

머신러닝을 위한 의료영상기반 학습 데이터 지원 플랫폼 구축 및 근감소증 데이터 AI 응용

김지연¹, 임동욱¹, 유영주¹, 노시형¹, 이충섭¹, 김태훈^{1,2}, 정창원^{1,2}

¹원광대학교 의료융합연구센터

²원광대학교 병원 스마트사업팀

e-mail:{kakasky112, dw316, yeriel.ryu, nosij123, cslee99, tae_hoonkim, mediblue}@wku.ac.kr

Construction of Medical Image-Based Learning Data Support Platform for Machine Learning and Its Application of Sarcopenia Data AI

Ji-Eon Kim¹, Dong Wook Lim¹, Yeong Ju Yu¹, Si-Hyeong Noh¹,
ChungSub Lee¹, Tae-Hoon Kim^{1,2}, Chang-Won Jeong^{1,2}

¹Medical Convergence Research Center, Wonkwang University

²Smart Business Team, Wonkwang University Hospital

요 약

의료산업은 진단 및 치료 위주의 기술개발이 진행되어왔다. 최근 의료 빅데이터를 기반으로 진단, 치료 및 재활뿐만 아니라 예방과 예후관리까지 지원하는 의료서비스에 대한 패러다임이 변화되고 있다. 특히, 여러 의료 중심의 플랫폼 기술 가운데 객관적인 진단지표를 가지고 있는 의료영상을 기반으로 인공지능 학습에 적용하여 진단 및 예측을 중심으로 한 플랫폼 개발이 진행되고 있다. 하지만, 인공지능 연구에는 많은 학습 데이터가 요구될 뿐만 아니라 학습에 적용하기 위해서는 데이터 특성에 따른 전처리 기술과 분류 작업에 많은 시간 소요되어 이와 같은 문제점을 해결할 수 있는 방법들이 요구되고 있다. 따라서, 본 논문은 인공지능 학습까지 적용하기 위한 의료영상 데이터에 대한 확장 모델을 개발하여 공통적인 조건에 따라 의료영상 데이터가 표준화되어 변환하며, 자동화 시스템 구조에 따라 데이터가 분류·저장되어 인공지능 학습까지 지원할 수 있는 플랫폼을 제안하고자 한다. 그리고 근감소증 학습데이터 관리 및 적용 결과를 통해 플랫폼의 수행성을 검증하였다. 향후 제안한 플랫폼을 통해 의료데이터에 대한 전처리, 분류, 관리까지 지원함으로써 CDM 확장 표준 의료데이터 플랫폼으로 활용 가능성을 보였다.

1. 서론

과거 의료기술 산업은 질환을 진단하고 치료를 중점적으로 연구되었다. 그러나, 현재 4차 산업의 시대 변화로 의료기술 산업의 패러다임은 병원에 축적된 환자 데이터(의료영상, 생체신호, 유전체 등)와 데이터를 학습하고 특정 패턴을 분석하고 인식하기 용이한 인공지능과 같은 첨단기술이 융합되어 유의미한 임상적 결과를 나타낸다고 보고되고 있다. 의료데이터는 인공지능 기술과 밀접하게 관련되어 연구와 기술개발이 활발하게 진행되고 있다[1,2]. 특히, 의료데이터를 기반으로 한 인공지능 연구 가운데 가장 객관적이고 특정 패턴을 분석하는데 가장 많은 정보를 포함하고 있는 의료영상 데이터가 인공지능 연구에

사용되고 있다. 하지만, 이와 같은 인공지능 학습까지 적용되기 위해서는 선행적으로 해결해야 하는 문제점들이 있다[3]. 먼저, 인공지능 모델을 개발하기 위해서는 많은 의료영상 데이터가 필요할 뿐만 아니라 의료영상 장비 특성에 따라 전처리 과정과 인공지능 데이터 분류에 많은 시간이 요구된다. 또한, 의료 데이터를 전처리, 분류하기 위한 기준점이 모호할 뿐만 아니라 인공지능 학습까지 지원할 수 있는 플랫폼이 필요하다.

