

목 차

1. 요구 조건 및 제약 사항 분석	1
1-1. 요구 조건	1
1-2. 제약 사항 분석 및 수정 사항	2
2. 설계 상세화 및 변경 내역	3
2-1. 데이터 수집	3
2-2. 데이터 전처리	3
2-3. 데이터 분석 및 예측	3
3. 갱신된 과제 추진 계획	4
4. 구성원 별 진척도	5
5. 보고 시점까지의 과제 수행 내용 및 중간 결과	6
5-1. 데이터 수집	6
5-2. 데이터 분석 및 예측	7

1. 요구 조건 및 제약 사항 분석

1.1 요구 조건

뇌 MRI 영상 데이터와 MGMT 프로모터 메틸화 존재 여부와의 관계를 분석하여 교모세포종의 유전적 아형을 예측한다.

- 데이터 수집
 - MRI 영상 데이터 수집(dicom 파일)
 - 촬영 방식에 따른 MRI 영상 데이터 분류
- 데이터 전처리
 - 영상 데이터의 이미지 방향, 모양 및 크기에 대한 처리
 - 이미지가 같은 공간에 위치할 수 있도록 학습 데이터 처리
- 데이터 학습
 - CNN 방식을 사용한 데이터 학습
 - 데이터 학습 소요 시간 및 정확도 측정
- 데이터 예측
 - 학습된 모델을 바탕으로 하여 입력받은 데이터에 대한 예측
 - MGMT 프로모터 메틸화 여부를 시각적으로 표시

1.2 제약 사항 분석 및 수정 사항

- 각 환자마다 다른 영상 데이터 개수
- 촬영된 뇌의 크기, 모양, 위치에 대한 문제

-> 수집 데이터 중 필요한 부분만 사용할 수 있도록 데이터 전처리 모듈 개선

-> FreeSurfer 및 FSL 사용 예정

- 수집된 데이터의 총 용량 문제

-> 학습 모델 테스트를 위해 일부 데이터를 Google Drive에 업로드 후 Google Colab으로 테스트 진행, 이후 전체 데이터를 학과 서버에 업로드한 후 전체 데이터에 대해 테스트 진행 예정

- 데이터 예측 결과 값 확인 문제

-> 테스트 데이터에 대한 메틸화 존재 여부가 확인되지 않아, 학습 모델 검증을 위해 scikit-learn에 포함된 train_test_split 모듈 활용 예정

-> 현재 모듈에 대한 검증은 accuracy와 loss만 측정하고 있는데, 이후 sensitivity, specificity, F1-score, Precision, Confusion matrix 값 측정 후 전체적인 모듈의 성능 검증 예정

2. 설계 상세화 및 변경 내역

2-1. 데이터 수집

- 585명의 환자에 대한 약 35만 장의 MRI 영상 데이터 수집 완료
- 수집된 전체 데이터를 서버에 업로드 중

2-2. 데이터 전처리

- FreeSurfer 및 FSL을 활용하여 전체적인 데이터 전처리 모듈 개선 예정

2-3. 데이터 분석 및 예측

- CNN 모듈 및 ResNet 모듈 작성 완료
- 학습 후 모델에 대한 accuracy 값과 loss 값 측정
- 학습 데이터 학습 후 테스트 데이터에 대한 메틸화 여부 예측
- 메틸화 확률이 높다고 예측된 순서로 정렬 후 출력

3. 갱신된 과제 추진 계획

5월			6월				7월				8월					9월			
16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	29	5	12	19	26
관련 지식 습득																			
			학습 모델 개발 및 테스트																
									중간 보고서 작성										
										데이터 결과 시각화									
										데이터 전처리 모듈 수정									
										학습 모델 최적화 및 오류 수정									
																최종 테스트			
																최종 보고서 작성			

4. 구성원 별 진척도

이름	역할
이섬재	기본 학습 모델 작성 완료 데이터 결과 시각화 진행 예정 데이터 전처리 모듈 수정 예정
한혜린	학습 모델 테스트 진행 학습 모델 최적화 및 오류 수정 예정
최연희	학습 모델 테스트 진행 학습 모델 최적화 및 오류 수정 예정

