

## Hadoop기반의 공개의료정보 빅 데이터 분석을 통한 한국여성암 검진 요인분석 서비스

박민희<sup>1</sup> · 조영복<sup>2</sup> · 김소영<sup>3</sup> · 박종배<sup>4</sup> · 박종혁<sup>1\*</sup>

### Analysis of Factors for Korean Women's Cancer Screening through Hadoop-Based Public Medical Information Big Data Analysis

Min-hee Park<sup>1</sup> · Young-bok Cho<sup>2</sup> · So Young Kim<sup>3</sup> · Jong-bae Park<sup>4</sup> · Jong-hyock Park<sup>1\*</sup>

<sup>1\*</sup>Department of Convergence in Health and Biomedicine Program in Health Policy Graduate, School, Chungbuk National University, Cheongju, 28644, Korea

<sup>2</sup>Department of Computer & Information Security, Daejeon University, Daejeon 34520, Korea

<sup>3</sup>Department of Public Health and Preventive Medicine, Chungbuk National University Hospital, Cheongju, 28644, Korea

<sup>4</sup>Department of Radiology, Chungbuk Health & Science University, Cheongju, 28150, Korea

#### 요 약

본 논문에서는 공개의료정보 빅데이터 분석을 위해 클라우드 환경에서 아파치 하둡 기반의 클라우드 환경을 도입하여 컴퓨팅 자원의 유연한 확장성을 제공하고 실제로, 로그데이터가 장기간 축적되거나 급격하게 증가하는 상황에서 스토리지, 메모리 등의 자원을 신속성 있고 유연하게 확장을 할 수 있는 기능을 포함했다. 또한, 축적된 비정형 로그데이터의 실시간 분석이 요구되어질 때 기존의 분석도구의 처리한계를 극복하기 위해 본 시스템은 하둡(Hadoop) 기반의 분석모듈을 도입함으로써 대용량의 로그데이터를 빠르고 신뢰성 있게 병렬 분산 처리할 수 있는 기능을 제공한다. 빅데이터 분석을 위해 빈도분석과 카이제곱검정을 수행하고 유의 수준 0.05를 기준으로 단변량 로지스틱 회귀 분석과 모델별 의미 있는 변수들의 다변량 로지스틱 회귀분석을 시행 하였다. ( $p < 0.05$ ) 의미 있는 변수들을 모델별로 나누어 다변량 로지스틱 회귀 분석한 결과 Model 3으로 갈수록 적합도가 높아졌다.

#### ABSTRACT

In this paper, we provide flexible scalability of computing resources in cloud environment and Apache Hadoop based cloud environment for analysis of public medical information big data. In fact, it includes the ability to quickly and flexibly extend storage, memory, and other resources in a situation where log data accumulates or grows over time. In addition, when real-time analysis of accumulated unstructured log data is required, the system adopts Hadoop-based analysis module to overcome the processing limit of existing analysis tools. Therefore, it provides a function to perform parallel distributed processing of a large amount of log data quickly and reliably. Perform frequency analysis and chi-square test for big data analysis. In addition, multivariate logistic regression analysis of significance level 0.05 and multivariate logistic regression analysis of meaningful variables ( $p < 0.05$ ) were performed. Multivariate logistic regression analysis was performed for each model 3.

**키워드** : 건강검진, 여성암, 유방암, 자궁경부암, 지역건강조사 빅 데이터.

**Key word** : Health medical examination, Female cancer, Breast cancer, Cervical cancer, Local Health Survey Big Data.

Received 3 May 2018, Revised 24 July 2018, Accepted 20 August 2018

\* Corresponding Author Jong-Hyock Park(E-mail: jonghyock@gmail.com, Tel:+82-43-261-2873)

College of Medicine / Graduate School of Health Science Business Convergence, Chungbuk National University, Chungbuk 28644, Korea

Open Access <http://doi.org/10.6109/jkiice.2018.22.10.1277>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

최근 컴퓨팅 패러다임이 클라우드 환경으로 전환되면서 빅 데이터의 처리에 대한 관심이 고조되고 있다. 보건 의료 분야에서도 공공데이터의 개방이 빠르게 이루어지면서, 다양한 서비스 개선을 위한 공공 의료 빅 데이터가 민간 및 공공분야에서 많이 사용되고 있다.

보건 의료 분야의 빅 데이터는 데이터 자체의 특성과 의료 분야라는 특수성으로 인해 다른 분야의 일반적인 빅 데이터와는 차별점이 많다. 따라서 보건 의료 분야에서는 데이터 분석에 앞서 다양한 종류와 형태의 의료 데이터를 의미적으로 융합할 수 있는 방법이 전제되어야 한다[1].

우리나라 사망원인 1위인 암은 2016년 전체 사망자의 27.8%(78,194명)를 차지하고, 모든 암의 연령표준화 발생률도 1999년 219.9명에서 2015년 275.8명으로 증가율을 보였다. 이 중 남성은 1999년 291.9명에서 2015년 301.2명으로 증가율을 보인 반면, 여성은 1999년 173.3명에서 2015년 266.1명의 증가율을 보여 여성의 암이 빠르게 증가하고 있음을 알 수 있다[2]. 세계보건기구(WHO)는 암의 1/3은 예방 가능하고, 1/3은 조기 진단만 되면 완치가 가능하며, 1/3의 환자도 적절한 치료를 하면 완화가 가능한 것으로 제시하고 있다. 자궁경부암의 경우 전 암 단계의 병변을 발견하여 치료하면 암 발생 자체를 줄일 수 있고, 유방암도 조기진단만 되면 유방모양을 그대로 유지하며 암을 완치시킬 수 있어 두 암종 모두 완치율과 생존율을 높이는 조기검진의 필요성이 크게 대두되고 있다.

