Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 34, No. 3, 231-242, 2016
http://dx.doi.org/10.7848/ksgpc.2016.34.3.231

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

병원위치정보를 이용한 지리적 거리기반의 대기환경 데이터셋 구축

A Construction of Geographical Distance-based Air Quality Dataset Using Hospital Location Information

김형수[1] · 류근호[2]
Kim, Hyeongsoo · Ryu, Keun Ho

Abstract

As of late, air quality information has been actively gathered and investigated in order to find possible environmental risk factors that may affect the onset of cardiovascular disease. Nevertheless, existing studies are limited in the detailed analysis because they take advantage of the air quality information of the macro statistics divided into administrative districts. This paper proposes the construction of distance-based air quality dataset using a domestic hospital's geographical location information as a reliable data gathering step for a more detailed analysis of environmental risk factors. For the construction of the dataset, air quality information was obtained by utilizing the geographical location of a hospital—in which a patient with cardiovascular disease had been admitted—and then matching the hospital with a meteorological and air pollution station in its vicinity. An air quality acquisition system based on GMap.net was devised for the purpose of data gathering and visualization. The reliability of the experiment was confirmed by evaluating the matching rate and error of air quality values between the acquired dataset with existing area-based air quality datasets from matched distances. Therefore, this dataset, which considers geographical information, can be utilized in multidisciplinary research for the discovery of environmental risk factors that can affect not only cardiovascular diseases but also potentially other epidemic diseases.

Keywords : Geographical Information, Distance-based, Air Quality, Acquisition System, GMap.NET, Hospital Location

초 록

최근 심혈관계질환의 발병에 영향을 미칠 수 있는 환경적 위험요인을 찾기 위해 대기환경정보를 활용한 다양한 연구가 활발히 진행되고 있다. 그러나 기존 연구들은 대기환경정보에 대해 행정구역 단위로 구분된 광의적인 통계자료를 활용하였기 때문에 세밀한 분석을 하기에는 다소 제한적이다. 이에 본 논문에서는 보다 세밀한 분석을 위한 신뢰성 높은 데이터셋을 수집 및 구축하는 단계로써, 전국 주요병원의 지리학적 위치정보를 이용한 거리기반의 통합 대기환경 데이터셋 구축 방법을 제안한다. 데이터 구축을 위해, 전국 병원들 중 심혈관계질환자가 내원한 병원을 기준으로 인접한 대기환경관측소와의 거리계산을 통해 매칭하였으며, GMap.net 기반의 대기환경정보 획득시스템 개발을 통해 데이터 획득 및 시각화에 활용하였다. 또한, 기존 행정구역으로 구분한 지역기반 대기환경정보와의 비교평가를 통해 거리, 매칭률 및 정보의 오차에 대한 신뢰성을 확보하였다. 이렇게 구축된 통합데이터는 심혈관계질환 뿐 아니라 다양한 발병에 영향을 미칠 수 있는 환경적 위험요소 발견을 위한 다각적 연구에 보다 신뢰성 있는 기초정보로 활용될 수 있을 것으로 기대된다.

핵심어 : 지리정보, 거리기반, 대기환경정보, 획득시스템, GMap.NET, 병원위치

Received 2016. 02. 12, Revised 2016. 03. 21, Accepted 2016. 05. 17

1) Member, Department of Computer Science, College of Electrical and Computer Engineering, Chungbuk National University (E-mail: hskim@dblab.chungbuk.ac.kr)
2) Corresponding Author, College of Electrical and Computer Engineering, Chungbuk National University (E-mail: khryu@dblab.chungbuk.ac.kr)

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.

1. 서 론

최근 발표된 세계보건기구의 보고서(WHO reports, 2015)에 따르면, 최근 10년간 질병에 의한 사망 원인 중 심혈관계질환은 세계적으로 상위에 랭크되어있으며, 그것에 의한 사망률은 꾸준히 증가하고 있다. 그럼에도 불구하고, 심혈관계질환은 발병에 영향을 미칠 수 있는 위험요소들을 관리함으로써 예방이 가능하기 때문에, 발병을 최소화하기 위한 다양한 연구가 진행되고 있으며, 주요 연구결과를 통해 심혈관질환에 영향을 미치는 위험인자가 최근까지 진단지표로 활용되고 있다(Wilson *et al*., 1998; Yusuf *et al*., 2004).

