

대사증후군기반 의료 빅데이터 표준화 시스템의 설계

김지언^{O*}, 이기택^{*}, 정창원^{*}, 김규겸^{*}, 김태훈^{*}, 유중현^{*}, 전홍영^{*}, 장미연^{*},
이윤오^{**}, 조은영^{**}, 유태양^{***}, 김대원^{****}, 윤권하^{*****}

^{O*}원광대학교 영상의학기반 폐 및 골질환 연구센터, ^{**}원광대학교병원 전산정보팀, ^{***}원광대학교병원 내과,
^{****}원광대학교병원 내분비-대사 내과, ^{*****}원광대학교의과대학 신경외과, ^{*****}원광대학교의과대학 영상의학과
e-mail: {kakasky112^O, tlzld12345, mediblue, kgkim, tae_hoonkim, jhryu, zip80, wkdaldus5}@wku.ac.kr^{*},
wonduke@empas.com^{**}, love6970@wku.ac.kr^{***}, yutaeyang@gmail.com^{****}, {kimdw@wku.ac.kr^{*****},
khy1646@wku.ac.kr^{*****}

Design of Medical Bigdata Standard System Based on Metabolic Syndrome

Ji-Eon Kim^O, Gi-Taek Lee^{*}, Chang-Won Jeong^{*}, Kyu Gyeom Kim^{*}, Tae-Hoon Kim^{*}, Jong-Hyun Ryu^{*}, Hong
Young Jun^{*}, Mi Yeon Jang^{*}, Yun Oh Lee^{**}, Eun Young Cho^{***}, Tae Yang Yu^{****}, Dae Won Kim^{*****},
Kwon-Ha Yoon^{*****}

^{O*}Imaging Science-based Lung and Bone Disease Research Center, Wonkwang University

^{**}Computing&Information Team, Wonkwang University hospital

^{***}Dept. of Internal Medicine, Wonkwang University School of Medicine and Hospital

^{****}Division of Endocrinology and Metabolism, Dept. of Medicine, Wonkwang University School of Medicine

^{*****}Dept. of Neurosurgery, Wonkwang University School of Medicine and Hospital

^{*****}Dept. of Radiology, Wonkwang University School of Medicine and Hospital

● 요약 ●

본 논문에서는 대사증후군관련 의료 빅데이터 표준화 시스템의 환경과 소프트웨어 환경을 설계한 사항에 대해서 기술한다. 이를 위해 임상데이터를 기반으로 의료 빅데이터를 수집하고 국제 표준화된 공통 데이터 모델로 수집된 데이터를 ETL하여 통합 데이터베이스에 저장하였다. 본 연구를 통해 구축된 의료 빅데이터 표준화 시스템은 향후 의사결정 보조시스템 개발과 연계하여 효과적인 검색과 다양한 통계 분석을 지원할 계획이다. 또한 병원의 다양한 임상 연구를 지원하기 위한 주요 시스템으로 자리매김할 것으로 기대한다.

키워드: 의료 빅데이터(medical bigdata), 표준화(standard), 통합 데이터베이스(integration database)

I. Introduction

국내의 인구의 식생활 습관과 인구의 고령화로 인하여 대사증후군 환자가 급증하고 있다[1,2]. 특히 한국에서 성인 인구의 18~20%가 대사증후군 환자이며, 30세 이상 성인의 유병률은 약 30%에 이르는 것으로 보고되고 있다. 이와 관련하여 본 논문에서는 대사증후군 환자의 임상 데이터기반의 빅데이터화 및 분석을 통해 난치성 간질환을 규명하고자 한다. 이를 위해 먼저 의료정보의 표준화가 선행되어야 한다. 그러나 현재 의료정보 표준화 현황으로 병원간뿐만 아니라 병원내 표준 미적용으로 협천하기 어려운 실정이다. 최근 빅데이터 관련 기술적 수요와 관심이 높아짐에 따라 표준화 기구들은 각각의 도메인 영역에서 빅데이터를 정의하고, 기존 표준화 노력의 연장선상

에서 도메인 내 상호운영성을 이루기 위한 노력을 하고 있으나 초기 단계로 해결해야할 문제들이 많다.