국내에서는 자궁경부암은 만 30세 이상 연령에 자궁경부 세포검사를, 유방암은 만 40세 이상을 대상으로 유방촬영술을 각각 2년에 한 번씩 받도록 권장하고 있고, 2016년 국내 여성암 수검률은 자궁경부암 38.9%(미국 77.9%, 영국 78.9%, 일본 18.9%), 유방암 48.2%(미국 66.5%, 영국 73.3%, 일본 17.6%)로 미국, 영국에 비해서는 더 낮고, 일본에 비해서는 높은 수검률을 보였다. 미국, 캐나다 등 서구 선진국들과 일본에서는 자궁경부암 검진을 20대부터 시행하도록 권고하거나, 국가검진으로 제공하고 있고 국내에서도 2016년부터 만 20대 이상 연령에 자궁경부암검진을 2년에 한 번씩 받도록 권장하고 있다[3]. 본 연구에서는 미국, 캐나다, 선진국들과 일본처럼 국내에서도 암 검진이 20대부터 시행하도록 권

고되거나, 조기에 이루어져야 하는 타당성을 분석하고자 하였다. 국내 암 검진의 연령을 20대로 낮추기 위해 조기 검진의 타당성을 설명하기 위해 기준으로 2010년도 데이터를 사용하였다. 또한 여성 암 수검률을 높이기 위한 연구는 여전히 유효하고 가능요인들에 대한 연구가 요구되었다. 그러나 지금까지 암 수검 관련요인에 대한 선행연구는 대부분 일부 지역 또는 일부 병원 자료만이 사용되었고, 암 검진 수검 행동도 다양한 요인들과 관련성이 있으나 주로 일부요인의 상관관계나 단편적인 관련성 연구로 결과를 일반화하기에는 제한점이 있어, 이는 대표성 있는 자료를 통해 확인 될 필요가 있다 [4]. 따라서 본 연구에서는 건강신념모델[5], PRECEDE 모델[6], 사회생태학적 모델[7], 인지모델[8] 등을 참고한 ‘암 수검 관련요인에 관한 연구모형’을 통해 수검 관련요인들을 분류하고, 대표적인 한국 여성암인 유방암과 자궁암 검진에 미치는 영향을 빅 데이터 분석모델을 기반으로 여러 가지 가능 요인을 분석하고 규명하고자 한다.

## II. 관련연구

### 2.1. 보건 의료 빅 데이터

보건 의료 분야에서 빅 데이터의 활용은 개인이 건강을 관리하고 의료를 선택하는 방식을 변화시켜 보건의료시스템 전반에서 건강 결과와 지출의 효율을 높이는 혁신의 동력으로 기대되고 있다. 빅 데이터 활용 사업은 첨단 기술적 요소뿐 아니라 데이터가 담고 있는 정보의 품질이 높아야 가치를 확대시킬 수 있다. 빅 데이터의 품질은 여러 가지 공적 프로그램에서 행정적으로 수집된 데이터를 연결하고 개인의 동의와 참여를 기반으로 유전체, 생의학, 행태 정보 등이 연결될 때 정보의 완성성 측면에서 높아진다[1,9,10,11].

다양한 데이터 소스에서 수집된 빅 데이터를 처리·분석하여 지식을 추출하고 이를 기반으로 지능화된 서비스를 제공하기 위해서는 빅 데이터 플랫폼이 필요하다. 그림 1은 UNC 유방암센터의 여성 암 관리 시스템을 도식화 한 것이다. 그림1에서도 암 검진 수검에 영향을 미치는 요인 대부분의 연구가 인구사회학적 요인들을 포함하였으나, 많은 경우 이러한 요인들이 암 수검에 유의하지 않다는 결과를 보고하였다. 기존 관련연구에서

도 유의하다고 보고된 인구사회학적 요인들도 유의성의 방향이 연구마다 상이하다고 분석되었다[12-17]. 인구사회학적 요인과 건강행태학적 요인에 가능 요인을 추가한 연구들은 다음과 같다. 민간 암보험 가입자는 비가입자에 비해 암 검진도 많이 받았고[13-14], 암 검진 장애요인으로 는 본인이 ‘건강 하다고 생각하여서’가 가장 많았다[13-20]. 또한 기존 관련 연구를 통해 건강 검진의 장애요인과 암 검진의 장애요인을 분석한 결과 ‘필요성을 몰라서’가 가장 큰 빈도를 보였고, 가능요인을 추가하여 살펴본 결과, 유방암 대상자의 경우 건강검진을 받을수록, 다른 암 검진을 받을수록, 민간 암 보험에 가입할수록 수검률이 높았으며, 자궁경부암 대상자의 경우 건강검진을 받을수록, 민간 암 보험에 가입할수록 수검률이 높아, 건강에 대한 관심이 있는 사람일수록 암 검진을 많이 받은 것으로 나타났다[17-21].