5. 보고 시점까지의 과제 수행 내용 및 중간 결과

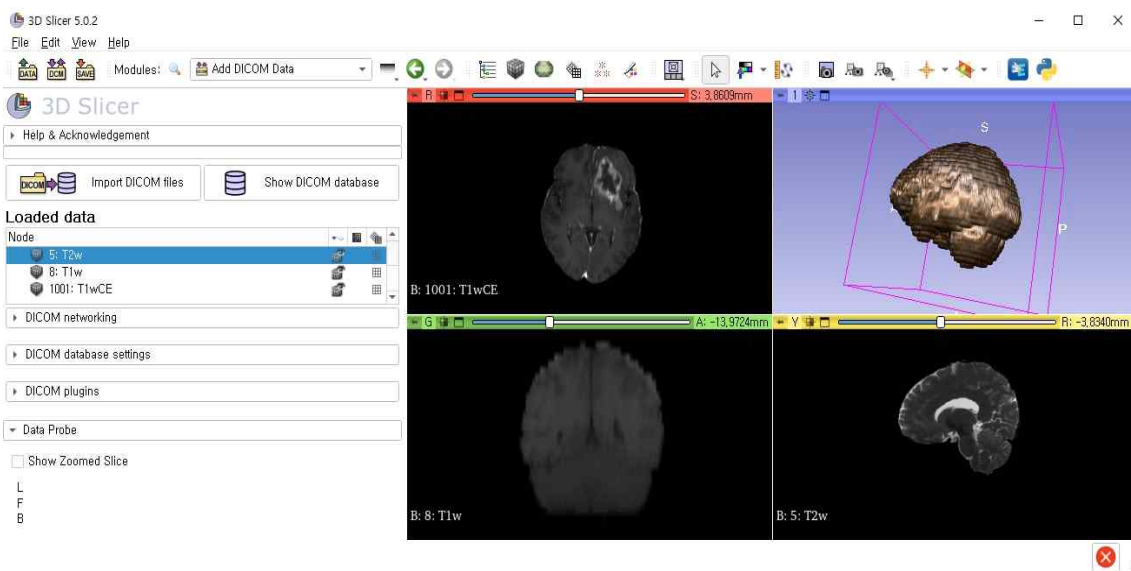
5-1. 데이터 수집



파일 348,642, 폴더 2,927

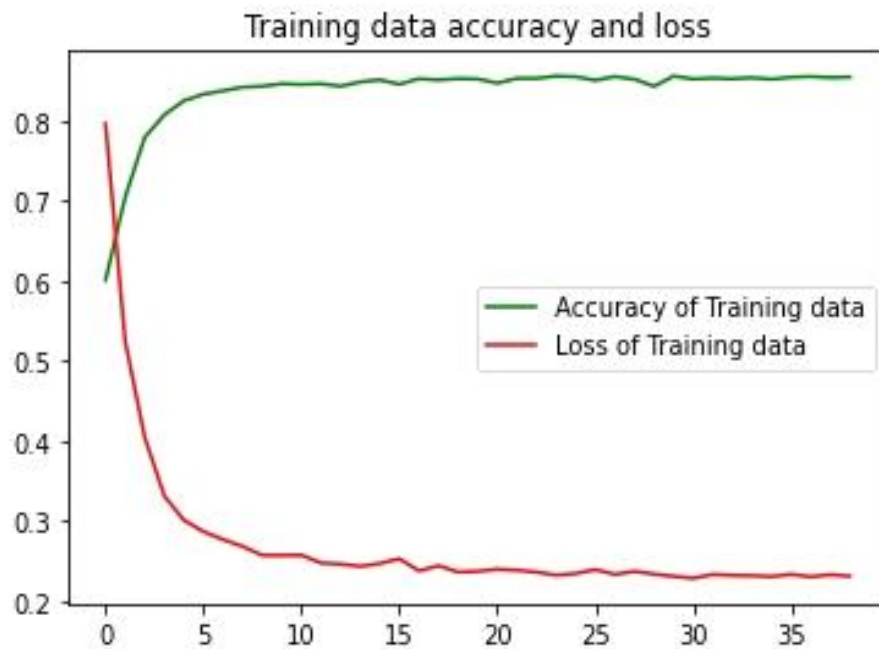
종류: 파일 폴더 형식의 모든 파일
위치: D:\의 모든 파일
크기: 110GB (118,850,629,649 바이트)
디스크 할당 크기: 111GB (119,894,556,672 바이트)

<그림 1> 학습 데이터 개수 및 용량



<그림 2> 3D Slicer를 활용한 영상 데이터 분석

5-2. 데이터 분석 및 예측



<그림 3> CNN 모듈의 accuracy 및 loss 값

	BraTS21ID	MGMT_value
4	37	0.992248
0	1	0.914729
6	79	0.899225
9	91	0.852713
5	47	0.837209
1	13	0.798450
2	15	0.671053
7	80	0.588710
3	27	0.575658
8	82	0.527132

<그림 4> 학습 완료된 모듈의 테스트 데이터 예측 결과