본 논문에서는 이러한 병원내 임상 데이터의 표준화를 위한 일련의 과정과 예상되는 시스템의 결과를 보이고자 한다.

II. Related Works

대사증후군에 관한 질환에 따르는 위험요소와 치료의 복잡한 연관 관계에 따르는 관련 데이터의 통합과 표준화가 요구된다. 다음 그림 1은 Ben van Ommen이 정의한 시스템 바이오로직의 다양한 연관성을 보이고 있다[3].

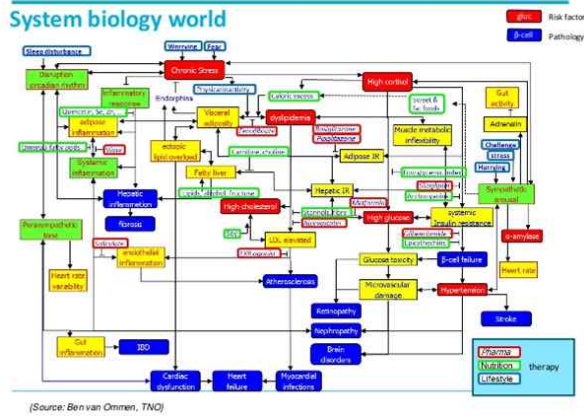


Fig. 1. System biology

의료 빅데이터의 수집과 저장 그리고 활용에 대한 개념도는 다음 그림 2와 같다.

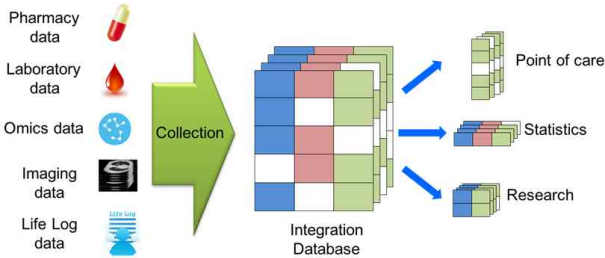


Fig. 2. Medical Bigdata Collection and utilization

그림 2에서 보인바와 같이 다양한 임상데이터를 수집하여 통합된 데이터베이스에 저장하고 연구자, 의사 그리고 관련 기관담당자들에 따라 연구 또는 진단 그리고 통계 분석에 따르는 의사결정 등에 활용할 수 있다. 그러나 체계적인 표준화 부재에 의한 시스템 구축으로 인하여 데이터 통합뿐만 아니라 표준화된 용어, 비정형 데이터의 처리 문제 등을 갖고 있다. 최근 들어서 국내외에서 비정형 데이터의 정형화 연구와 표준화된 용어 그리고 ETL(Extraction, Transformation, Load)에 관련된 연구가 활발하게 진행되고 있다[4, 5]. 우리는 병원의 임상데이터를 중심으로 표준화하며, 기타 유전체 정보 및 국내 코호트 DB를 포함하여 통합 수집하고 이를 기반으로 표준화하여 검색 시스템을 개발하고자한다.

III. The Proposed System

제안하는 의료 빅데이터 표준화 시스템의 구성은 그림 3과 같다. 병원의 인프라를 활용하여 정형 데이터와 비정형 데이터의 구조화와 OHDSI의 CDM 모델로 변환하여 데이터베이스에 저장하고 이에 대한 검색 엔진을 통해 검색하기 위한 시스템으로 구성하였다.