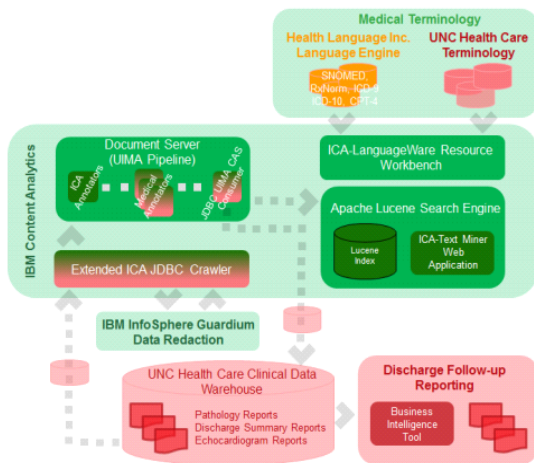


Fig. 1 UNC Healthcare's Women's Cancer Diagnostic System

2.2. 빅 데이터 분석 플랫폼

하둡은 빅 데이터를 저장하는 분산파일 시스템(Hadoop Distributed File System, HDFS)과 분산 병렬 처리하는 맵리듀스(MapReduce)로 구성되어 있다. 저가 장비 및 스토리지 활용으로 저비용으로 방대한 양의 데이터 저장 및 처리가 가능하다. HDFS를 통해 다양한 형태의 초대용량 데이터를 분산 저장하며 맵리듀스로 빅 데이터를 초고속으로 처리할 수 있다. 하둡은 자바 언어로 작성되어 있으며 크게 아래의 3가지 모듈로 구성되어 있다[9,10,11].

- 하둡 공통 (Hadoop Common): 다른 모듈들을 지원하는 공통 유틸리티 모음. 하둡을 기동하는 스크립트나 분산 파일 시스템에의 접근을 돕는 유틸리티 등이 포함된다.
- 하둡 분산 파일 시스템 (Hadoop distributed file system; 이하 HDFS): 하둡 공통을 이용하여 대용량 데이터를 다수의 컴퓨터에 분산시켜 고속으로 처리하기 위한 분산 파일 시스템.
- 하둡 맵리듀스 (Hadoop MapReduce): 분산 파일 시스템에 저장된 대용량 데이터의 병렬 처리를 위한 소프트웨어 프레임워크.

III. Hadoop기반의 공개의료정보 빅 데이터 분석을 통한 한국 여성암 검진요인분석 서비스

제안 논문은 지역사회건강조사 자료를 분석하여 국내 여성암 검진 요인을 분석한다.

3.1. 연구자료

2010년 8월 1일부터 2010년 10월31일 까지 지역사회건강조사에 참여한 대상자 229,229명 중 성인여성 124,654(50.5%)명을 기반으로 복합표본추출을 활용한 자료를 사용하였다. 조사된 성인 여성 중 “2010 한국 암 검진권고안”에 따라 자궁경부암은 만 30세 이상, 유방암은 만 40세 이상 여성을 대상으로 결측 자료가 있는 경우 분석에서 제외하였다. 분석모델은 그림 2와 같다.

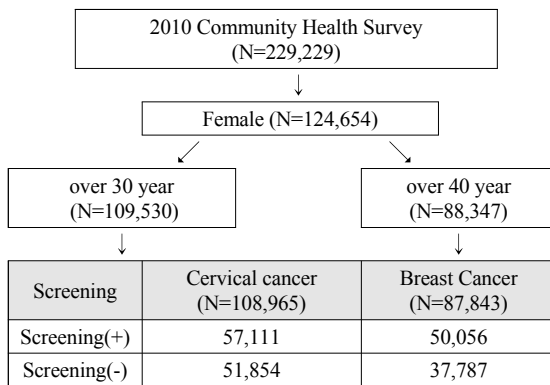


Fig. 2 Flow chart of study design

### 3.2. 빅 데이터 분석을 위한 요인 구성

제안 논문의 시스템 구조는 그림 3과 같다. 의료정보 빅 데이터 분석을 위해 제안 논문은 아파치 하둡 기반에서 데이터를 군집화 하고 분류하였으며 분산처리의 효율성을 위해 맵리듀스를 활용하였다.

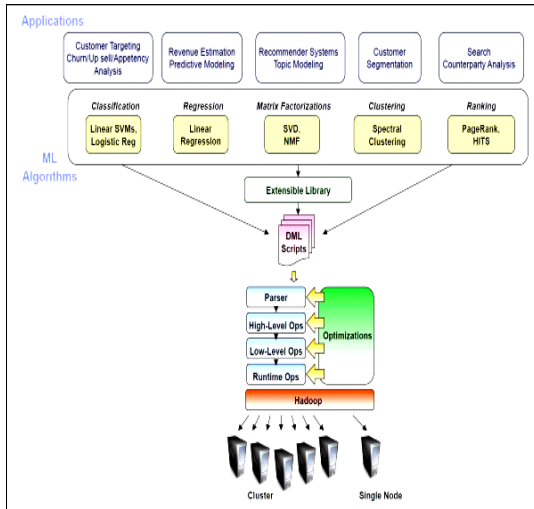


Fig. 3 System Architecture

「2010년 지역사회건강조사」 자료를 바탕으로 [12]를 참조로 여성암 수검에 따른 연구대상자의 수검 행동과 관련된 요인 분석을 위해 인구사회학적 요인, 건강행태학적 요인, 가능요인으로 구분하였다. 본 논문에서는 빅 데이터 분석을 위해 하둡 기반의 맵리듀스를 활용해 그림 4와 같이 분석하였다.

