

Joelle Fitzgerald

HIDS 509

Homework 3: Brain Cancer Image Segmentation Pipeline on Seven Bridges

Question 1) Write a short summary about the (a) why segmentation is important, (b) various segmentation tools you learnt about , and (c) possible challenges of such workflows (6 points)

- a. Segmentation is the process of dividing digital images into subgroups called image segments. Labels are then assigned to pixels to identify important objects or elements of a given image. Image segmentation is important because it allows investigators to identify different regions of an image, reduces the complexity of images for easy analysis, and is a significant step for computer vision technologies and algorithms.
- b. One segmentation tool used for skull stripping or brain extraction in order to get a clear image of the brain while also preserving patient confidentiality (proper de-identification) is the Brain Mask Generator (BrainMaGe). This tool uses deep learning segmentation architecture to strip brain images of bone, marrow, and blood vessels that are not needed for analysis. This is challenging as some non-brain material may be hard to remove computationally and result in affected segmentation outputs. Another tool used during the brain segmentation process is called SUSAN or Smallest Univalve Segment Assimilating Nucleurs. SUSAN uses nonlinear filtering to reduce the noise of images while preserving the structure and quality of the image contents. This tool is helpful for smoothing all volumes and reduces the possibility of high frequency intensity variations in an image. Additionally, Seven Bridges uses the tool 'FAST' with FSL brain segmentation that relies on the intensity of in the radio frequency to identify subregions of the brain which can be useful for highlighting cerebrospinal fluid (CSF) and other portions of the brain.
- c. The main challenge of segmentation workflows is maintaining the integrity and quality of a given image while removing and refining certain aspects for proper analysis. If there is non-brain elements left over from the brain extraction step or too much/too little noise left over from the denoising step, segmentation outputs suffer and may not be as useful to gain insight from. Imaging technology can introduce a lot of noise and when handling the data, there is risk for noise to be added and error introduced. With Statistical Parametric Mapping (SPM), which produces similar outputs to FSL, the brain extraction and skull-stripping step is not used. Instead, SPM uses probability maps to understand and identify subregions of the brain such as grey matter, white matter, CSF, etc. However, predicting the boundaries of each voxel is challenging.

Question 2) Write a short summary in simple words what the FSL brain segmentation workflow in Seven Bridges does (1 point)

- a. The FSL brain segmentation workflow works through five steps of image segmentation processing. First, images are re-oriented using LPS (left posterior superior) coordinate system (a requirement for GLISTRboost) so that they can be properly read. Then, images are registered using FSL FLIRT and skull stripped using FSL BET to de-noise and clear out all non-brain elements (image normalization). Finally images are segmented by FSL FAST and de-noised using SUSAN as discussed above. Files (HTML) are outputted with grey matter, white matter, CSF, and other tissue (FLAIR images only).

Question 3) Can you think of any real world use cases/applications that would use such an image segmentation workflow ? (2 point)

- a. Some real-world applications for brain image segmentation include measuring and visualizing the brain's anatomical features for research purposes or education, analyzing brain changes, delineating pathological regions of the brain, and surgical preparation and image-guided procedures. Additionally, as we discussed in class, tumor size and location is easy to analyze and predict using image segmentation which can help clinical teams in providing personalized patient care and decision making.

Question 4) (6 points) Run the FSL Brain segmentation workflow on a TCGA-GBM patient # TCGA-19-5958.

1. Selecting image files for patient #TCGA-19-5958 in Data Browser. Copy to project.

The screenshot displays the Seven Bridges Genomics Data Browser interface. At the top, the navigation bar includes 'Projects', 'Data', 'Public Apps', 'Public Projects', and 'Developer'. The 'Data' tab is active, showing a 'New query' button and a search bar. Below the navigation bar, a 'Case' entity is selected, showing details for 'Disease type: Glioblastoma Multiforme' and 'ID: TCGA-19-5958'. A 'File' entity is also selected, showing details for 'File name', 'Access level', and 'Body Part Examined'. The 'Add entity' dropdown menu is open, showing 'Investigation' and 'Study' options. The main content area displays a list of files for the selected case, with columns for 'File', 'Details for', and 'Connections'. The 'File' column lists several files with their IDs. The 'Details for' column shows the file's properties, including 'Access level', 'Body Part Examined', 'Data category', 'Data format', 'Disease type', 'Image count', 'Manufacturer', and 'Manufacturer Model Name'. The 'Connections' column shows the file's inbound and outbound connections.

