

MediCRAG: 검색증강생성기반의로이미지분류성능향상

연세대학교 컴퓨터과학과 권세인

2024년 12월



과제명: IoT 환경을 위한 고성능 플래시 메모리 스토리지 기반 인메모리 분산 DBMS 연구개발

과제번호: 2017-0-00477



과학기술정보통신부
Ministry of Science and ICT



연세대학교
YONSEI UNIVERSITY



정보통신기술진흥센터
Institute for information & communications Technology Promotion

MediCRAG: 검색 증강 생성 기반 의료 이미지 분류 성능 향상



Index

01 서론

02 시스템 구조

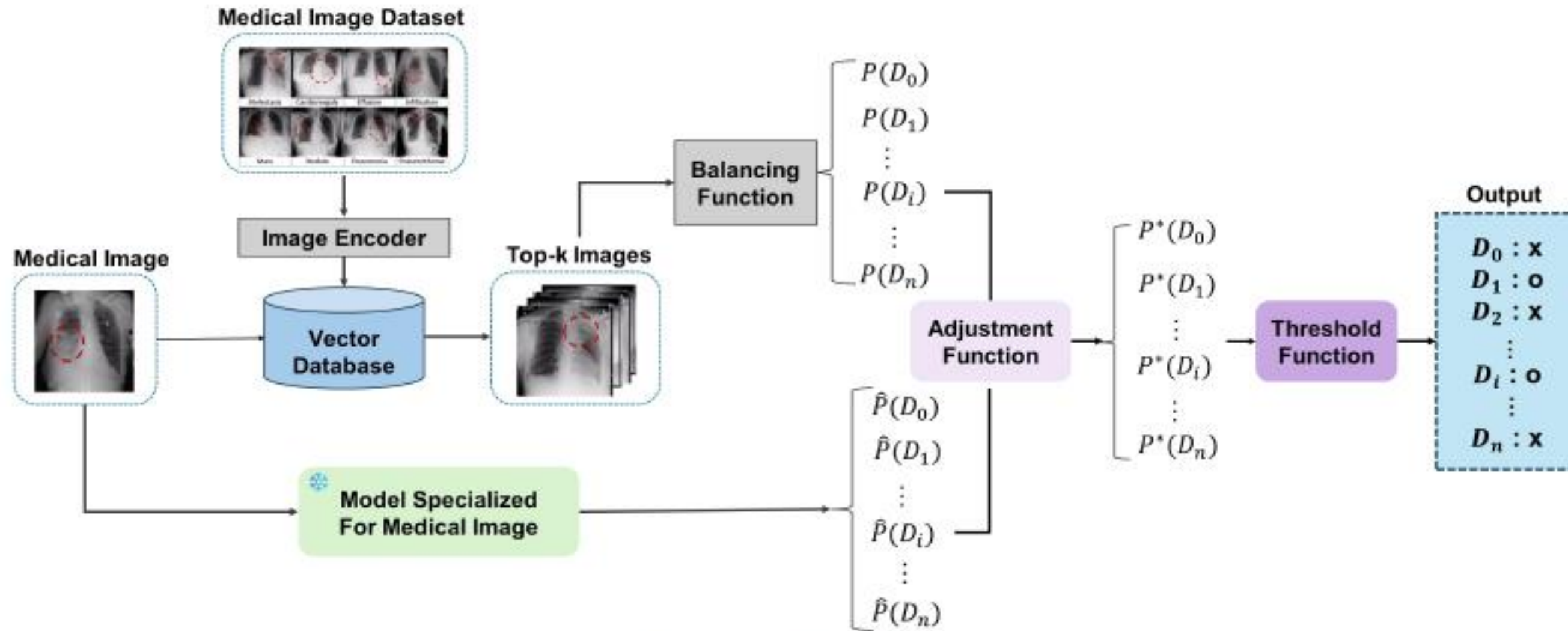
03 실험 및 결과

04 결론

1. 서론

- 의료 이미지에 대해 사전 학습된 딥러닝 모델: 질병 분류 및 병변 탐지와 같은 작업 수행
- **한계점**
 - 의료 데이터 불균형 문제로 발생 빈도가 낮은 질병 과소평가될 수 있음
 - 모델 예측 근거 제공의 어려움
 - 의료 도메인에 특화된 오픈소스 기반 검색 시스템의 부족으로 실제 의료진이 기술을 효과적으로 활용하기 어려움
- **RAG(Retrieval-Augmented Generation)**
 - 외부 지식 베이스나 벡터데이터베이스에서 유사한 정보나 사례를 검색
 - 최신 정보와 전문 지식을 모델 생성 과정에 통합하여 정확성과 신뢰성을 높임
 - RAG를 의료 이미지에 특화된 모델에 적용한 검색 증강 분류 시스템인 MediCrag 제안
- **논문의 주요 기여**
 - 의료 이미지 도메인에 특화된 오픈소스 모델과 의료 데이터를 활용하여, 질병 분류의 정확도를 높이는 검색 증강 시스템을 제안함
 - 벡터데이터베이스를 활용한 유사 의료 이미지 검색과 가중치 기반 조정 방식을 도입해 데이터 불균형 문제로 인한 모델 성능 저하를 해결함
 - 모델 단독 사용과 MediCrag 시스템의 비교 실험을 통해, MediCrag가 분류 정확도를 유의미하게 개선함을 확인함