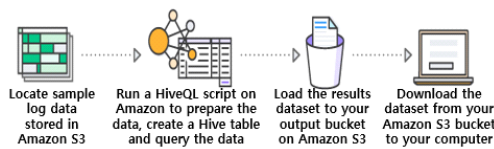


Fig. 4 Map-reduce for big data analysis

인구사회학적 요인으로 연령, 교육수준, 거주 지역, 결혼, 직업, 소득수준 상태를 선정하였고, 건강행태학적 요인으로는 규칙적인 식사, 규칙적인 수면, 흡연, 음주, 신체적 활동, 만성질환유무, 체질량지수, 스트레스지수, 건강상태(EQ-5D)등을 구분하여 분석하였고, 가능요인

으로 인플루엔자 예방접종, 다른 암 검진(위암, 간암, 대장암), 민간의료보험가입 여부, 건강 검진의 장애요인, 암 검진의 장애요인 등으로 구분하였다. 분석모델을 기준으로 여성암 검진 대상자의 일반적인 특성은 표 1과 같이 정의하고 분석에 활용하였다.

Table. 1 General characteristic variables

<b>Variable</b>	Age, Level of education, Residence, Marriage, Job, Income, Smoking, Drinking, A regular meal, Regular Sleep, Physical activity, BMI, Chronic disease, Stress Index, EQ-5D index, Vaccination, Health screenings, Other cancer screenings, Private insurance.
-----------------	--

### 3.3. 분석방법

본 논문에서는 공공의료정보 데이터 분석을 위해 하둡 플랫폼을 구성하고 통계적 분석을 실시하였다. 대상자의 일반적인 특성 및 건강관련 특성을 비교하기 위해 빈도분석을 실시하고, 여성암 수검자와 미수검자 간의 차이를 분석하기 위해 카이제곱검정을 시행하였다. 또한 여성암 수검에 대한 행위 관련 요인에 미치는 영향을 알아보기 위해 유의 수준 0.05를 기준으로 단변량 로지스틱 회귀분석과 의미가 있는 변수를 Model 1-Model 3으로 나누어 다변량 로지스틱 회귀분석을 빅 데이터 플랫폼에서 수행하였다.

## IV. 실험결과

### 4.1. 여성암 검진 요인 일반적 특성

한국 암 검진 권고안에 따라 2년 이내에 유방암 검진을 받을 수 있는 연령을 40세 이상, 자궁암 검진연령은 30세 이상으로 제한하고 있다. 암검진 연령을 선진국과 같이 20대로 낮추기 위한 타당성 조사를 위해 유방암 검진 대상자인 40세 이상 여성(87,843명) 중 수검자 50,056(57.0%)명, 미수검자 37,787(43%)명이고, 자궁암 검진 대상 30세 이상 여성(109,530명) 중 수검자 57,111(53.9%)명, 미수검자 51,845(45.0%)명을 대상으로 하였다. 여성암 검진 요인 일반적 특성은 APPENDIX 표 2와 같은 결과를 보였다. 건강검진과 암 검진 장애요인의 분석 결과 각각 ‘필요성을 몰라서’가 가장 높았으며, 두 변수의 상관관계에서는 암 검진 장애요인과 건강

검진 장애요인, 암검진과 검진이 각각 0.64, 0.62로 높은 상관관계를 보였고, 만성질환과 나이, 교육과 나이가 각각 0.58, 0.53 으로 높은 상관관계를 갖는 것으로 나타났다.

4.2. 여성암 검진 요인 단변량 분석 결과

공공의료 빅 데이터를 기반으로 여성암 검진의 요인을 다변량 분석결과 유의한 차이를 보이는 변수와 기준에 관련성이 있다고 알려진 변수를 기준으로 그림 5와 같이 로지스틱 회귀분석 결과 인구사회학적 요인과 건강행태학적 요인으로 의료 빅 데이터를 분석하였다.

4.3. 여성암 검진 요인 다변량 분석결과

제안 논문에서는 의미가 있는 변수들을 비교하기 위해 여성암 검진 수검을 위한 요인을 모델별로 나누어 비교 분석한 결과 APPENDIX 표 3과 같은 결과를 보였다. Model 1은 인구사회학적 요인, Model 2는 인구사회학적 요인과 건강행태학적 요인, Model 3은 기존 연구에서 제시되지 않았던 가능요인을 기반으로 모든 요인을 포함시켜 분석한 결과로 Model 1은 50-60세 미만의 여성이 기혼일수록, 직업은 무직인 경우보다 직업이 있는 경우 수검할 가능성이 높았고, 육체적 노동을 할수록, 전문대 이상의 학력과 수입이 높을수록 수검률이 높았다. Model 2에서는 50-60세 미만의 여성이 기혼일수록,

직업은 무직인 경우보다 직업이 있는 경우 수검할 가능성이 높았고, 육체적 노동을 할수록, 전문대 이상의 학력과 수입이 높을수록, 음주를 조금 하며 규칙적인 식사와 신체활동과 BMI가 정상인 경우, 만성질환이 있는 경우, 삶의 질이 높을수록 수검률이 높았다.