File	Details for	Connections
1.3.6.1.4.1.14519.5.2.1.5826.4001.116850591555011833581600...	TCIA Access level: Open Body Part Examined: Brain Data category: Imaging Data format: DCM Disease type: Glioblastoma Multiforme Image count: 38 Manufacturer: SIEMENS Manufacturer Model Name: Verio	Inbound: Case 1 81622677-9742-40ea-ad60-5f9b2a624ea7 Outbound: No outbound connections

2. Selecting FLAIR image files for the patient to be used in FSL workflow.

cgcsbgenomics.com

Select files for "Archived DICOM Images"

Current Project Projects Public Files

Files

Search Extension: 8 selected Sample ID: All Task ID: All Tags: All Save selection

Copy

Experimental strate...	Path	Platform	Paired-e...	Task ID	Created on	Extension	Size	Reference geno...	Case ID	Sample ID	Series date	Series descripti
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	318.3 KiB	-	TCGA-19-5958	-	2004-03-09	dti_for_nsurg_fa
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	21.4 MiB	-	TCGA-19-5958	-	2004-03-09	dti_for_nsurg
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	10.6 MiB	-	TCGA-19-5958	-	2004-03-09	volumetric_axial
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	10.9 MiB	-	TCGA-19-5958	-	2004-03-09	volumetric_axial_gd
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	263.0 KiB	-	TCGA-19-5958	-	2004-03-09	dti_for_nsurg_tracew
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	475.1 KiB	-	TCGA-19-5958	-	2004-03-09	dti_for_nsurg_colfa
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	3.4 MiB	-	TCGA-19-5958	-	2004-03-09	flair_axial
Files	Files	-	-	-	Apr. 5, 2023 08:49	ZIP	339.2 KiB	-	TCGA-19-5958	-	2004-03-09	dti_for_nsurg_adc
Files	Files	-	-	066d34c3-f40...	Mar. 29, 2023 16:58	TAR.GZ	2.7 MiB	-	TCGA-02-0009	-	1997-06-14	flair_axial
Files / Input-TCGA-02-0009	-	-	-	-	Mar. 28, 2023 13:34	ZIP	1.7 MiB	-	TCGA-02-0009	-	1997-06-14	axial_fse_t2
Files / Input-TCGA-02-0009	-	-	-	-	Mar. 28, 2023 12:34	ZIP	1.5 MiB	-	TCGA-02-0009	-	1997-06-14	flair_axial

Refresh

Showing 1-20 of 20

3. Number of tissue type classes = 4. Image Type = FLAIR.

cgcsbgenomics.com

Projects Data Public Apps Public Projects Developer

Last update by jk160 on Apr. 5, 2023 08:50

App: FSL Brain Segmentation - Revision: 0

Task Inputs Execution Settings

Inputs

Batching On

Archived DICOM Images Change selection

Batch by: File

This will create one task for each selected item.

1.3.6.1.4.1.14519.5.2.1.5826.4001.1511326555131565368

88042112741.zip (1 item)

App Settings

Edit parameters Show editable

FSL 5.0 FAST (#FSL_5_0_FAST)

Number of tissue type classes

4

Type of image * (Type of image)

FLAIR

SBG Segmentation Class Viewer (#SBG_Segmentation_Class_Viewer)

Type of image * (Type of image)

FLAIR

Dicom2nifti (#Dicom2nifti)