특히 최근에는 임상정보와 다양한 분야에서 발생할 수 있는 위험요인들을 함께 고려한 융합 연구가 중요시됨에 따라, 다양한 분야와 연계된 통합데이터 구축의 필요성이 대두되고 있다. 그 중 사람이 쉽게 영향을 받을 수 있는 대기환경정보는 다양한 형태로 구축되어 활용되고 있으며, 일부 연구를 통해 환경적 요인이 심혈관계질환의 발병에 영향을 미친다는 것을 결과로 제시하고 있다(Amiya *et al*., 2009; Cao *et al*., 2009; Goggins *et al*., 2013; Radišauskas *et al*., 2014; Vanos *et al*., 2014). 국내의 경우 기후와 대기오염이 미치는 영향에 대한 분석을 통해 의미 있는 연관성을 찾는 연구가 진행되었다(Joo, 2014; Lee *et al*., 2010; Park *et al*., 2015; Yang, 2004b).

하지만 대부분의 기존 연구들은 환자로부터 획득한 임상정보만을 대상으로 분석되었거나 대기환경정보의 경우 특정 기간 또는 지역으로 구분된 광의적인 통계자료를 이용하여 분석하였기 때문에 보다 신뢰성 있고 유용한 정보를 추출할 수 있는 세밀한 분석의 어려움이 있다. 따라서 세밀한 분석을 통해 보다 정확한 위험인자를 찾기 위해서는, 기존의 광의적인 통계자료가 아닌 신뢰성을 확보한 데이터를 수집할 수 있는 연구가 선행되는 것이 중요하다.

이에 본 연구는 심혈관계질환을 대상으로 기상기후, 대기오염 등 대기환경정보가 미치는 상관성 분석을 위한 데이터 수집 단계로써, 신뢰성 높은 대기환경정보를 수집하고 공유할 수 있는 데이터셋의 구축을 목표로 한다. 따라서 지리적 위치를 활용한 GMap.NET 기반의 대기환경정보 획득 시스템을 개발하고 이를 통해 데이터셋을 구축하였다. 데이터셋은 심혈관계질환자들이 내원한 병원을 대상으로 지리적 거리 및 위치정보를 이용하여 가장 인접한 대기환경관측소를 매칭하고, 해당 관측소로부터 관측된 대기환경정보를 매칭 조건에 따라 수집하여 구축하였다. 아울러, 생성된 데이터셋의 신뢰성을 평가하기 위해 기존 정보와의 비교평가를 실시하였으며, 공간적 연결성이나 인접성과 같은 공간분석을 위해 시스템을 통해 시각화하였다. 이렇게 구축된 통합 데이터셋은 심혈관계질환 뿐 아니라 다양한 질병과 관련된 다각적 연구에 적용하여 발병에 영향을 미칠 수 있는 보다 신뢰성 있는 환경적 위험요소를 발견하기 위한 기초정보로 활용될 수 있을 것으로 기대된다.

