

DELAB Intern Ahn Sung Hyun

---

# BraTS 데이터셋 조사

<2023/01/04>

# PT

---

1. 데이터셋 설명
2. 데이터셋 구조
3. 데이터 시각화
4. 다운로드 및 평가

---

# 데이터셋 설명

- Brain Tumor Segmentation (BraTS)
  - <http://braintumorsegmentation.org/>
  - Brain Tumor (뇌종양:뇌와 뇌를 둘러싸고 있는 뇌척수막에서 생기는 종양) 에 대한 Dataset
  - Task1 (Segmentation) 과 Task2 (Classification) 로 Dataset이 나뉘어진다

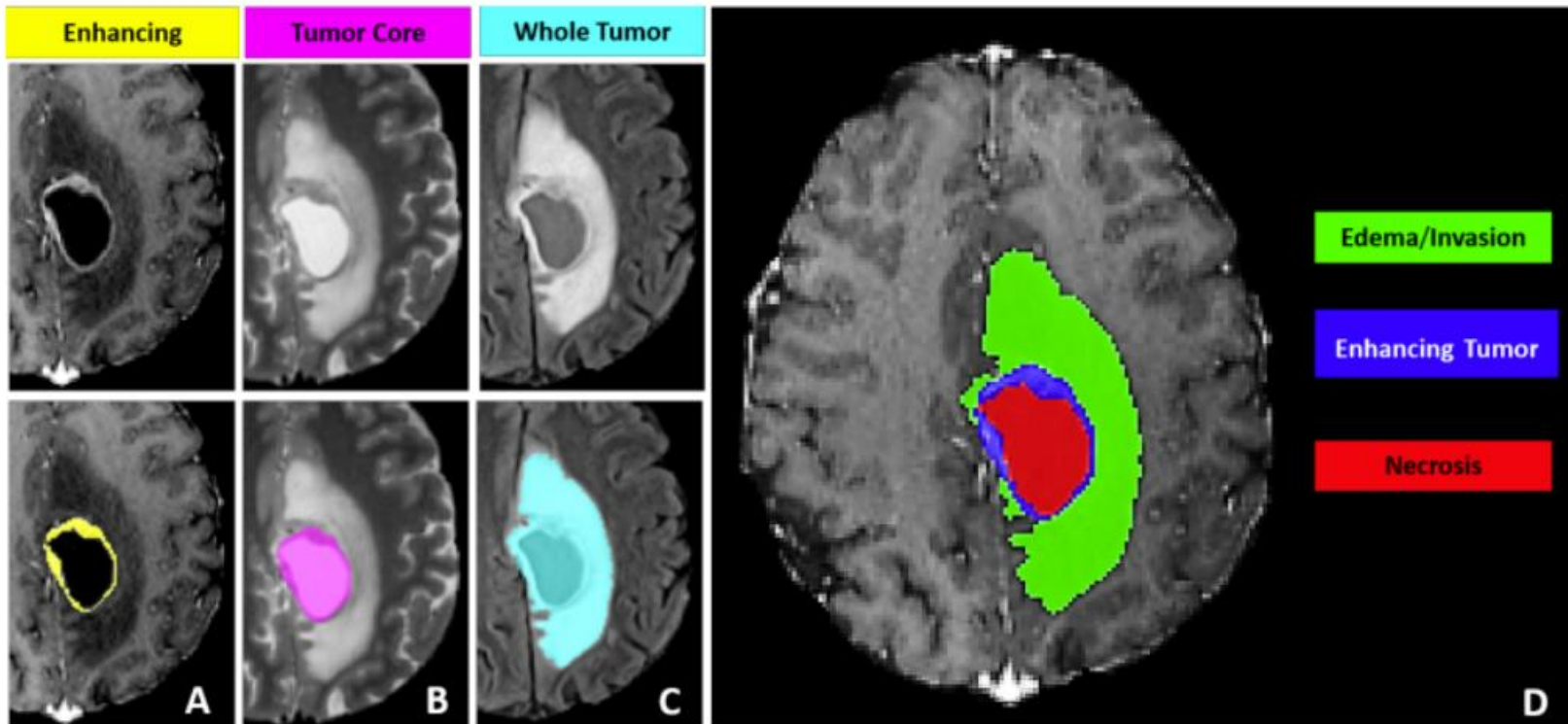


**Task 1: Brain Tumor Segmentation in mpMRI scans.**

**Task 2: Radiogenomics: Prediction of the MGMT promoter methylation status in mpMRI scans.**

# 데이터셋 설명

- BRaTS Task 1 Dataset
  - Brain Tumor의 세 가지 범주에 대한 3D Segmantation Dataset
  - **Category for Data:** Edema(ED), Enhancing Tumor(ET), Necrosis(NCR)
  - **Category for Segmentation:** Enhancing(ET), Tumor Core(TC), Whole Tumor(WT)
  - **ET = ET, TC = NCR + ET, WT = ET + NCR + ED**