Output basename

FLAIR

Output Settings

A Base64 NIFTI gallery No value

A Compressed NIFTI gallery No value

BiasField No value

Binary Brain Segmentation NIFTI images No value

Brain skull removed NIFTI image No value

Homogeneous NIFTI image No value

Original NIFTI image No value

Partial Volume Maps Brain Segmentation NIFTI images No value

Forum Terms Privacy Data Use

© 2023 Seven Bridges Genomics

4. Workflow complete. This workflow took 6 minutes and cost \$0.05.

The screenshot displays the Data Studio interface for a workflow named "FSL Brain Segmentation run - 04-05-23 12:50:31". The workflow is marked as "COMPLETED" and took 6 minutes to run, costing \$0.05. The interface includes a navigation bar with options like Projects, Data, Public Apps, and Public Projects. A table lists the task details, including the task name, submitted by (jkt60), submitted on (Apr. 5, 2023 08:53), app (FSL Brain Segmentation), duration (6 minutes), and status (COMPLETED). The footer shows a "Refresh" button and a "Showing 1-1 of 1" indicator.

Task Name	Submitted by	Submitted on	App	Duration	Status	Acti
FSL Brain Segmentation run - 04-05-23 12:50:31: file: 1.3.6.1.4.1.14519.5.2.1.5826.4001.151132655513156...	jkt60	Apr. 5, 2023 08:53	FSL Brain Segmentation	6 minutes	COMPLETED	

Output Settings

▼ A Base64 NIfTI gallery

 [TCGA-19-5958_csf.page_0.b64html](#)

 [TCGA-19-5958_white_matter.page_0.b64html](#)

 [TCGA-19-5958_gray_matter.page_0.b64html](#)

 [TCGA-19-5958_other_tissue_0.page_0.b64html](#)

The files outputted from running the FSL workflow with FLAIR images include cerebrospinal fluid, white matter, grey matter, and other tissue.

Grey Matter:

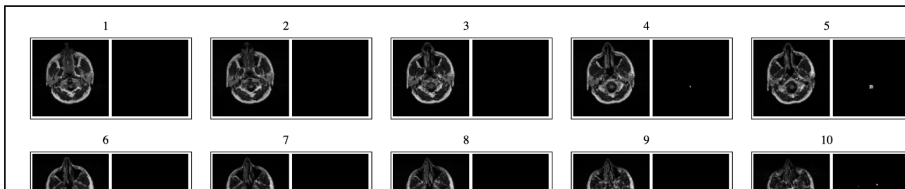
TCGA-19-5958_gray_matter.page_0.b64html

1.3 MiB (1,337,930 bytes) · Produced on April 5, 2023 09:00 (Eastern Daylight Time), by [FSL Brain Segmentation run - 04-05-23 12:50:31](#): [file: 1.3.6.1.4.1.14519.5.2.1.5826.4001.151132655513156536888042112741.zip](#) · Hosted on AWS (us-east-1)

Metadata Raw View **Preview**

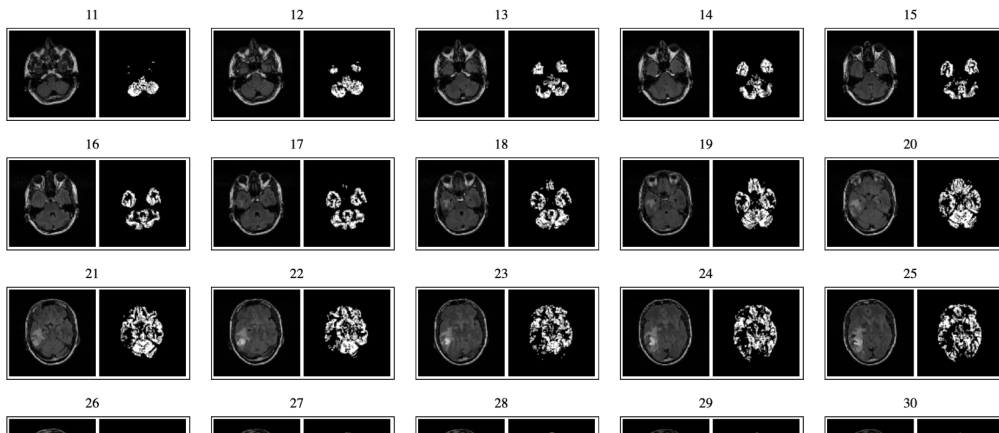
Investigation: TCGA-GBM
Case ID: TCGA-19-5958
Primary site: Brain
Age at diagnosis: 56
Gender: male
Imaging modality: MR
Image type: FLAIR
Series description: flair_axial
Slice Number: 54
Segment: gray matter