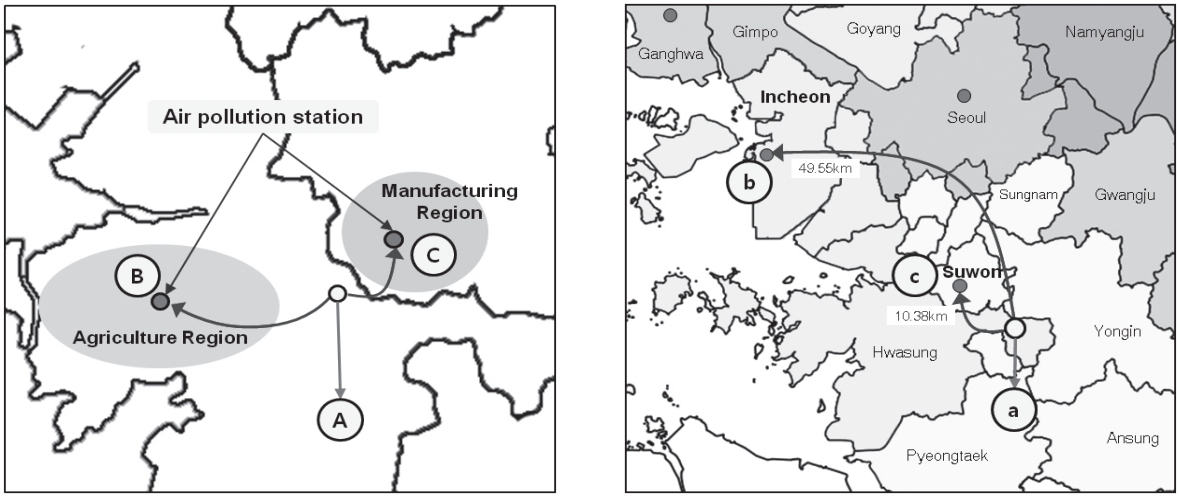
2. 관련연구

2.1 이론적 배경

최근에는 국내외적으로 단순 임상정보만을 이용한 연구가 아닌 다양한 분야에서 획득한 정보를 함께 적용한 융합적인 연구가 활발히 이루어지고 있다. 기온, 습도, 대기오염물질 등과 같은 대기환경정보는 일상생활에서 쉽게 영향을 받을 수 있기 때문에 다양한 형태로 각 연구 분야에 적용되고 있으며, 다양한 대기환경 속성이 각종 질병과 연관성이 있다는 연구결과가 도출되고 있다.

Radišauskas *et al*.(2014)는 몇 가지 기상요소를 분석한 결과 기온과 급성심근경색 사망률 사이의 상관관계가 있으며, 풍속이 여성 및 고령인구의 사망률에 영향을 미치는 것을 발견하였다. 또한, 월별 및 계절의 변화를 통해 겨울철에 사망률이 높은 결과를 보여준다. Cao *et al*.(2008)와 Amiya *et al*.(2009)는 심혈관계질환과 기상조건간의 관계에 대한 연구를 통해 일교차가 심혈관계질환에 영향을 미치는 독립적인 위험요인임을 제안하였으며, 일교차를 통해 발병을 예측할 수 있음을 보였다. Goggins *et al*.(2013)은 기상조건과 대기오염의 연관분석을 통해 임계온도를 기준으로 1℃가 감소하거나, 이산화질소 농도가 10㎎/㎥ 증가할 때 발병률이 증가한다는 결과를 보여주고 있다. Vanos *et al*.(2014)는 캐나다 주요도시를 대상으로 계절별 날씨유형 및 대기오염에 대한 연구로 고온 건조한 날씨유형에 발생하는 대기오염 농도가 심혈관계질환의 사망률에 가장 큰 영향을 미친다는 연구결과를 보여준다.

국내의 경우, Park *et al*.(2015)은 미세먼지와 이산화황이 심혈관계질환의 일별 사망에 미치는 영향을 추정하였으며, SO2 농도가 11.67ppb 증가함에 따라 심혈관계 사망 위험도가 8.6% 증가한다는 연구결과를 제시하였다. Yang(2004a)은 서울시에 대한 대기오염이 질병별 사망자 수에 미치는 영향분석을 통해 미세먼지가 1㎍/㎥ 증가할 때 사망자 수가 0.1% 증가한다는 분석을 통해 미세먼지농도가 질병에 의한 사망률에 영향을 미친다는 결과를 보여준다. Joo(2014)는 주요 국가 산업단지를 대상으로 대기오염이 호흡기질환 및 심혈관질환의 일별 사망에 미치는 영향 추정을 통해 대기오염의 농도에 대한 관련성을 제시하였다. Lee *et al*.(2010)은 심혈관계질



(a) A case of air pollution information

(b) A case of meteorological information

Fig. 1. An example of projecting unreliable air quality information

환에 대한 기상 매개 변수와의 관계에 대한 연구를 통해 기상매개변수가 심혈관계질환의 발병에 영향을 미친다는 결과를 보여준다.

하지만, 위에서 기술한 기존 연구들은 몇 가지 단점을 가지고 있다. 첫째, 상관성 분석에 적용한 대기환경정보의 경우 특정 기간의 평균값을 활용하고 있다. 둘째, 특정 지역 또는 행정구역단위의 통계자료를 활용함으로써 지역기반의 광의적인 분석만을 진행하였다. 셋째, 질병에 대한 사망과의 상관성을 분석함으로써 환자에 관한 분석 또는 예방차원의 분석이 다소 제한적이다. 이렇듯, 대부분 광의적인 분석을 통해 대기환경정보가 심혈관계질환에 영향을 미치는 연관성에 대한 분석은 보다 신뢰성 있고 유용한 정보를 도출하는데 제한적이다. 따라서 이러한 문제점을 보완하기 위해서는 객관적인 신뢰성을 확보할 수 있는 데이터 구축에 관한 연구가 필요하다.

2.2 기존 대기환경데이터 분석

최근 정부3.0의 공공데이터포털을 통해 다양한 대기환경정보를 제공하고 있지만, 광의적인 수준의 통계자료 공유에 그치고 있으며, 지자체간 데이터 형식과 내용의 차이 등으로 인해 통합된 데이터로 재구축해야 하는 필요성이 있다. 특히, 행정구역을 기반으로 적용된 데이터가 대부분이기 때문에 행정경계에 인접한 지역의 대기환경정보에 대한 모호성이 발생할 수 있다. Fig. 1은 지리적 위치를 고려하지 않고 단순히 행정구역별로 구분하여 대기환경정보를 적용할 사용할 경우, 정보의 신뢰성이 모호한 두 가지 예를 보여준다.