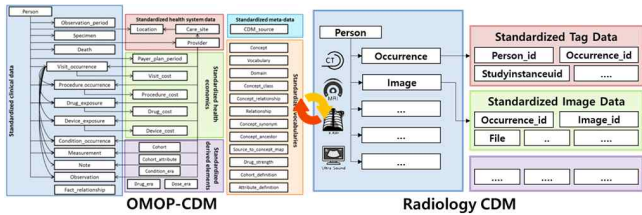
따라서, 본 논문은 의료영상 데이터를 표준화된 기준에 따라 변환하기 위해 OMOP_CDM 기준에 따라 의료영상 데이터에 대한 확장 모델을 개발하여 의료영상 데이터를 공통데이터 모델로 변환 및 관리가 가능할 뿐만 아니라 인공지능 학습에 필요한 데이터셋을 지원 및 결과까지 확인할 수 있는 플랫폼을 제안하고자 한다.

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술 연구개발사업(HI18C1216) 그리고 한국연구재단 (2020R111A1A01074256)(2021R1A5A8029876) 지원에 의하여 이루어진 것임.

2. 제안한 시스템

2.1 의료영상 확장 모델

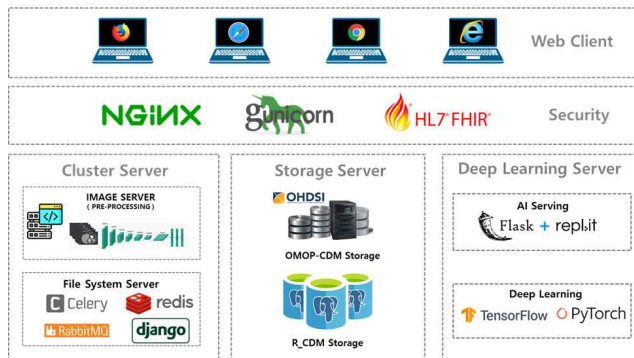
OMOP_CDM은 의료데이터의 공통데이터모델의 대표적인 표준화 모델이다. 그림 1과 같이 의료영상 데이터를 변환하기 위해 의료용어 표준과 공통데이터 모델과 연계를 고려하여 의료영상 표준화 모델인 R_CDM(Radiology Common Data Model)을 정의하였다. 그리고 머신러닝을 위한 의료영상데이터의 표준화 데이터셋 구성에 필요한 메타데이터의 ETL 정의와 Auto ETL 프로그램을 개발하였다.



(그림 1) R_CDM Model Structure

2.2 시스템 소프트웨어 아키텍처

본 논문에서 제안하는 웹 기반 통합 플랫폼 구조는 그림 2와 같다. 접근성이 용이한 웹 기술로 구성되어 있어 의료영상데이터 변환 및 관리 그리고 인공지능까지 적용된 결과를 확인할 수 있다. 그리고 기존에 의료영상 데이터의 용량이 매우 커서 변환하는 시간이 많이 소요되었으나 본 논문에서 제안하는 플랫폼은 소규모 및 대용량 의료영상 데이터에 대한 트래픽을 안정적으로 처리할 수 있도록 서버 구성을 하였을 뿐만 아니라 환자 개인정보와 서버 전달과정에서의 민감 정보 유출문제를 해결하기 위한 보안 요소로 HL7(Health Level 7)에 기준에 의한 의료데이터에 적용하였다.