Images shown: 1-54

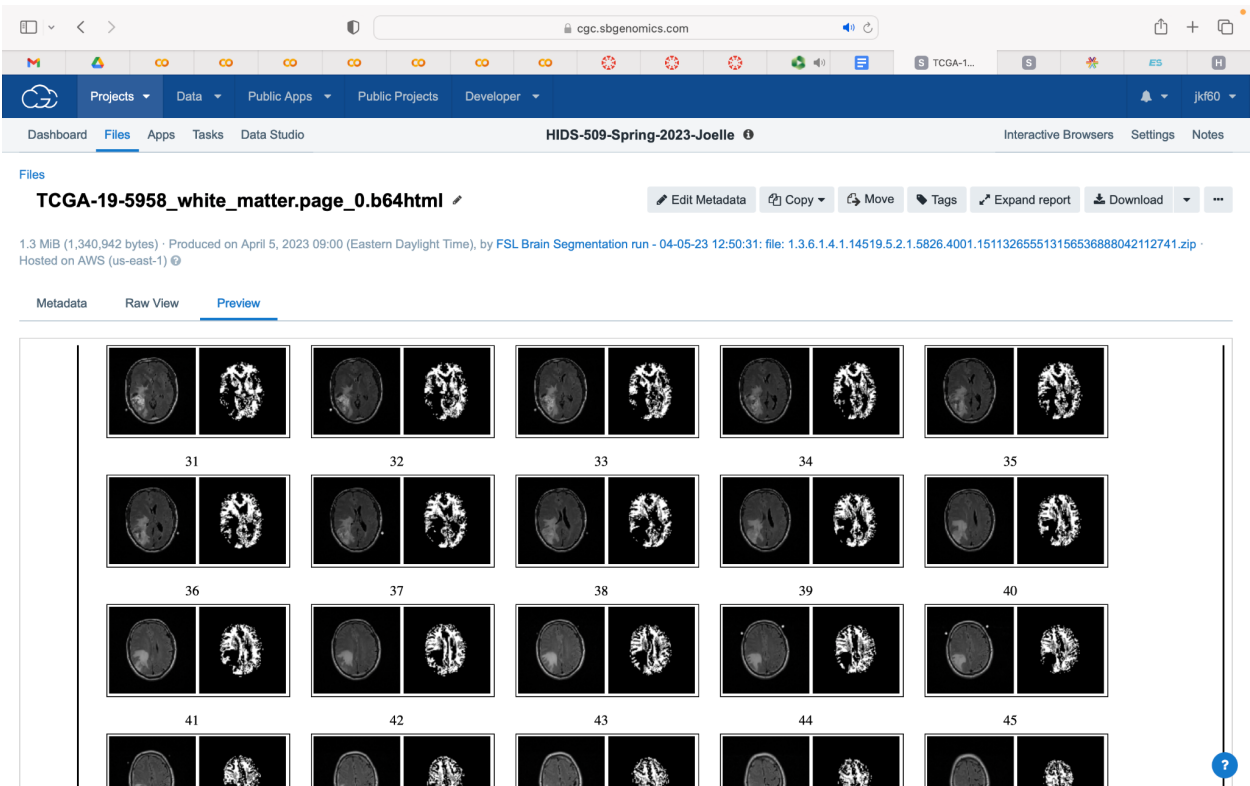


TCGA-19-5958_gray_matter.page_0.b64html [Edit Metadata](#) [Copy](#) [Move](#) [Tags](#) [Expand report](#)