Model 3에서는 50-60세 미만의 여성이 기혼이고 육체적인 노동을 할수록, 규칙적인 식사와 신체활동이 정상인 경우 건강검진, 암 검진을 받을 수록, 민간 암 보험을 가입할수록 수검확률이 높은 것으로 실험 결과 나타났다. 자궁경부암 수검 관련된 요인 비교 시 Model 1은 50-60세 미만의 여성이 기혼일수록, 육체적 노동을 할수록, 전문대 이상의 학력과 수입이 높을수록, 수검률이 높았고, Model 2에서는 50-60세 미만의 여성이 기혼일수록, 육체적 노동을 할수록, 음주를 조금하며 규칙적인 식사와 신체활동이 정상인 경우, 만성질환이 있는 경우, 삶의 질이 높을수록 수검률이 높았다. Model 3에서는 30-39세 미만의 여성이 도시에 살수록, 기혼일수록 비육체적인 노동을 할수록, 규칙적인 식사와 신체활동이 정상인 경우, 건강검진을 받을수록(OR 1.66, 95% CI 1.45 - 1.89), 민간 암 보험에 가입할수록 (OR 1.53, 95% CI 1.41 - 1.67) 통계적으로 유의하게 수검가능성이 높은 것으로 나타났다.

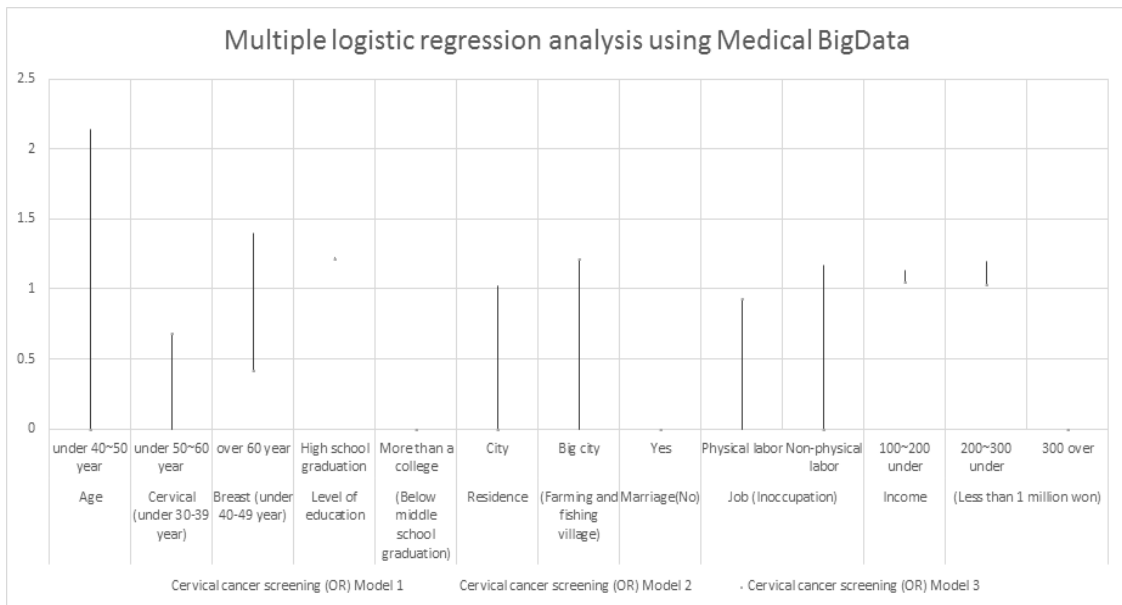


Fig. 5 Multiple logistic regression analysis using Medical BigData

## V. 결 론

본 연구는 의료 빅 데이터를 기준으로 여성암 수검 요인 분석을 위해 「2010년 지역사회건강조사」 자료를 바탕으로 우리나라 여성의 대표적인 유방암과 자궁경부암 검진 수검에 영향을 미치는 요인을 하둠 환경에서 분석한 단면조사 연구이다. 여성암 수검에 관련된 요인을 조사하기 위해 각각의 암 검진 대상연령에 해당하는 여성들을 추출하여 의미 있는 변수들을 Model 1~Model 3으로 나누어 HDFS를 통해 다양한 형태의 초대용량 데이터를 분산 저장하고 병렬처리를 지원하는 맵리듀스로 초고속으로 처리한 결과 분석속도는 4.5배 정도 빠르게 분석이 가능해졌다. 분산처리 시스템의 고성능을 제공하기 위해 리소스에 따른 성능 기반으로 실시한 결과 분석 수행속도는 약 4.5배 빠른 결과로 높은 성능을 보였다. 따라서 의료정보와 같은 빅데이터 자원을 분석하기 위해서 고성능을 분석이 기반이 되어야 하고 하둠과 스파크등의 자원에 따른 우선순위를 부여함으로 분석 노드의 수를 적절히 분산하는 경우 기존 보다 높은 성능의 분석 성능을 제공함을 확인하였다. 또한 분석 결과의 정확도로 분석내용의 결과 한국여성암 검진 수검관련 요인이 이 전 연구결과와 동일하게 나타났으나 본 연구에서는 만성질환을 앓고 있는 경우 수검을 받을 확률이 높은 것으로 조사되어 기존의 연구결과와 다르게 나타났다. 특히, 건강에 대한 관심도가 높아 건강검진이나 다른 암 검진을 받은 사람일수록, 다른 요인들에 비해 여성암 수검률이 월등히 높음을 감안할 때, 건강에 대한 관심도가 낮은 저 연령층과 소외된 계층을 겨냥한 정기적인 교육과 대중매체의 홍보, 검진 후 돌봄 관리 시스템은 여전히 유효하다고 할 수 있다. 또한 여성암은 조기에 만 발견되면 완치율과 생존율이 높다는 필요성을 증대시킨 강력한 메시지가 필요하다. 이는 Model 3에서 나온 결과를 토대로 젊은 여성일 수록 수검가능성이 높다 것을 감안한다면 암 검진은 20대부터 시행하도록 권고되거나, 조기에 이루어져야하는 타당성을 뒷받침하고 있다. 자칫 부담스러울 수 있는 여성암검진의 특수한 검사환경도 편안한 환경조성을 통해 더 이상 불편한 검사가 아니라는 인식도 전환시킬 필요도 있다. 따라서 이러한 연구 결과가 새로운 프로그램 개발을 위한 자료로 활용되고, 여성암 조기검진의 교육 및 관리가 건강에 대한 관심과 이해의 폭이 넓은 방향으로 편성되길 기대해