Fig. 1(a)는 대기오염정보와 관련하여 특정 병원(Ⓐ)이 농경지역과 공업지역의 행정경계에 위치한 경우로써, Ⓐ는 농경지역이 포함된 행정구역으로 구분되어 대기오염관측소(Ⓑ)에서 관측된 정보가 적용된다. 하지만, 지리적 위치와 대기오염 범위 및 대기확산 등을 고려하였을 때 Ⓐ는 공업지역(Ⓒ)의 영향을 보다 많이 받을 수도 있다.

Fig. 1(b)는 기상정보와 관련하여 병원이 행정경계 근처에 위치한 경우, 실제 종관기상관측소에 대한 정보의 적용에 대한 모호성을 예로써 보여준다. 예를 들어, 특정병원(ⓐ)의 경우 행정구역상 “화성시”에 위치해 있기 때문에 “경기남부서해안” 지역으로 구분되어 인천 관측지점(ⓑ)의 기상정보가 적용된다. 하지만 ⓐ는 기존에 적용된 인천지점으로부터 49.55km에 위치해 있으며, 거리상 인접한 수원지점(ⓒ)으로부터 10.38km에 위치하고 있다. 이러한 경우 행정구역은 다르지만 거리상으로 인접한 관측소의 정보를 적용하는 것이 정보의 신뢰성을 높일 수 있는 방법이 될 수 있다. 따라서, 이 논문에서는 관측된 대기환경정보의 오차를 최소화하고 신뢰성을 높이기 위해 지리적 거리를 고려하여 데이터셋을 구축하였다. 또한, 위에서 언급한 두 가지 예제에 대해 5장에서 실제 데이터를 적용하여 기존 데이터와 비교하였다.

3. 연구 방법

본 논문에서는 2006년 1년간의 기간을 기준으로 KAMIR (Korean Acute Myocardial Infarction Registry) 데이터에 포

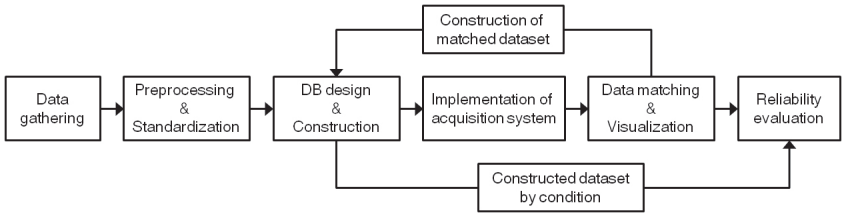


Fig. 2. Research process for reliable air quality dataset

함된 전국 주요병원을 대상으로 하였다. 의료기관의 경우 크게 1~3차 의료기관으로 분류될 수 있다(Yang, 2004a). 평가에 사용된 병원정보는 3차 의료기관에서 수집한 데이터이며, 1, 2차 의료기관으로부터 후송된 정보를 포함하고 있다. 특정 병원으로부터 거리에 대한 적용범위는 An(2006)과 Bang(2012)의 응급후송 시간 및 거리와 Lee and Park(2004)의 진료권 분석을 토대로 하여 반경거리를 최대 20km로 정하였다. 대기환경정보는 OpenAPI를 통해 각각의 관측소로부터 측정된 공공데이터를 활용하였으며, 데이터셋 구축은 Fig. 2와 같이 체계적인 단계를 통해 진행되었다. 특히, 본 연구에서 개발된 대기환경정보 획득 시스템을 통해 새롭게 생성된 데이터셋은 재사용성을 고려하여 데이터베이스에 새로운 테이블로 구축되어 실험평가에 활용되었다.

3.1 사용데이터 정의

병원에 대한 위치기반의 대기환경정보를 획득하기 위해서, OpenAPI를 통해 기상청, 건강보험심사평가원, 국립환경과학원에서 제공하는 공공데이터를 수집하였다. 데이터베이스 구축에 활용된 각각의 원시데이터에 대한 설명은 다음과 같으며, Table 1은 데이터 속성에 대해 요약하여 보여준다.