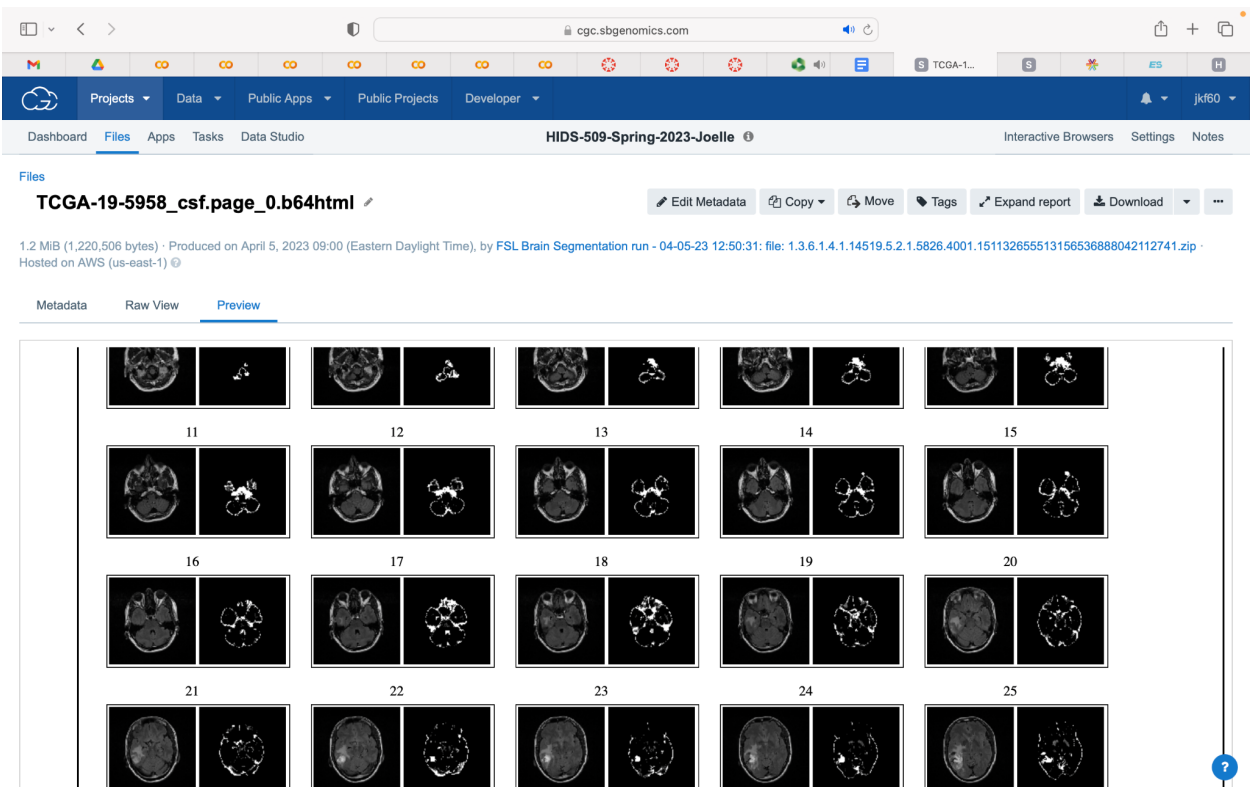
1.3 MiB (1,337,930 bytes) · Produced on April 5, 2023 09:00 (Eastern Daylight Time), by [FSL Brain Segmentation run - 04-05-23 12:50:31](#); file: [1.3.6.1.4.1.14519.5.2.1.5826.4001.151132655513156536888042112741.zip](#) · Hosted on AWS (us-east-1) 

[Metadata](#) [Raw View](#) [Preview](#)