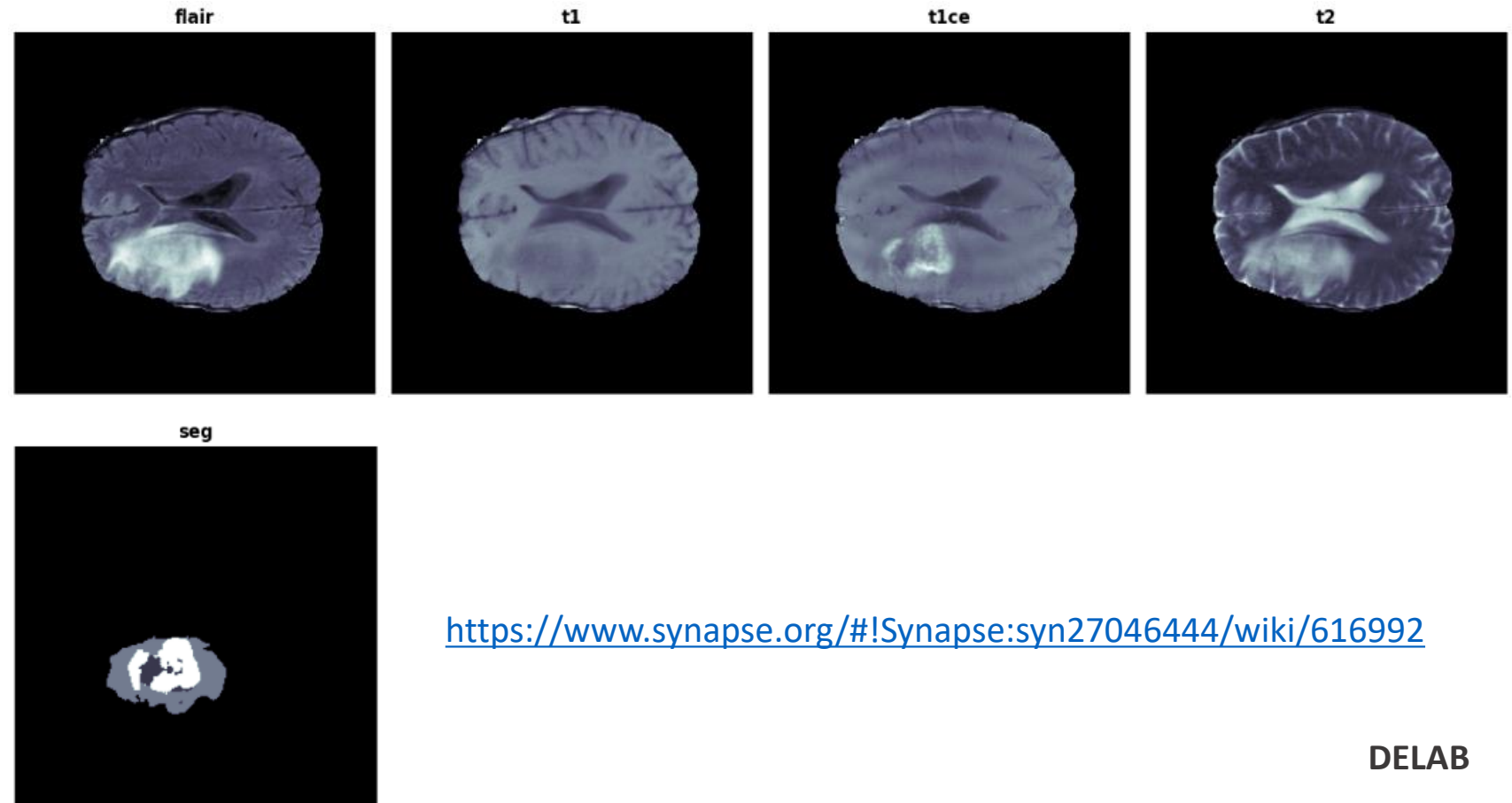
<https://arxiv.org/pdf/2107.02314.pdf>

# 데이터셋 설명

- Dataset Component

- 환자별로 다섯 개의 NifTI 파일 (.nii.gz) 을 가지고 있음 -> **flair, t1, t1ce, t2, seg**
- NifTI:** Brain MRI 영상 등을 표현할 때 자주 쓰이는 형식
- flair ~ t2**는 같은 뇌에 대해 각각 다른 스캐너를 이용해서 출력한 Data, **seg**는 Ground Truth Data (GT : Tumor Area)