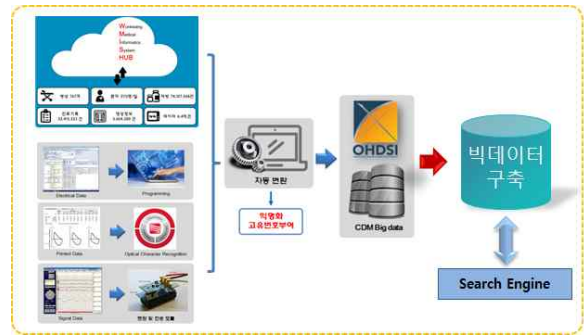


Fig. 3. Medical Bigdata Standard System structure

의료 표준화 시스템을 개발하기 위한 과정은 그림 4와 같다. 먼저 의료 빅데이터의 확보를 위해 다양한 임상데이터를 비롯하여 생활습관 데이터와 영상데이터 그리고 코호트 DB를 포함하여 수집한다. 비정형 데이터의 경우 정형화를 진행하고, 수집된 데이터의 탐색을 통해 새로운 데이터를 추가하는 일련의 과정을 거친 후 데이터 정의가 완료된 후 통합 데이터베이스를 구축한다.



Fig. 4. Development procedure for medical bigdata standard system

그리고 효율적인 빅데이터 검색을 위한 기술을 기반으로 검색엔진을 개발하여 의료 빅데이터 표준화 시스템을 개발하고자 한다. 이후, 대시중후군에 관련된 다양한 데이터 분석을 통해 난치성 질환의 진행정도에 따르는 진단 및 치료 그리고 예측 연구를 진행하여 의사결정정보 시스템 개발하고자 한다.

제안한 시스템의 소프트웨어 구성요소는 그림 5와 같이 데이터 수집, 저장과 처리 그리고 데이터베이스 부분으로 정의하여 설계하였다.

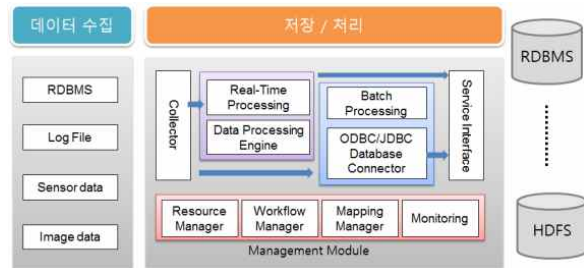


Fig. 5. Software architecture for medical bigdata standard system

IV. Conclusions

최근 의료빅데이터에 대한 관심이 높아지고 있으며, 병원의 다년간의 임상데이터의 활용을 높이기 위한 방안으로 빅데이터의 표준화와 연구목적의 Clinical Data Warehouse를 구축하고 있는 추세이다.

본 논문에서는 제안하는 의료 빅데이터의 표준화 시스템에 대한 전체 시스템 구성과 소프트웨어 환경을 기술하였다. 그리고 병원내 임상 데이터의 표준과 함께 비정형 데이터의 구조화와 표준화를 위해 OHDSI의 CDM 모델을 기반으로 하였다. 현재 정형화된 병원의 임상데이터를 OHDSI의 CDM 모델링을 진행하였고, 물리적인 시스템 환경을 구축하였다.

향후 연구로는 생활습관 데이터와 기타 다른 의료정보의 통합을 위한 추가 연구를 진행해야 한다. 또한 이와 함께 대사증후군 환자의 임상데이터와 의료영상데이터 그리고 생활습관 데이터간의 연관관계를 분석하기 위한 연구를 진행하고자 한다.

Acknowledgment

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF- 2016M3A9A7918501)

References

- [1] Yun-Jung Kang. "A Prospective Cohort Study on Predictive Risk Factors Causing Metabolic Syndrome within the First Two Years" Korean Journal of Obesity 24.3, 2015.
- [2] Sung Hoon Ki, Sung Ae Ye, Yun Kyung Song, "Original Article : The Study on Correlation between Abdominal Fat Area and Obesity Index, Metabolic Syndrome Components in Obese Adult Women", Journal of oriental rehabilitation medicine, 26.1, pp. 103-111, 2016.
- [3] Data for personalized health(care) : time for quality, not quantity, https://www.ihe-nl.org/uploads/bestanden/Alain_van_Gool.pdf
- [4] Hripcsak, George, et al. "Characterizing treatment pathways at scale using the OHDSI network." Proceedings of the National Academy of Sciences (2016): 201510502.
- [5] OHDSI(Observational Health Data Sciences and Informatics, <http://www.ohdsi.org/>