**Introduction**

의료 영상 데이터를 위한 국제 표준 DICOM은 메타데이터와 이미지데이터를 동시에 보유

* 메타데이터는 매우 불완전하며 값을 기술하기 위해 정해진 표준화된 형식이 없음
* 따라서 연구자들이 데이터베이스에서 필요한 데이터를 추출하는데 비효율 초래

다기관 데이터의 통합 및 효율적인 재사용을 위한 R-CDM이라는 새로운 의료 영상 데이터 표준화 기법 개발, 적용

**Materials & Methods**

가명화된 연구 데이터 준비

* 의료 영상 데이터는 환자의 개인정보를 포함
* 연구에 관련 없는 제3자인 의료 빅데이터 센터가 익명화 과정을 수행한 자료를 수령하여 연구를 진행
* 개인정보를 포함하는 메타데이터 12개 항목 삭제, 환자번호 대체
* 아주대병원 의료 영상 백업 데이터베이스에서 ~에서 ~동안 촬영된 ~TB에 해당하는 데이터를 가명화하여 준비

아주대병원 의료 영상 데이터의 문제점 파악

* 메타데이터 값을 기술하는 용어가 표준화되어 있지 않음
* 메타데이터에 틀린 정보들이 기록되어 있는 경우가 있음
* 특정 종류의 정보가 다른 메타데이터에 기록되어 있는 경우
* 중요한 의학 정보가 메타데이터에 기록되지 않은 경우

R-CDM 표준화 용어 시스템 설계

* RadLex 소개
* RadLex와 LOINC, OMOP vocabulary의 결합을 통한 국제적 표준화 달성

R-CDM 표준화 구조 설계

* 2개의 테이블 설계
* OMOP-CDM과의 연결

아주대병원 의료 영상 데이터의 R-CDM 변환

* 의료 영상 데이터의 표준화 변환 과정의 설명
  + 메타데이터 ETL -> 딥러닝 이미지 분류기 적용
* 메타데이터를 활용한 ETL
  + 유의미한 정보를 지닌 메타데이터 16개 항목 결정 후 기록된 값들을 추출, 추출된 값을 ETL 하여 R-CDM의 용어와 구조로 표준화한 다음 데이터베이스에 적재
  + 메타데이터 ETL의 한계점 (오류, 결측치 존재 및 세부 정보 추출 불가)
* ETL의 정확도, 정밀도를 높이기 위한 딥러닝 이미지 분류기 개발 및 적용
  + Brain CT 딥러닝 이미지 분류기가 가지는 기능
  + 데이터 전처리, ResNet 활용한 multiclass classification, 분류 결과 데이터베이스 업데이트 자동화 파이프라인 구축
  + 데이터 준비 과정

**Results**

메타데이터 ETL의 정확도

* 촬영 기기 종류, 촬영 기기 제조 회사, 이미지 해상도에 관한 정보는 100% 매핑
* 촬영한 신체 부위에 대한 정보는 ~% 정확도로 매핑
* 조영제 투여 상황, 이미지 촬영 방향에 대한 매핑은 불가

이미지 분류기의 분류 정확도

* test set을 100%의 정확도로 분류
* F1 score ~점
* ~장의 Brain CT의 촬영 방향 및 조영제 투여 상태 예측, Image table의 Image Occurrence ID에 적재
* Image clustering 결과

OMOP-CDM과 R-CDM 연동 예시

* 응급실로 내원하여 당일 Brain CT를 촬영한 환자의 pre contrast image 추출

**Conclusion**

OMOP-CDM의 확장 모델인 R-CDM은 의료 영상 데이터의 구조와 용어를 표준화

* DICOM 메타데이터의 ETL
* 이미지 딥러닝 분류기의 적용을 통한 ETL의 정확도, 정밀도 향상

R-CDM과 OMOP-CDM의 연결을 통한 의료 영상 데이터와 임상 데이터의 효율적 연계

R-CDM을 통한 다기관의 대규모 의료 영상 데이터의 효율적 확보로 해당 분야의 딥러닝 연구 촉진 기대