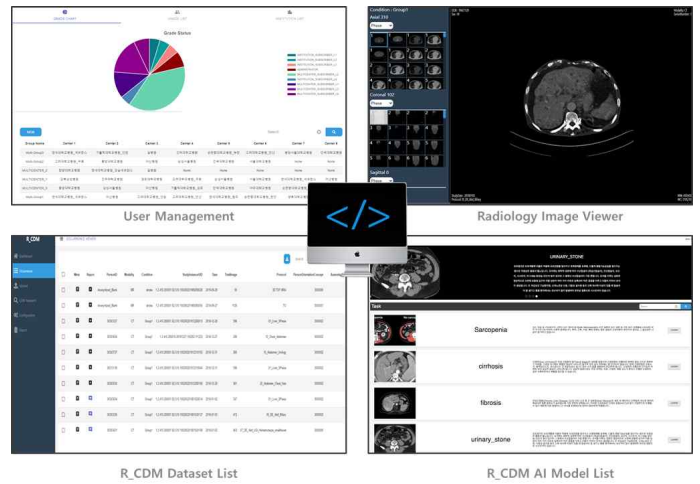
(그림 2) System Architecture

또한, 의료영상데이터를 변환 및 관리할 뿐만 아니라 인공지능 서버를 통해 표준 의료영상 데이터셋으로 변환된 데이터를 통해 인공지능 학습과 학습결과 까지 확인이 가능하다.

3. 시스템 구현 결과

3.1 웹 플랫폼기반 데이터 관리

웹 기반의 데이터 관리를 위한 인터페이스는 그림 3과 같다. 개발한 내용 가운데 사용자 관리는 플랫폼을 사용하는 개인연구자를 위한 사용자와 다기관 공동연구를 위해서 병원 기관을 관리할 수 있다. 또한, 표준 데이터셋 변환 결과는 R_CDM Dataset List로부터 확인할 수 있으며 Radiology Image Viewer를 통해 이미지 정보를 자세히 확인할 수 있다. 이와 관련하여, 기본적인 이미지 영상 처리가 가능하기 때문에 다양한 관점에서 질환을 관독할 수 있으며, 질환 관독 결과에 따라 인공지능 학습에 적용 가능 여부는 R_CDM AI Model LIST를 통해 확인할 수 있다.



(그림 3) Web Platform Develop Result

제안하는 플랫폼은 데이터셋을 관리하는 방법이 가장 중요하기 때문에 데이터셋의 현황과 데이터셋 사용에 대한 로그를 확인할 수 있도록 Dataset Report 기능을 그림 4와 같이 포함하고 있다.



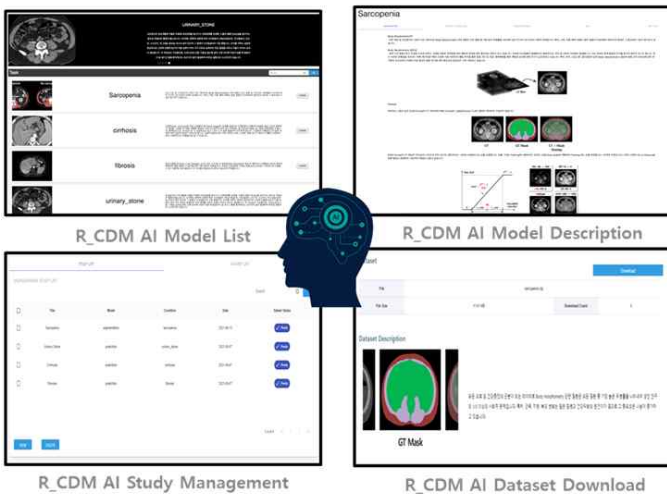
(그림 4) Dataset Activity Report

Dataset Report 주요 기능 가운데 Weekly Dataset Report는 주간 데이터셋 변환 및 다운로드 된 데이터셋에 대한 통계 정보를 확인할 수 있다. 또한, Condition Dataset Report는 인공지능 기능과 밀접한 기능으로서 인공지능 학습과 밀접한 데이터셋 정보를 확인할 수 있다. 마지막으로, 저장된 데이터셋에 대한 출처 및 공유 정보는 Upload / Download Dataset Report를 통해 유지정보, 변환결과 및 공유 데이터셋 타입 등 자세하게 정보를 확인할 수 있다.

