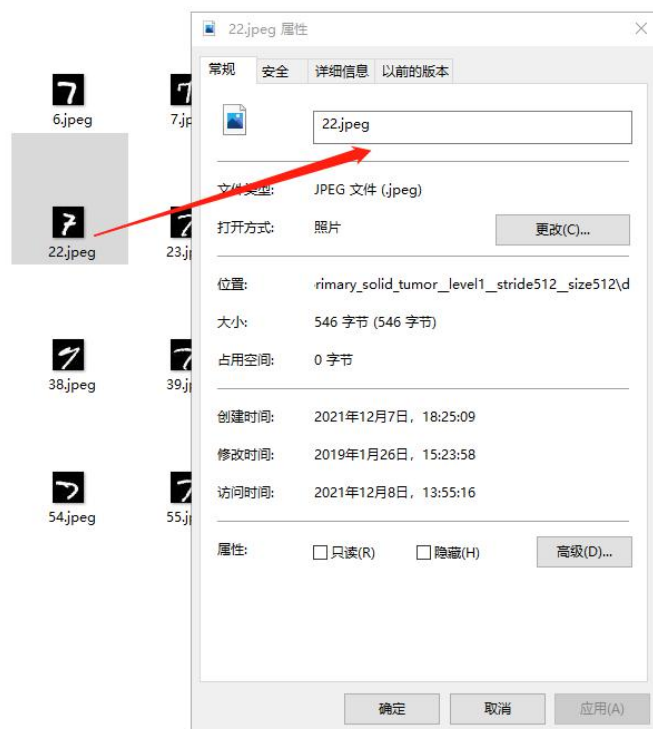
```
RuntimeError: DataLoader worker (pid(s) 16268, 17188, 8468, 13016) exited unexpectedly
0%|          | 0/9 [00:05<?, ?it/s]
```

According to the online tutorial, I set *num_workers* to 0:

```
True, num_workers=0,
e, num_workers=0, co
```

But then there was a problem that the program could not find the directory, even though the path did have picture, even if it was modified to a full path, the problem still exists:

```
FileNotFoundError: [Errno 2] No such file or directory: '../Images/all_cropped_patches_primary_solid_tumor_level1_stride512_size512/d/22.jpeg'
0%|          | 0/9 [00:00<?, ?it/s]
```



I tried to debug, then I found *train_data_loader* may have problems:

```
0%|          | 0/9 [00:00<?, ?it/s]Traceback (most recent call last):
File "C:/Users/chenyuzhou/Desktop/ISRTPMs-main/LUAD/mil_dpfr_regression/train.py", line 125, in <module>
for images, targets in train_data_loader:
```

Through further analysis, the problem may still in *num_workers*, and debugging has entered a dead end:

```
# DataLoader object so that workers can be reused
if self.persistent_workers and self.num_workers > 0:
    if self._iterator is None:
        self._iterator = self._get_iterator()
    else:
        self._iterator._reset(self)
    return self._iterator
else:
    return self._get_iterator()
```

I think there are two reasons for the current situation. First, my environment is Anaconda+Pytorch under Windows system. I noticed that the experiments were run on a Linux machine. System differences may cause program conflicts. Second, there may be problems with the data setting. I speculated about the possible data structure through the instructions on GitHub. There may be discrepancies or omissions. However, this conjecture can be confirmed only when I download the WSIs data completely and compare the two structures.

Due to tight time constraints and forbidden to ask for help, I failed to implement a simpler version. I have uploaded relevant files to GitHub, hoping to show my ability in analyze and solve problems and my efforts. I am eager to join NTU to study BMDS, I hope I still have the opportunity to fight for all of this.