보며, 공공의료정보의 활성화와 빅 데이터 분석을 위한 기본 모형을 제공할 수 있는 기회가 될 수 있을 것으로 사료된다.

향후 연구로는 리소스자원의 급속한 발전을 기반으로 클라우드 환경에서 의료빅데이터를 분석가능한 환경을 조사하고, 다양한 기존자료를 활용한 국내 여성암 검진 수검 관련 요인의 정확한 요인 분석 도출을 진행하고자 한다.

## ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) in 2018 (No.2016R1A2B4011045)

## REFERENCES

- [ 1 ] J. K. Son, S. A. Sin, T. W. Han, "Life care Trend using Big Data," *Journal of Korean Information Society (Information and Communication)*, vol. 32, no. 11, pp. 3-7, Oct. 2015.
- [ 2 ] National Cancer Center. [Internet]. Available: <https://www.cancer.go.kr/lay1/S1T639C640/contents.do>.
- [ 3 ] Ministry of Health and Welfare. [Internet]. Available: <http://www.mohw.go.kr/react/index.jsp>.
- [ 4 ] Y. Y. Chang, B. L. Cho, K. Y. Son, D. W. Shin, H. K. Shin, H. K. Yang, A. S. Shin and K. Y. Yoo "Determinants of gastric cancer screening attendance in Korea: a multi-level analysis," *International Journal of BioMed Cancer*, vol.15, no.1, pp. 336-343, May. 2015.
- [ 5 ] I. S. Cho and Y. S. Park, "A study on regular cancer screening behavior among middle-aged women," *Journal of Korean Academy of Nursing*, vol. 34, no. 1, pp.141-149, Feb. 2004.
- [ 6 ] G. Jacklyn, K. Howard, L. Irwig, N. Houssami, J. Hersch and A. Barratt, "Impact of extending screening mammography to older women: Information to support informed choices," *International journal of cancer*, vol. 141, no. 8, pp. 1540-1550, Jul. 2017.
- [ 7 ] M. N. Suh, S. H. Song, H. N. Cho, B. Y. Park, J. K. Jun, E. J. Choi, Y. E. Kim and K. S. Choi, "Trends in Participation Rates for the National cancer screening program in Korea 2002-2012," *Journal of Korean Cancer Association*, vol. 49, no. 3, pp. 798-806, Jul. 2017.

