

의료 프로젝트 보고서

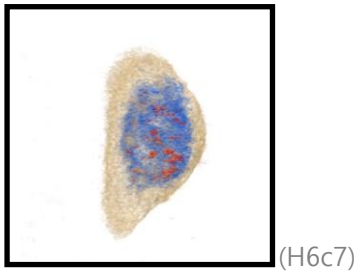
20193238학번 이은서

▪ **실험 목표** 정상 췌장세포와 췌장암 세포 구별

▪ **데이터 구성**

: 네 개의 class의 동영상 데이터와 meta data로 구성.

- 정상 세포: H6c7
- 암 세포 Capan-2, PSN-1, BxPc-3



▪ **실험 과정**

(1) 동영상 파일에서 이미지 프레임 추출

데이터 전처리: 10 초 분량의 원본 영상에서 5 초 분량만 사용해 프레임별로 잘라서 학습에 사용한다. 일정 각도마다 프레임을 뽑아 영상별로 37 장을 추출한다.

(2) 이미지 학습을 위한 딥러닝 모델 구축

네트워크의 깊이(depth), 너비(width), 입력 이미지의 해상도(resolution)를 균형 있게 조절하는 compound scaling이 적용되는 EfficientNet을 사용해 학습하였다.

(3) 교차검증 구현

K-fold cross validation: 데이터를 k개의 부분 집합으로 나누고, 이 중 한 개의 부분 집합을 검증 세트(validation set)로, 나머지 k-1개의 부분 집합을 학습 세트(training set)로 사용하여 모델을 학습시키고 평가한다. 이 과정을 k번 반복하며, 매번 다른 부분 집합을 검증 세트로 선택하는 과정을 거쳤다.

```
for fold in range(args.k_fold):  
    for fold in folds:  
        run(fold, df_train, meta_features, n_meta_features, transforms_train, transforms_val, target_idx)
```

위 과정을 통해 모든 데이터가 적어도 한 번씩은 학습과 검증에 사용되어 데이터를 효율적으로 사용할 수 있고, 여러 번 모델을 검증하면서 과적합을 방지할 수 있다.

(4) meta-data 학습 모델 구현

세포별 두 개의 csv파일을 취합하여 PPT에서 제공된 대로, Concentration와 Surface area 항목의 데이터를 함께 학습에 이용하였다.

(5) Result grouping 구현

환자별 예측 값 37개의 평균을 내어 하나의 결과를 도출하는 과정을 거쳤다.

■ 실험 결과

	Accuracy	AUC_all_raw	AUC_all_rank
Image	95.09866	0.98511	0.97982
Image + meta data	95.26984	0.98837	0.98173
Image + result grouping	97.86667	0.99806	0.97984
Image + meta data + result grouping	97.60000	0.99644	0.98158

▪ 실험 세팅

k-fold: 5, Epochs: 30, image size: 256, batch size: 16, learning rate: 3.00E-05, random seed: 2359

lr_scheduler: CosineAnnealingWarmRestarts, optimizer: Adam, loss: CrossEntropyLoss

1. Image data만 사용한 경우
2. Image data와 각 세포의 Concentration과 surface area항목의 데이터를 함께 학습
3. Image data만 사용한 학습 후 환자의 37장 예측값들을 평균내어 하나의 결과를 도출
4. 2번과 동일하게 학습 후 37장의 예측값들을 평균내어 하나의 결과 도출한 결과

: 결과적으로, Image data와 result grouping을 이용한 방법이 Accuracy, AUC 모두 좋은 성능을 보였다.

* AUC_all_raw와 AUC_all_rank의 차이점

'AUC_all_raw'는 모델이 직접 출력한 예측 확률(0~1)을 사용하는 방법이고, 'AUC_all_rank'는 예측값들을 순위로 변환하여 사용한다. 따라서 순위 기반 AUC는 데이터의 스케일에 덜 민감하며 다양한 데이터 분포에서 일관성을 유지할 수 있다.