-  BraTS2021\_00000\_flair.nii
-  BraTS2021\_00000\_seg.nii
-  BraTS2021\_00000\_t1.nii
-  BraTS2021\_00000\_t1ce.nii
-  BraTS2021\_00000\_t2.nii

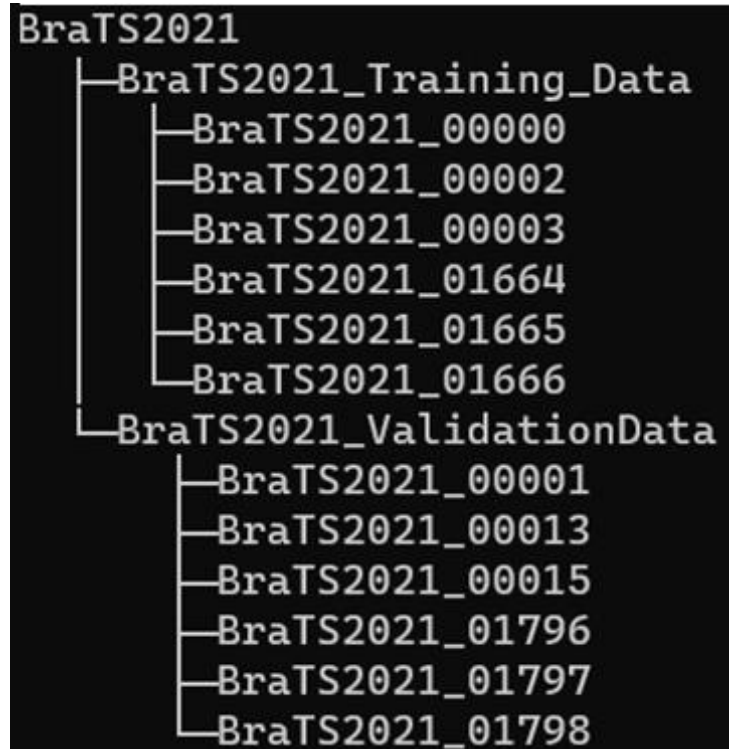


<https://www.synapse.org/#!/Synapse:syn27046444/wiki/616992>

# 데이터셋 구조

- Dataset Structure

- Training Data에는 1251명의 환자 정보가 있음 (12.4 GB)
- Validation Data에는 219명의 환자 정보가 있음 (2.2 GB)
- Test Dataset은 따로 제공하지 않음 (논문에서도 Validation Set으로 Evaluation을 진행함)