3.2 인공지능 학습 지원

본 논문에서 제안하는 인공지능 학습 지원 인터페이스는 그림 5와 같다. 표준 데이터셋을 기반으로 개발된 다양한 의료영상 기반의 인공지능 모델을 시연할 수 있다.

R_CDM AI Platform



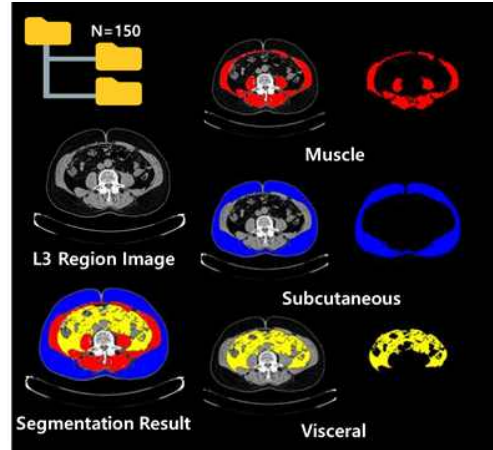
(그림 5) R_CDM AI Platform Result

또한, R_CDM AI Model Description을 통해 인공지능 모델 설계, 데이터셋 모델 구성 및 학습 결과 등 자세한 설명을 확인할 수 있어 표준화된 데이터셋에 대한 중요성과 신뢰성있는 인공지능 모델을 제공하고자 하였다.

3.3 근감소증 학습데이터 적용 결과

인공지능 모델 가운데 근감소증의 L3레벨의 진단 지표영역을 영상분할 모델에 적용하였다. 이에 환자 150명에 대한 의료영상 데이터를 표준화된 데이터셋으로 변환하여 근감소증에 대한 인공지능 모델을 그림 5와 같이 적용하여 수행 결과를 확인하였다. 개발한 모델 결과를 확인한 결과 질환 진단에 필요한 진단지표 영역인 근육, 피하지방 및 내장지방을 분

류한 결과를 확인하였다.



(그림 6) Sarcopenia Segmentation Result

4. 결론

의료빅데이터는 많은 학습 데이터가 필요한 인공지능 기술과 밀접하게 연구되고 있다. 하지만, 기존의 의료빅데이터는 서로 상이한 조건에 따라 저장된 데이터가 대부분이며 표준화 데이터로 변환하기 위한 기준점이 모호한 실정이었다. 본 논문은 의료데이터 가운데 가장 객관적인 지표를 가지고 있는 의료영상데이터를 의료데이터 표준화 방법에 대해 가장 활발하게 사용되고 있는 OMOP-CDM 기준에 따라 의료영상 데이터를 표준화 할 수 있는 모델을 정의하고 Auto ETL 프로그램을 개발하여 플랫폼에 적용하였다. 특히, 사용자의 접근성이 용이하고 다양한 플랫폼과 연동이 가능한 웹 기술을 이용하여 의료영상데이터를 표준화할 뿐만 아니라 관리 그리고 인공지능 기술과 적용할 수 있도록 개발하여 다양한 의료영상 기반의 인공지능 연구를 지원 할 수 있도록 개발하였다. 향후, 다양한 다기관 실증 지원 연구를 통해 개발된 플랫폼의 기능 개선과 활용성을 높일 수 있도록 OHDSI 네트워크와 연계할 계획이다.

참고문헌

[1] KIM, Seung-seob, et al. 인공지능 기반 임상의학 결정 지원 시스템 의료기기의 성능 및 안전성 검증을 위한 간 중앙 표준 데이터셋 구축, 대한영상의학회지, 2021;82:e106.
 [2] 김동엽, et al. 인공지능 의료기기 소프트웨어의 IEC 62304 국제표준 준수 가이드라인 개발 및 적용. 정보과학회논문지, 2021, 48.1: 71-81.
 [3] 노시형, et al. 웹기반 머신러닝 기술을 이용한 간 경화증 진단 시스템 구축. 한국정보처리학회 학술대회논문집, 2021, 28.1: 19-21.