2. 시스템 구조



< 그림 1. 시스템 구조 >

2. 시스템 구조

2.1 쿼리 이미지의 유사한 이미지 검색

- 흉부 X-ray 이미지 데이터셋이 XRV-ResNet50 모델의 이미지 인코더를 통해 feature vector로 변환되고, Milvus 벡터 데이터베이스에 저장됨
- 새로운 흉부 X-ray 이미지가 입력된 경우
 - 1) 이미지는 XRV-ResNet50 이미지 인코더를 통해 특징 벡터로 변환된 후 Milvus에 전달됨
 - 2) Milvus는 벡터 간 Euclidean 거리 유사도를 계산하고 유사도가 가장 높은 상위 k개 이미지를 반환함
- 검색된 이미지는 의료 이미지 특화 모델의 예측 확률을 보완하여 모델 분류 성능을 향상시키는데 활용됨

2.2 Top k 유사 이미지의 질병 클래스 비율 추출

- MediCrag은 의료 이미지 특화 모델이 단일 쿼리 이미지와 유사 사례 기반 병명 분포를 함께 고려하게 하여 예측의 신뢰도 높임
- 질병 간 데이터 불균형으로 데이터셋 내 비중이 낮은 질병이 과소평가될 수 있음
- 이를 보완하기 위해 데이터셋 내 클래스 비중에 반비례하는 가중치 w_i 적용 (수식 (1))
 - N : 전체 이미지 수, n_i : 질병 클래스 D_i 에 속하는 이미지 수

$$w_i = \frac{1}{n_i/N} \quad (1)$$

- 수식 (1)을 적용하여, 데이터셋 내 비중이 적은 클래스일수록 높은 가중치 부여함
- 검색된 유사 이미지의 각 클래스 빈도 계산 시, 단순 개수 대신 w_i 가 곱해진 빈도 사용
- 질병 클래스 비율 $P(D_i)$: 모델의 분류 예측 확률을 조정하는 추가적인 정보로 사용됨

2. 시스템 구조

2.4 유사 이미지 비율에 따른 예측 조정

- XRV-DenseNet121 모델의 초기 확률은 벡터 데이터베이스에서 검색된 유사 이미지의 질병 비율에 따라 조정 함수(*Adjustment Function*)을 통해 조정됨
- **Adjustment Function** (수식 (2))

$$P^* = \hat{P} \times (1 + g(r) * C_f) \quad (2)$$

- 조정 함수 $s(r)$: 검색된 유사 이미지 비율 r 에 따라 동적으로 정의됨 (수식 (3))

$$s(r) = f(T_l, P) \quad (3)$$

- T_1 이상이면 모델의 초기 예측 확률을 크게 증가시키고, T_2 이하이면 모델의 초기 예측 확률을 크게 감소시키는 방식으로 조정함

2.5 임계값 비교를 통한 질병 유무 판단

- 각 질병 클래스에 대해 조정된 확률 $P^*(D_i)$
 - 임계값 함수(*Threshold Function*)를 거쳐 최종 클래스 분류 이루어짐
 - 임계값(Threshold): XRV-DenseNet121의 학습 과정에서 최적화된 값
 - 각 클래스의 임계값은 ROC(Receiver Operating Characteristic) 곡선에서 최적 분류 지점을 기준으로 설정됨
- 조정된 확률 $P^*(D_i)$ 가 임계값 초과하면 해당 질병을 양성(Positive)으로 판별함

3. 실험 및 결과

3.1 실험 데이터셋

- NIH ChestX-ray14 데이터셋을 활용하여 실험 진행
 - 해당 데이터셋은 미국 국립보건원(NIH) 클리니컬 센터의 PACS 데이터베이스에서 추출된 이미지들, 14가지 흉부 질환에 대한 다중 레이블이 포함됨
- 데이터셋의 테스트 데이터셋 중 5000개의 샘플을 무작위로 선택하여 모델의 성능 평가

3.2 실험 결과

- 의료 이미지 특화 모델의 단독 사용과 MediCRAG의 성능을 비교
 - 1) XRV-DenseNet121의 초기 예측 확률에 기반한 질병 분류를 수행
 - 2) MediCRAG를 적용하여 조정된 확률을 사용한 질병 분류를 수행
 - 3) 두 모델의 성능을 비교하기 위해, 각 이미지에 대해 14개 클래스의 예측 결과와 실제 라벨을 비교하고 정확히 예측된 클래스 수를 누적 합산하여 최종 정확도를 산출
- 실험 결과
 - 기본 모델을 사용한 경우 정확도(Accuracy)는 0.70으로 측정됨
 - MediCRAG를 적용한 경우 정확도가 0.85로 향상됨
 - 재현율(Recall)과 특이도(Specificity) 값을 측정한 결과, 재현율은 0.30에서 0.32로 소폭 향상되었으며, 특이도는 0.71에서 0.88로 크게 개선됨

< 표 1. 실험 결과 >

| Model | Accuracy | Recall | Specificity |
|-----------------|----------|--------|-------------|
| XRV-DenseNet121 | 0.70 | 0.30 | 0.71 |
| MediCRAG | 0.85 | 0.32 | 0.88 |

4. 결론

- 의료 데이터 불균형 문제를 해결하고, 의료 이미지 특화 모델의 성능을 향상시키기 위해 RAG을 도입한 MediCRAG를 제안함
- 실험 결과
 - 의료 이미지 특화 모델을 단독으로 사용한 경우보다 MediCRAG를 적용한 경우 분류 정확도가 유의미하게 개선되었음을 확인함
 - 유사 이미지 검색을 통한 확률 보정이 의료 이미지 특화 모델의 예측 성능을 보완하여, 의료 도메인에서 보다 신뢰성 있는 의사결정을 가능하게 함을 시사함
 - 현재 모델의 재현율 향상이 뚜렷하지 않은 점에 관하여 향후 연구에서는 양성 사례의 재현율 향상을 위해 데이터의 다양성과 유사 이미지 검색 알고리즘을 최적화하는 방안을 탐색할 예정