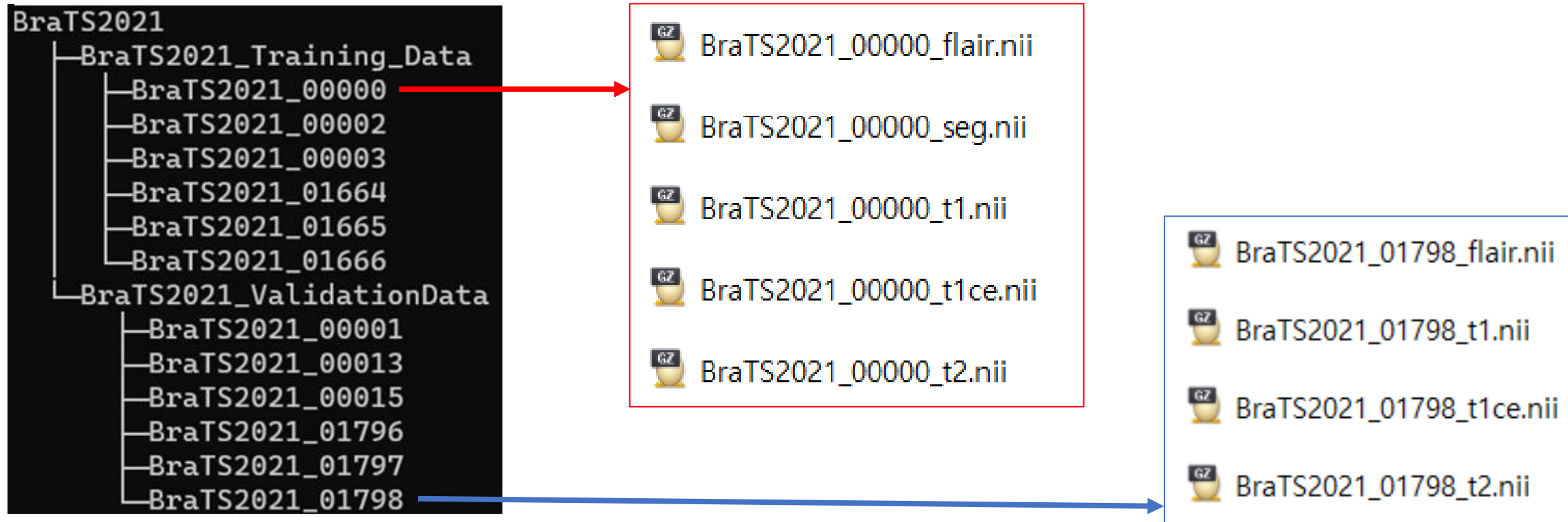


크기:	12.4GB (13,383,266,621 바이트)
디스크 할당 크기:	12.4GB (13,395,988,480 바이트)
내용:	파일 6,255, 폴더 1,251

크기:	2.22GB (2,392,941,647 바이트)
디스크 할당 크기:	2.22GB (2,394,722,304 바이트)
내용:	파일 876, 폴더 219

# 데이터셋 구조

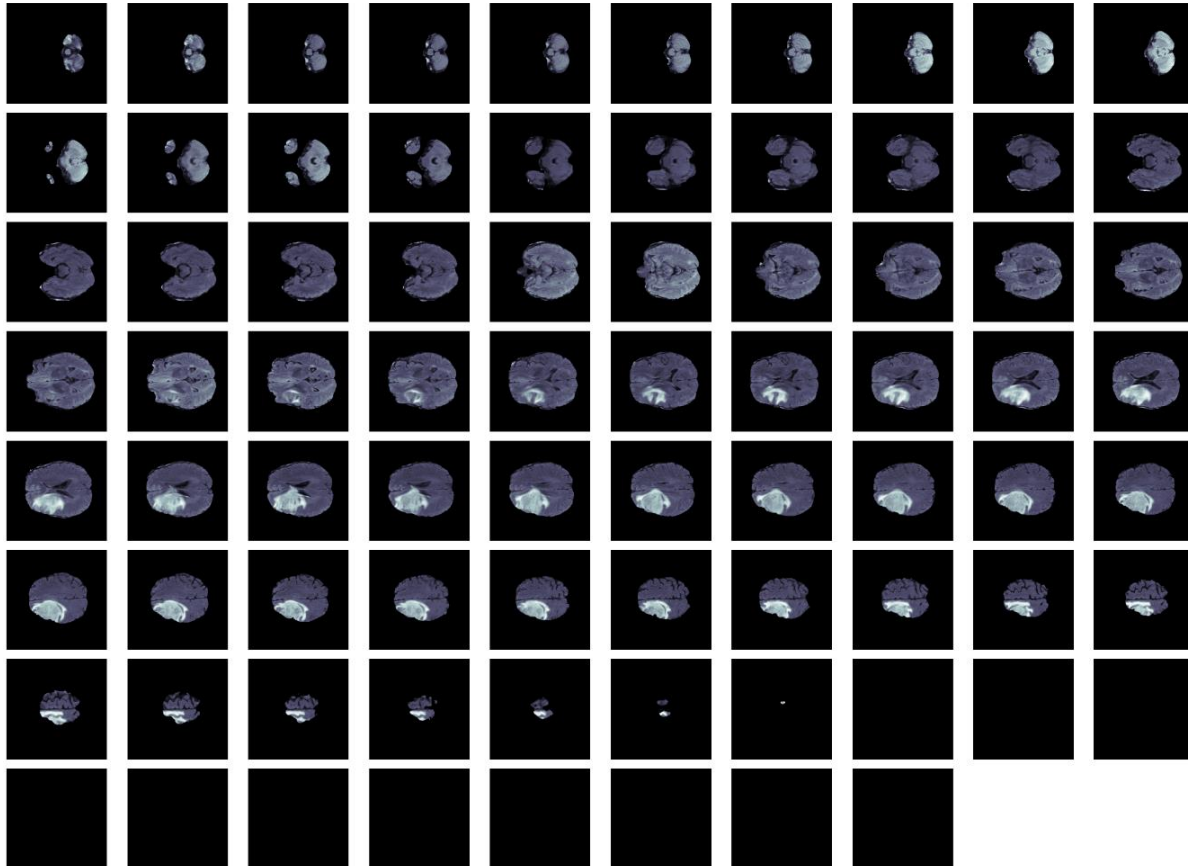
- Dataset Singular Point
  - Training Set에는 GT(seg file)가 있지만, Validation Set에는 GT가 없음
  - Validation Set으로 평가할 때는 [벤치마킹 사이트](#)를 이용해야 함 (하루에 2번 업로드 가능)