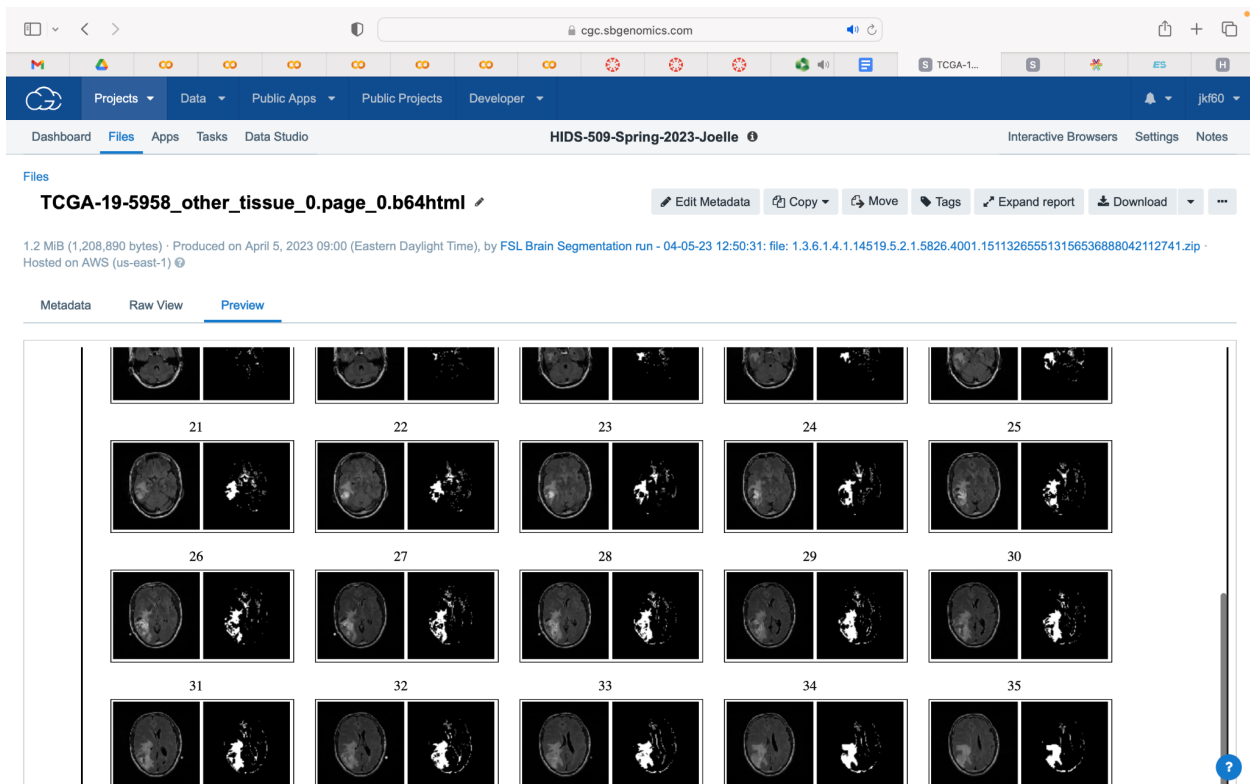
White Matter:



CSF:



Other Tissue:



From these screenshots above, you can see that the workflow was able to identify regions of the brain (grey matter, white matter, and CSF) as well as other tissue which could be a tumor. Other tissue elements are only produced with FLAIR images. In the second screenshot on the above page, the FSL workflow identified the region of tumor growth in this patient (big white blob in image).

Question 5) This TCGA Glioblastoma dataset we used in class today contains not only imaging data but also genomics data. For example, there is gene expression data from the same patients. This is represented as a matrix with features (genes) as rows and samples as columns. The numerical data are normalized gene expression values. As a data scientist, how would you integrate/connect the two data types (imaging and gene expression data) ? (Bonus 1 point)

As a data scientist, I would be interested in understanding which genes are differentially expressed in certain regions of the brain and then studying patient images where abnormalities are produced and linking which genes may be responsible (over or under expressed) for that abnormality. Connecting gene expression data with imaging data may be useful for

understanding why a tumor is growing in a certain region of the brain and highlighting patterns in cancer growth for certain patients.