- [ 8 ] M. G. Marmot, D. G. Altman, D. A. Cameron, J. A. Dewar, S. G. Thompson and M. Wilcox, "The benefits and harms of breast cancer screening: an independent review," *British journal of cancer*, vol. 108, no. 11, pp. 2205-2240, Jun. 2013.
- [ 9 ] Y. B. Cho S. H. Woo and S. H. Lee, "The Big Data Analysis and Medical Quality Management for Wellness", *Journal of the Korea Society of Computer & Information*, vol. 19, no. 12, pp. 101-109, December 2014.
- [10] M. K. Kim and Y. B Cho, "An Analysis of Factors Affecting Quality of Life through the Analysis of Public Health Big Data", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 6 pp. 835-841, Jun. 2018.
- [11] M. K. Kim and Y. B Cho, "A Secure Telemedicine System in Smart Health Environment using BYOD", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 19, no. 10 ,pp. 2478-2480, Oct. 2015.
- [12] E. Altobelli and A. Lattanzi, "Breast cancer in European Union: an update of screening programmes as of March 2014(review)," *International Journal of Oncology*, vol. 45, no. 5, pp. 1785-1792, Nov. 2014.
- [13] Y. I. Jung, H. S. Kim and D. S. Choi, "Factors Associated with Cancer Screening Among Korean Adults: A Literature Review," *Korean Journal of Clinical Health Promotion*, vol. 10, pp. 185-195, Oct. 2010.
- [14] J. H. Kang, "The Analysis of the Association Factors Which Influence on Cancer Screening in Korean women," Ph. D. dissertation, Chungbuk National University Graduate School, Department of Medicine, 2012.
- [15] N. Y. Choi and B. S. Lee, "Factors of Breast and Cervical Cancer Screening Behaviors in Married Female Immigrants," *Journal of the Korea Contents Association*, vol. 15, no. 6, pp. 326-336, Jun. 2015.
- [16] N. H. Yu, S. M. Kwon, H. Y. Lee, E. C. Park, K. S. Cho and M. S. Kwak, "Factors Affecting Satisfaction of National Cancer Screening Program," *Korean Journal of Health Policy & Administration*, vol. 19, no. 1, pp. 31-48, Mar. 2009.
- [17] Y. H. Yang, "Relationship between Knowledge about Early Detection, Cancer risk Perception and cancer screening Tests in the General Public Aged 40 and over," *Journal of Nursing & Asian Oncology Nursing*, vol. 12, no. 1, pp. 52-60, Feb. 2012.
- [18] S. O. Lee, E. S. Sim, and S. H. Ahn "Factors Affecting Periodic Screening Behaviors for Breast Cancer among Hospital Nurses," *Korean J Woman Health Nursing*, vol. 16, no. 1, pp. 390-398, Dec. 2010.
- [19] Y. Laitman, D. M. Feldman, M. Sklair-Levy, A. Yosepovich, I. BarshackI, M. Brodsky, O. Halshtok, A. Shalmon, M. M. Gotlieb and E. Friedman "Abnormal Findings Detected by Multi-modality Breast Imaging and Biopsy Results in a High-risk Clinic," *Journal of clinical Breast Cancer*, Dec. 2017.
- [20] J. Jung, S. M. Moon, H. C. Jang, C. I. Kang, J. B. Jun, Y. K. Cho, S. J. Kang, B. J. Seo, Y. J. Kim, S. B. Park, J. Lee, C. S. Yu and S. H. Kim, "Incidence and risk factors of postoperative pneumonia following cancer surgery in adult patients with selected solid cancer: results of "Cancer POP" study", *International Journal of Cancer Medicine*, vol. 7, no. 1, pp. 261-269, Jan. 2018.
- [21] K. J. Min, Y. J. Lee, M. N. Suh, C. W. Yoo, M. C. Lim, J. Y. Choi, M. Ki, Y. M. Kim, J. W. Kim, J. H. Kim, E. W. Park, H. Y. Lee, S. C. Lim, C. H. Cho, S. R. Hong, J. Y. Dang, S. Y. Kim, Y. Kim, W. C. Lee and J. K. Lee, "The Korean guideline for cervical cancer screening," *Journal of gynecologic oncology*, vol. 26, no. 3, pp. 232-239, Jul. 2015.

APPENDIX

**Table. 2** General characteristics of female cancer screening

Variable	Division	Cervical cancer screening			Breast Cancer screening		
		N	Check rate(%)	p-value	N	Check rate(%)	p-value
<b>Demographic factor</b>							
Age	under 30-39 year	21,166	43.0	<.0001			<.0001
	under 40~50year	23,754	62.3		23,754	58.3	
	under 50~60year	22,818	64.3		22,820	66.0	
	over 60 year	41,227	44.9		41,269	51.2	
Level of education	Below middle school graduation	57,465	49.7	<.0001	56,932	54.8	<.0001
	High school graduation	31,291	56.1		21,729	60.4	
	More than a college	19,910	54.6		8,906	62.9	
Residence	Big city	28,911	53.8	<.0001	22,065	56.4	0.4904
	City	41,363	53.3		31,840	56.6	
	Farming and fishing village	38,691	50.4		33,938	57.7	
Marriage	Yes	104,896	53.4	<.0001	86,624	57.2	<.0001
	NO	4,041	25.9		1,196	9.2	
Job	Non-physical labor	12,051	55.8	<.0001	5,705	64.9	<.0001
	Physical labor	34,481	57.2		30,567	61.8	
	Inoccupation	55,698	48.7		45,956	52.8	
Income	Less than 1 million won	24,961	44.7	<.0001	24,219	52.2	<.0001
	100~200 under	19,730	51.2		16,843	57.4	
	200~300 under	18,016	52.4		13,090	57.8	
	300 over	36,519	59.2		25,636	62.5	
<b>Health Behavioral Factors</b>							
Smoking	Yes	3,713	38.9	<.0001	3,079	40.9	<.0001
	NO	105,210	52.9		84,732	57.6	
Drinking	Yes	67,833	54.4	<.0001	49,988	58.9	<.0001
	NO	41,086	49.2		37,811	54.4	
A regular meal	Yes	94,120	52.9	<.0001	78,930	57.5	<.0001
	NO	14,845	49.1		8,913	52.8	
Regular Sleep	Yes	61,566	52.0	0.0004	46,854	56.9	0.6874
	NO	47,399	53.0		40,989	57.1	
Physical activity	Nomal	21,590	58.6	<.0001	18,058	63.9	<.0001
	lack	86,976	50.9		69,474	55.3	
BMI	under 18.5,Underweight	14,008	39.5	<.0001	11,289	44.8	<.0001
	under 18.5-25(normal)	74,307	54.4		58,588	58.6	
	25over(obesity)	20,650	54.1		17,966	59.5	
Chronic disease	Yes	46,699	51.6	0.0012	45,744	56.7	<.0001
	NO	62,099	53.1		41,952	57.4	
Stress Index	Very much	27,988	51.2	<.0001	22,061	55.9	<.0001
	A little	58,046	53.8		45,314	58.6	
	Almost none	22,672	50.8		20,210	54.9	
EQ-5D index	0.2under	1,016	17.8	<.0001	997	19.1	<.0001
	under 0.2-0.5	823	28.1		811	34.1	
	> 0.5	107,126	52.9		86,035	57.6	
<b>Possible factor</b>							
Vaccination	Yes	48,022	53.8	<.0001	42,361	58.8	0.0004
	NO	60,723	51.4		45,300	59.1	
Health screenings	Yes	70,183	71.3	<.0001	61,432	77.1	<.0001
	NO	38,713	18.3		26,348	10.1	
Other cancer screenings	Yes	53,380	83.9	<.0001	49,532	89.9	<.0001
	NO	13,639	89.8		7,649	70.9	
Private insurance	Yes	69,982	59.1	<.0001	51,029	64.4	<.0001
	NO	38,684	40.5		36,526	46.7	