# 데이터 시각화

- Data Visualization

- Brain Tumor Data(3D)는 155장의 2D Data로 이루어져 있음
- 3차원 뇌를 155장의 단면으로 표현했다고 생각하면 됨
- 환자마다 H x W x D 의 데이터 (H,W: 해상도<240>, D: 뇌의 단면 개수<155>) 를 다섯 개 지님 (flair, t1, t1ce, t2, seg)



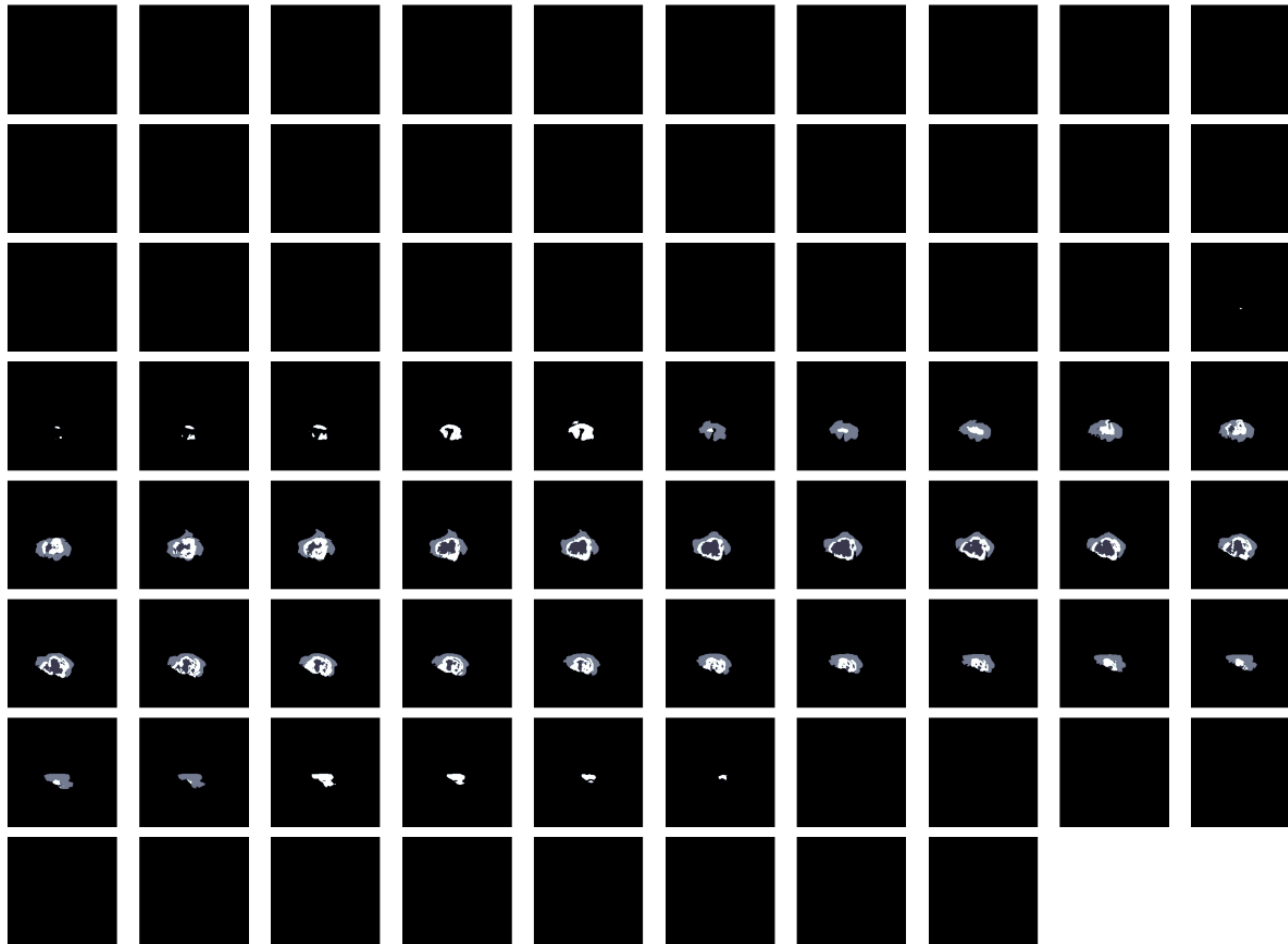
-> 한 환자에 대한 Flair Data 일부 스캔 ( $\frac{155}{2}$  Data)



