

기계학습을 위한 의료영상 데이터 표준화 및 응용 소프트웨어

김지언*, 한성민*, 박민기*, 김승진*, 노시형*, 전홍영*, 이충섭*, 김태훈*, 정창원**

*원광대학교 의료융합연구센터

e-mail:{kakasky112, zhsk528, pmg0612, koch369369, nosij123, zip80, cslee99,tae_hoonkim, mediblue}@wku.ac.kr

Medical Image Data Standardization for Machine Learning and Its Application Software

Ji-Eon Kim*, SeongMin Han*, Minki Park*, Seung-Jin Kim*, Si-Hyeong No*, Hong-Yong Jun*, Chung Sub Lee*, Tae-Hoon Kim*, Chang-Won Jeong**

*Medical Convergence Research Center, Wonkwang University

요 약

의료영상은 환자의 질병을 진단하고 치료방침을 결정하는데 중요한 도구로 자리매김하고 있다. 최근 의료영상을 인공지능 연구가 국내외에서 활발하게 진행되고 있다. 특히 대규모의 의료영상들을 학습시켜 질병과 상태를 정밀 진단할 뿐만 아니라 예측하는 소프트웨어를 개발 하는 상황이다. 그러나 의료영상은 DICOM 표준에 따르고 있지만 태그정보의 사용은 의료기기과 의료기관마다 상이하다. 따라서 의료영상에 대한 메타 데이터의 표준화에 어려움이 있다.

본 논문은 이러한 의료영상 데이터를 표준화 할 수 있는 방법을 제안한다. 그리고 제안한 표준화 데이터로 변환할 수 있는 ETL 소프트웨어의 수행결과를 보이고, 조건에 따라 머신러닝 학습 데이터셋을 생성하는 결과를 제공한다. 향후 제안한 의료영상 표준화와 ETL 소프트웨어는 다양한 수요자 중심의 표준화된 데이터셋을 제공할 수 있는 플랫폼의 주요기능으로 활용 될 것으로 기대한다.

1. 서론

의료분야에서 사용되는 의료영상은 환자의 질환을 정확하게 진단하고 치료 방침을 결정하기 위해 매우 중요한 진단 도구로 사용되고 있다. 이러한 의료영상들은 DICOM (Digital Imaging and Communications in Medicine) 국제 표준을 준수하여 저장되며 일반 영상파일과 달리 DICOM header 부분에 의료영상정보와 함께 저장하고 있다.

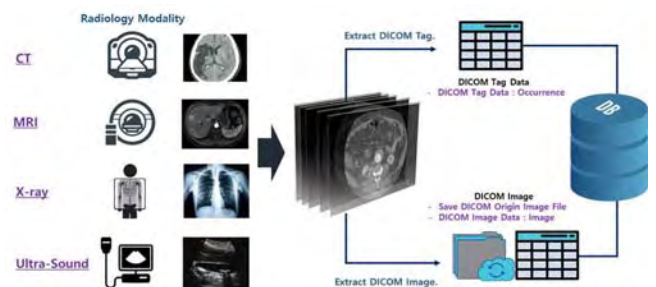
현재 환자를 대상으로 의료기관에서 수집되고 있는 의료영상은 CT, MRI 등 영상장비를 사용하여 각 질환 부위별로 생성된 영상들이다. 최근, 의료영상을 기반으로 기계학습을 통해 질병의 정밀진단 뿐만 아니라 조기 발병 진단과 예측하는 연구가 활발히 진행되고 있다. 그러나 수집된 의료영상을 이용하여 인공지능 학습 연구에 곧바로 적용하기 위해서는 해결해야 할 점들이 많이 남아 있다[1]. 최근, 의료영상을 기반으로 한 인공지능 학습을 위해서는 많은 양의 의료영상 데이터가 필요하다. 또한 각 병원에서는 의료영상을 DICOM 국제 표준을 준수하여 저장하고 있으나 각 질환별로 최적화된 임상프로토콜에 대한 선별, 핵심적인 의료영상에 저장되는 의료정보까지 표준화되어 저장되어야 한다. 이와 관련하여 수행된 연구는 국내외 미

흡한 실정이며 더욱이 의료기관별 의료영상의 표준화된 정보 없이 인공지능 학습 연구에 적용하기에는 어려움이 있다[2,3]. 따라서, 본 논문에서 제안하는 소프트웨어를 통해 수집되는 의료영상 데이터를 표준화 할 수 있는 방법을 제안하고 ETL 정의에 따라 표준화된 데이터로 변환할 수 있는 소프트웨어를 개발하고자 한다.

2. 시스템 제안

2.1 전체 시스템 구조

본 논문에서 제안하는 시스템 구조는 그림 1과 같다. CT, MRI, X-RAY, Ultra-Sound 등 의료 영상장비에서 생성된 의료영상데이터를 기반으로 표준화된 의료영상 데이터로 변환하여 DB에 저장한다.



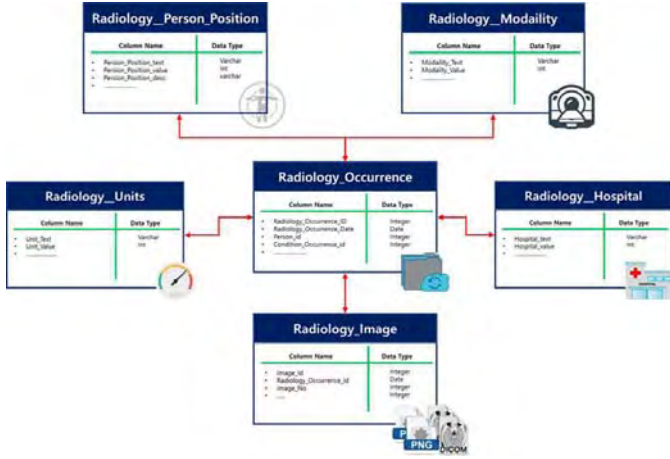
(그림 1) 의료영상 표준화 처리과정

* Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDD), funded by the Ministry of Health & Welfare (HI18C1216).