APPENDIX  
Table. 3 Multiple logistic regression analysis

variable	division	Cervical cancer screening (OR)			Breast Cancer screening (OR)		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Demographic factor							
		17,870,828	17,650,780	5,648,962	13,397,539	13,159,512	4,585,552
Age	AIC						
Cervical (under 30-39 year)	under 40-50 year	2.14	2.053	0.881*	1.596*	1.502*	1.289*
Breast (under 40-49 year)	under 50-60 year	2.73*	2.515*	0.684	0.902	0.862	0.853
	over 60 year	1.40	1.333	0.415	1.189	1.189	1.178
Level of education	High school graduation	1.22	1.213	1.217	1.359*	1.355*	1.239*
(Below middle school graduation)	More than a college	1.41*	1.392*	1.225*			
Residence	City	1.02	1.006	1.444*	0.812	0.792	0.798
(Farming and fishing village)	Big city	1.03*	1.019*	1.213	0.852	0.840	0.778
Marriage(No)	Yes	3.34*	3.146*	3.764*	1.911*	1.750*	1.116*
Job (Inoccupation)	Physical labor	1.24*	1.259*	0.935	1.296*	1.335*	1.051
	Non-physical labor	1.17	1.150	1.168*	1.199	1.182	1.088*
Income	100~200 under	1.13	1.102	1.054	1.099	1.056	0.884
(Less than 1 million won)	200~300 under	1.20	1.159	1.033	1.128	1.078	0.863
	300 over	1.52*	1.450*	1.127*	1.384*	1.305*	0.963*
Health Behavioral Factors							
Smoking (NO)	Yes		0.721	0.858		0.574	0.739
Drinking (NO)	Yes		1.166*	1.028*		1.116*	0.993
A regular meal(NO)	Yes		1.202*	1.056*		1.317*	1.122*
Regular Sleep(NO)	Yes		0.977	1.007*		0.966	0.985
Physical activity(lack)	Normal		1.220*	1.056*		1.260*	1.123*
BMI(under 18.5, Underweight)	under 18.5-25(normal)		1.370*	1.285*		1.514*	1.376
	25over(obesity)		1.278	1.197		1.481	1.537*
Chronic disease (NO)	Yes		1.148*	0.840		1.202*	0.933
Stress Index	A little		0.969	0.933		0.938	0.954
(Almost none)	Very much		0.994	1.068*		0.979	0.978
EQ-5D index	under 0.2-0.5		1.573	1.060		1.937	2.406*
(0.2under)	> 0.5		3.645*	1.389*		4.141*	2.230
Possible factor							
Vaccination (NO)	Yes			1.040*			0.956
Health screenings(NO)	Yes			1.659*			5.575*
Other cancer screenings (NO)	Yes			0.720			2.876*
Private insurance(NO)	Yes			1.532*			1.386*

by t-test or chi-square test (\*: p<0.05 o|h)



**박민희(Min-Hee Park)**

2007: 충북대학교 행정학과 행정학석사  
2015: 충북대학교 의학과 박사과정 수료  
현재: 박종범 내과 초음파실 근무

※ 관심분야: 의료영상처리, 초음파영상처리, 건강증진, 원격의료



**조영복(Young-Bok Cho)**

2005: 충북대학교 전자계산학과 공학석사  
2012: 충북대학교 전자계산학과 공학박사  
2016: 충북대학교 의학과 박사과정수료  
2012-2018: 충북대학교 소프트웨어학과 초빙교수  
현재: 대전대학교 정보보안학과 조교수

※ 관심분야: 의료영상처리, 정보보안, 의료정보보호, 모바일보안



**김소영(So Young Kim)**

2007: 서울대학교 의과대학원 의료관리학과 의학석사  
2010: 서울대학교 의과대학원 의료관리학과 의학박사  
현재: 충북대학교병원 교수

※ 관심분야: 건강형편성, 장애인보건의정, 보건의료빅데이터,



**박종배(Jong-Bae Park)**

2004: 대전대학교 생물학과 이학석사  
2011: 대전대학교 생물학과 이학박사  
현재: 충북 보건과학대학교 방사선과 교수

※ 관심분야: 생물정보학, 의료영상처리, 의료정보보호,



**박종혁(Jong-Hyock Park)**

2006: 서울대학교 보건대학원 보건정책학과 보건학석사  
2008: 서울대학교 의과대학원 의료관리학과 의학박사  
2009: 국립암 센터 보건복지정책고위과정 수료  
현재: 충북대학교 의과대학, 의생명과학경영융합대학원 교수

※ 관심분야: 암정책, 보건관리정책, 금연정책, 보건의료빅데이터.