---

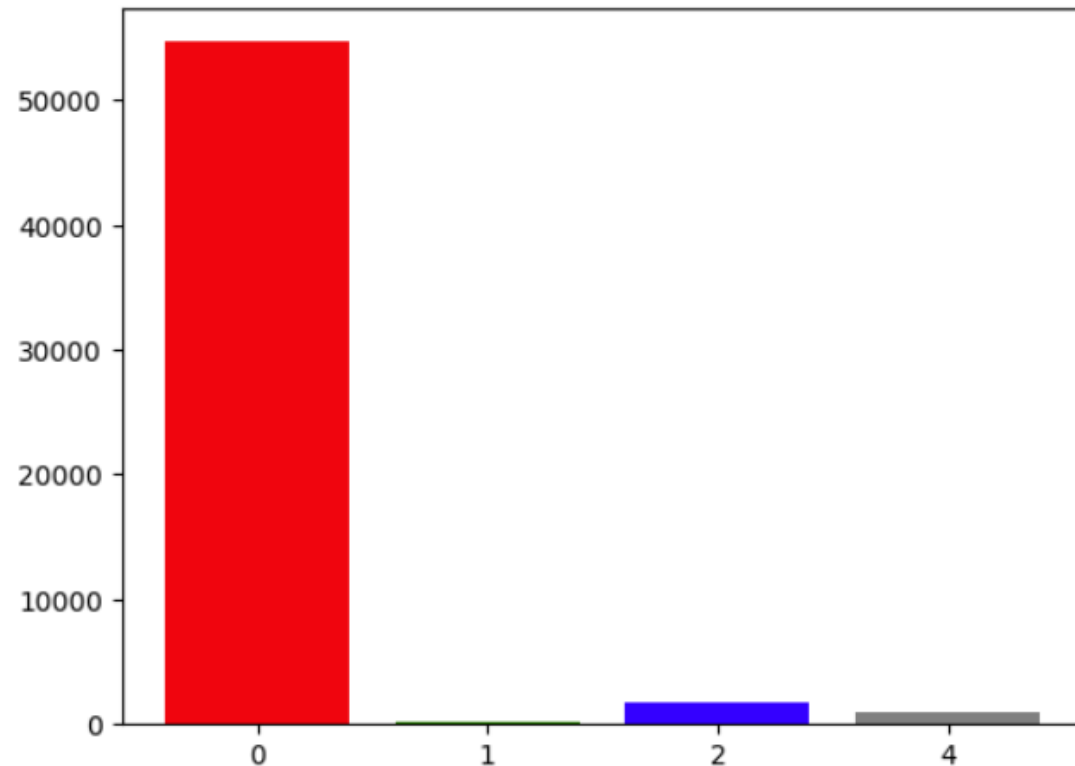
# 데이터 시각화

- Data Visualization
  - 한 환자에 대한 **Seg Data** 일부 스캔 (155/2 Data)



# 데이터 시각화

- Ground Truth
  - 한 환자의 Seg (240x240x155) 중 임의의 단면 정보 확인 (ex. Seg[:, :, 80])
  - 임의의 단면 정보(240x240)를 Flatten한 뒤 고유값을 계산해 봄 -> (0, 1, 2, 4가 54667, 275, 1677, 981개 나옴)
  - 0: Normal, 1: Necrosis(NCR), 2: Edema(ED), 4: Enhancing Tumor(ET)



```
array([0., 1., 2., 4.])
```

```
array([54667, 275, 1677, 981])
```