** 본 연구는 한국산업기술평가관리원의 바이오산업핵심기술개발사업(바이오핵심기술개발)의 일환으로 수행하였음(20001234, 선행공동데이터모델기반 분산형 바이오헬스 통합 데이터망 구축 기술개발)

2.2 ERD

그림 2는 제안하는 의료영상 표준화 방법에 대한 ERD 구조를 나타냈다. 표준화 의료영상정보는 저장된 DICOM Tag 정보를 추출하여 그림과 같이 표준화한다. 의료영상에 저장된 DICOM Tag 정보를 기반으로 표준화하기 위해 의료영상에 저장된 데이터뿐만 아니라 각 테이블에 표준화된 표준 데이터를 참조하여 데이터를 분류하여 저장한다.

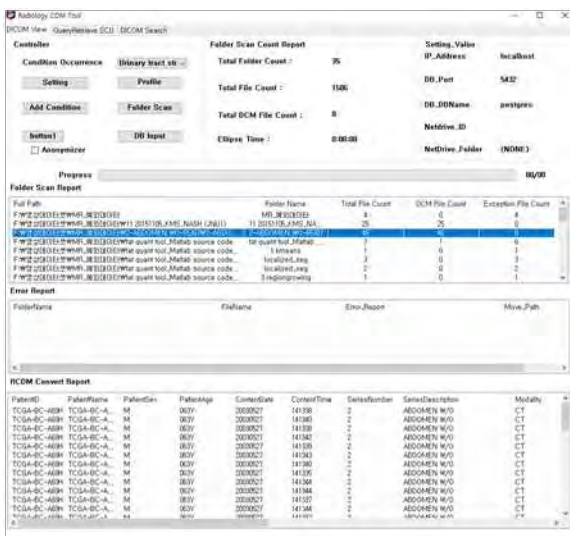


(그림 2) 의료영상 표준화 데이터의 ERD

3. 시스템 구축 결과

3.1 표준화 데이터 셋 변환 시스템 구축 결과

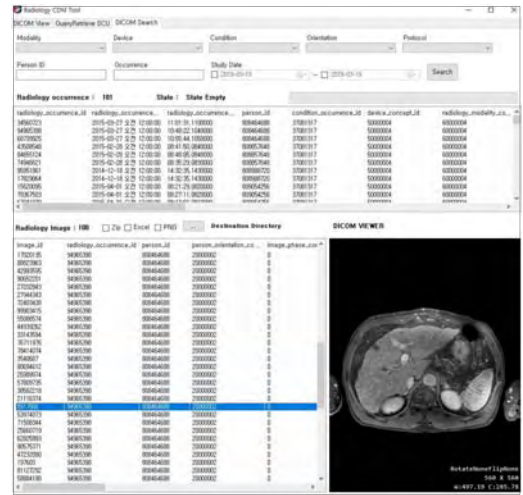
그림 3은 수집된 의료영상을 표준화된 데이터로 변환하여 저장하기 위한 시스템을 나타낸 것이다. 표준화 데이터 변환 시스템은 지정된 로컬에 저장되어있는 모든 의료영상데이터를 기준에 따라 분류할 수 있다. 데이터를 표준화하는 과정에서 생기는 문제점은 Report로 제공하기 때문에 데이터 변환 과정을 실시간으로 모니터링 할 수 있다.



(그림 3) 의료영상 표준화 소프트웨어 구현 결과

3.2 표준화 데이터 셋 검색 시스템 구축 결과

그림 4는 DB에 저장된 표준화된 의료영상데이터를 검색하여 나타낸 화면이다. 사용자가 원하는 데이터셋 검색 조건을 설정하여 검색이 가능하며, 검색 조건에 따라 원하는 이미지 정보와 표준화 데이터로 변환된 의료영상정보 및 이미지정보를 함께 확인할 수 있다.



(그림 4) 의료영상 표준화 소프트웨어 검색 결과

4. 결론

본 논문에서는 의료영상정보에 대한 표준화하는 방법을 적용한 시스템을 구현하였다. 제안하는 시스템을 활용하면, 의료영상 기반 임상연구를 위해 중요한 임상적 정보들을 용이하게 추출하여 활용할 수 있을 것이다. 또한 의료영상 표준화 시스템을 통해 의료영상 뿐만 아니라 의료정보도 함께 표준화된 데이터로 변환이 가능하기 때문에 다양한 임상연구들을 지원할 수 있는 기반이 될 것으로 기대한다. 뿐만 아니라 최근 관심이 높아지고 있는 인공지능 학습 연구에 표준화된 영상 데이터 셋을 제공하는데 적극 활용할 수 있을 것으로 확신한다. 향후 우리의 시스템을 통해 표준화된 질환별 데이터를 구축할 예정이며 의료영상 기반의 인공지능 연구와 임상 연구에 응용하고자 한다.

참고문헌

[1] KOTADIYA, Hiral; PATEL, Darshana. "Review of Medical Image Classification Techniques" In: Third International Congress on Information and Communication Technology. Springer, Singapore, 2019. p. 361-369.
 [2] Adrian V. Dalca, Katherine L. Bouman, William T. Freeman, Natalia S. Rost, Mert R. Sabuncu, Polina Golland, "Medical Image Imputation From Image Collections", IEEE transactions on medical imaging, 2019, 38.2: 504-514.
 [3] J Zhang, Y Xie, Q Wu, Y Xia, "Medical Image Classification Using Synergic Deep Learning ", Medical image analysis, 2019.