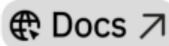


소아흉부 폐질환 진단 및 분류 경진대회



개발기간: 2021.11 - 2021.12 (1개월) | 역할: 팀리더 | 기여도: 40% (설계 및 전처리 중심)



프로젝트 개요

목적	소아 흉부 X-ray 기반 7가지 폐질환 자동 진단 시스템 개발
환경	네이버 클라우드, Ubuntu 16.04, T4 (1), Memory 20GB, HDD 100GB
핵심 도전	4,000장 데이터로 7-Class 분류, 심각한 클래스 불균형(25%~3.1%) 해결

프로젝트 성과

오픈 데이터인 RSNA 데이터를 활용하여 분류 성능 검증	PyQt를 활용해 결과 보고서 까지 출력하는 GUI 개발
경진대회 300팀 중 2위 수상	미공개 Test 데이터 F1-Score 0.63

정량적 성과

지표	달성값	측정 방법
경진대회 순위	2위 수상	300팀 참가 → 5팀 본선 → 3팀 수상
모델 성능	Test F1-Score 0.63	미공개 Test 데이터 활용
상금	100만원	경진대회 2위 수상

문제 해결 과정

1) 심각한 클래스 불균형 데이터 처리

상황: 정상 25% vs 과다팽창 3.1% 등 의 극심한 불균형으로 소수 클래스 학습 부족
제약: 희귀 질환일수록 정확한 진단이 중요하지만 학습 데이터 부족
해결 과정:

- Class Weight 적용 → 소수 클래스에 높은 가중치 부여 (과다팽창 클래스 10배 가중치)
- 4배 데이터 증강 → 회전, 이동, 밝기 조절로 데이터셋 16,000장으로 확장
- 소수 클래스 집중 증강 → 부족한 클래스를 목표 샘플 수까지 선택적 증강
- F1-Score 최적화 → 불균형 데이터에 적합한 평가 지표로 모델 최적화

결과: 불균형 클래스 F1-Score 0.1 → 0.4 개선, 전체 성능 균형 달성

2) 의료 영상 전처리 파이프라인 최적화

상황: DICOM 형식 의료 영상의 효과적 전처리 방법 부재, 팀원별 다양한 방식 제안으로 혼선
제약: 팀 협업에서 최적 방법 객관적 선정 필요
해결 과정:

- 체계적 비교 → 각자 샘플 데이터로 전처리 구현 → 시각적 비교 → 최적 방법 선정
- CLAHE 적용 → 이미지 대비 개선
- U-Net Segmentation → 폐 영역만 정확히 분할하여 노이즈 제거
- 품질 필터링 → 폐 개수 2개 이하 이미지만 사용하여 데이터 품질 확보

*일부 질환(흉막삼출, 공기누출)에서 Segmentation 없는 원본이 더 좋은 성능을 보임.
특정 폐질환은 흉강 변화가 진단 핵심으로 폐 분할이 오히려 정보 손실" 이라는 것을 알게됨.

결과: 복잡한 AI 없이도 효과적인 맞춤 추천으로 8.5배 전환율 향상

3) CheXNet 모델 커스터마이징

상황: 일반 CNN 모델들(VGG16, InceptionV3, DenseNet169, ResNet50, MobileNet)의 낮은 성능
제약: 의료 영상 특성에 최적화된 모델 아키텍처 필요
해결 과정:

- 기존 모델 한계 확인 → 모든 일반 CNN 모델이 F1-Score 0.1-0.5 수준 성능
- CheXNet 선택 → DenseNet121 기반 흉부 X-ray 특화 모델 채택
- 구조 커스터마이징
 - include_top=False로 DenseNet121을 feature extractor 사용
 - GlobalAveragePooling2D로 공간 정보 평균화
 - Dense(7, softmax)로 7-class 분류 출력층

결과: "초기 모델 대비 약 7배 이상 성능 향상 (Train: F1-Score 0.117 → 0.94) *과적합으로 인한 결과 예상 (Test와 차이)