<https://www.synapse.org/#!/Synapse:syn27046444/wiki/616992>

# 데이터 시각화

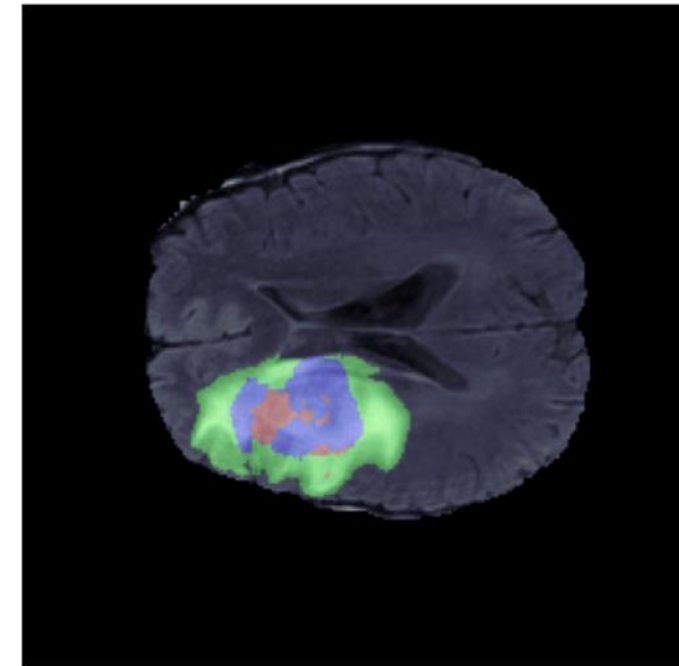
- Color Segmentation Map
  - Seg Data에는 어떤 픽셀이 어떤 범주인지에 대한 정보를 지니고 있음
  - 이 정보를 이용해서 **Color Segmantation Map (240x240x3)**을 구성할 수 있음
  - **1 = NCR -> Red**, **2= ED -> Blue**, **4 = ET -> Green**

```
flair=flair_data[:, :, 80]
seg=seg_data[:, :, 80]

color_segmentation=np.zeros((240,240,3))
gray_segmentation=seg

color_segmentation[gray_segmentation == 1] = [255,0,0] # Red (Necrotic Tumor Core, NCR)
color_segmentation[gray_segmentation == 2] = [0,255,0] # Green (Edema, ED)
color_segmentation[gray_segmentation == 4] = [0,0,255] # Blue (Enhancing Tumor, ET)

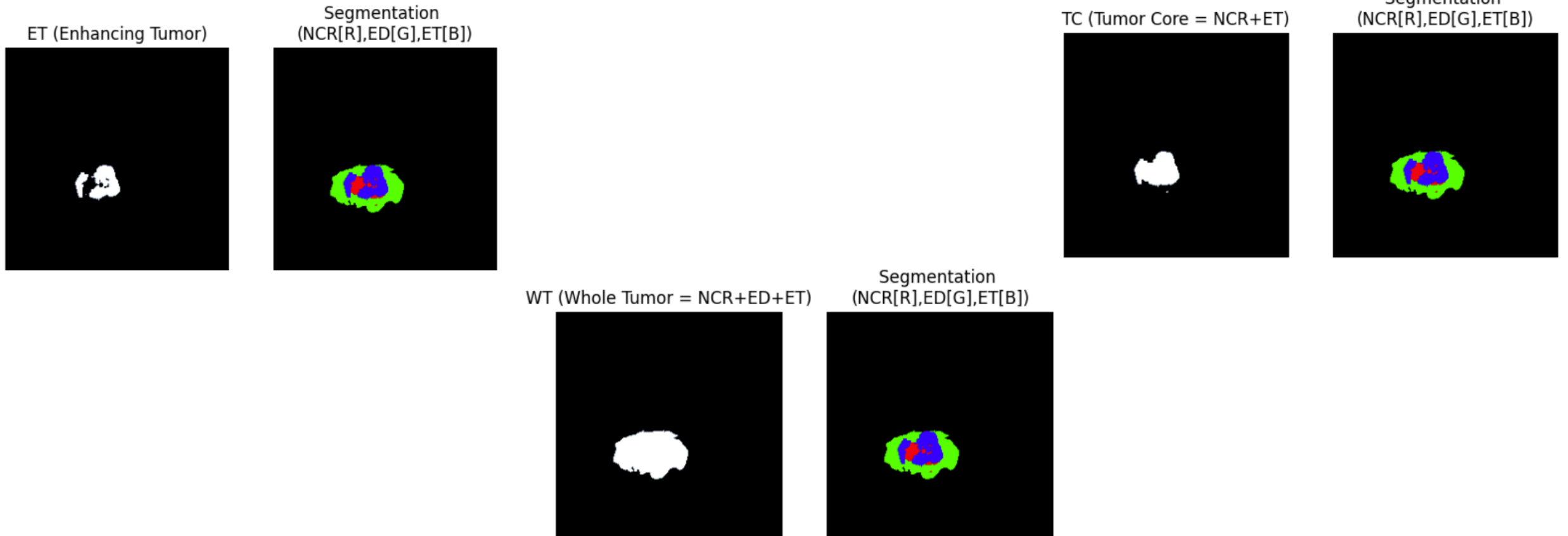
plt.title('color segmentation map with flair')
plt.axis('off')
plt.imshow(flair,cmap='bone')
plt.imshow(color_segmentation,alpha=0.3,cmap='bone')
```



# 데이터 시각화

- Segmentation Map for Deep Learning

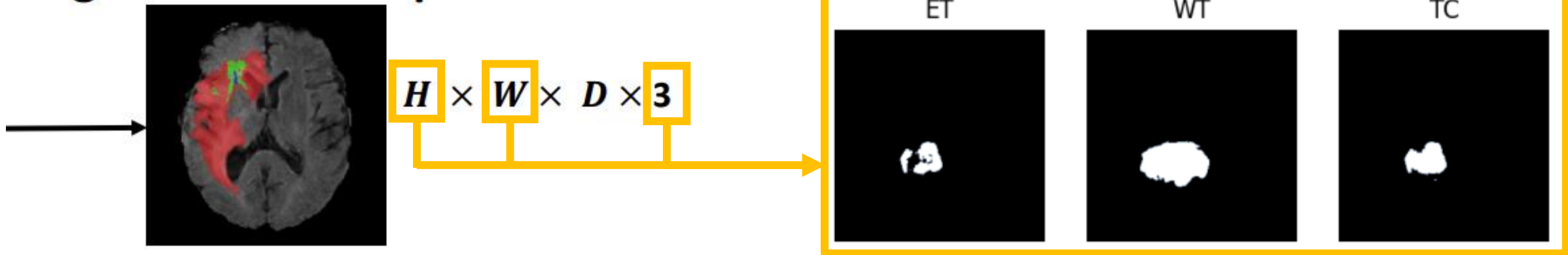
- 실제 딥 러닝으로 학습할 때는 (NCR, ED, ET) 범주 대신 (ET, TC, WT) 범주를 사용함
- 따라서 ET, TC, WT에 대한 Segmentation Map을 만들 필요가 있음



# 데이터 시각화

- Segmentation Map for Deep Learning
  - 딥 러닝 모델은 ET, WT, TC 에 대응되는 세 개의 채널을 가지는 Segmentation Output을 생성함
  - Brain Tumor Segmentation은 각 채널 별로 Pixel별 Tumor를 예측하는 Task

## Segmentation Output



Final segmentation output consists of 3 output channels corresponding to ET, WT and TC sub-regions.

---

# 다운로드 및 평가

- Dataset Download
  - <https://www.synapse.org/#!/Synapse:syn27046444/wiki/616992>
  - Register -> Request Data with Google Form -> Mail Check -> Data Download

1. Register for the challenge. You must be registered before completing the next steps.

2. Complete the data access request form. You will receive a confirmation email when you are given access to the training data.

Request Data Access

3. Download the Training Data.

Access the Data

---

# 다운로드 및 평가

- Validation Set Evaluation
- <https://www.synapse.org/#!/Synapse:syn27046444/wiki/616996>

1. Create a Synapse Project

2. Create your Segmentation Files

3. Upload your Submission Tarball -> 퀴즈 15개 맞춰야 업로드 가능!

4. Submit your Predictions

## Submission Limit

For this Challenge, you and your team are limited to two scored submissions a day. The daily submission limit will reset every day at 00:00:00 UTC.

---

# GitHub

[BRATS2021 Segmentation] 데이터 분석 코드는 GitHub를 참고

<https://github.com/SkiddieAhn/Study-Data-Analysis/blob/master/brats-seg/medical.ipynb>



DELAB Intern Ahn Sung Hyun

---

**Thank You**

<2023/01/04>