1)KAMIR 데이터 : 전국 주요 45개 대학 및 거점병원에서 수집한 6,345명의 급성심근경색환자 데이터로써 141개 속성의 임상정보를 포함하고 있으며, 다양한 분야에서 기초자료로 활용되고 있다(Ryu *et al*., 2015; Shon *et al*., 2013). 본 연구에서는 데이터구축에 필요한 병원명, 환자가 내원한 날짜, 시간 등 기본정보를 활용하였으며, 대기환경정보와 함께 추후 다양한 분석을 위한 기초 자료로 사용되어질 것이다.

2)기상 데이터 : 전국 79개 종관기상관측소와 476개의 방재기상관측지점에서 측정한 기상관련 데이터이며, 각각의 속성은 특정시간 간격으로 관측되어 있다.

3)대기오염 데이터: 도로변 대기정보를 제외한 5가지 오염물질에 대해 전국 257개의 관측소로부터 1시간 단위로 측정된 데이터이다.

4)전국병원 데이터: 전국 66,287개의 병원에 대한 위치정보를 포함하는 데이터이며, 이 중 대학병원, 개인병원, 요양병원 등 심장질환의 증상을 보일 경우 방문할 수 있는 33,828개의 병원들을 대상으로 적용하였으며, 치과와 성형외과에 대해서는 제외하였다.

5)대기환경관측소 위치 데이터: 지리학적 위치기반의 대기환경정보 수집을 위한 기상 및 대기오염관측소의 위치정보를 포함한다. 기존 원시데이터의 경우 좌표체계가 서로 상이하여, 데이터 전처리 단계에서 WGS84 좌표체계로 통일되게 변환하여 사용하였다.

Table 1. Summary of the used datasets

Dataset	Missing Values	Instances	Amount of Attributes			Description
			total	Nominal	Numeric	
KAMIR	Yes	6,345	141	114	27	Hospital location information about 6,345 cardiovascular disease patients.
Meteo_Info	No	194,571	36	5	31	Meteorological information about 555 observation stations.
Poll_Info	Yes	1,857,456	7	2	5	Air pollution information about 257 observation stations.
Hos_Info	Yes	33,828	14	12	2	Nation's hospital dataset
Meteo_STN	No	555	4	2	2	Location information of Meteorological observation station.
Poll_STN	No	257	4	2	2	Location information of Air pollution observation station.

3.2 데이터 표준화 및 전처리

대기환경정보 획득 시스템의 데이터베이스 구축을 위해 사용데이터를 분석한 결과, 각 기관으로부터 제공된 원시데이터의 서로 다른 좌표체계와 이름이 중복되는 병원 등 몇 가지 문제점을 발견하였으며, 위치정보 수집 및 배포, 활용과정을 고려한 표준화가 필요한 것으로 분석되었다. 따라서 본 연구에서는 표준화 및 데이터전처리 과정을 통해 통일된 좌표체계와 중복되는 병원들 중 실제 적용할 병원의 정확한 위치정보를 선택하였다. 통일된 좌표계를 통한 거리계산 및 위치표현을 위하여 좌표체계는 구글맵에서 적용하고 있는 WGS84 좌표계로 변환을 하여 모든 데이터의 좌표에 대해 표준화하였다. 좌표변환은 OpenAPI를 활용하여 변환하였으며, 실제 위치와 일치하는지 테스트를 통해 정확한 위치를 표현하는지 확인하였다.

동일한 병원명의 위치정보에 대한 전처리는 건강보험심사평가원으로부터 제공받은 전국의료기관의 위치좌표를 WGS84로 변환한 후 병원명 검색을 통해 대상병원의 위치를 결정할 수 있는 프로토타입을 개발하여 활용하였다. Fig.3은 개발한 프로토타입을 통해 동일한 이름의 병원에 대한 위치결정의 처리과정을 예로써 보여준다. 예를 들어, '하나병원'의 경우 4개의 동일한 이름을 가지는 병원이 검색되고, 이들의 주소(A) 및 지리적 위치(B)를 확인할 수 있다. 여기에서 임상정보를 입력한 3차 병원과의 인접성을 고려하여 가장 인접한 병원을 선택하여 위치정보를 결정하였다. 이는 응급 후송을 가정하였을 때, 일반적으로 가까운 거리를 이동해야 한다는 가정을 통해 적용하였다. 만약 3차병원이 '충북대학교병원(C)'이라면, 이는 1차 개인병원 또는 2차 일반병원으로부터 후송된 환자의 경우로써 후송에 소요되는 시간을 고려하였을 때 지리적으로 가장 인접한 청주지역의 하나병원을 선택함으로써 정확한 위치정보를 저장하도록 하였다. 또한, 인접한 2차병원과의 거리계산을 통해 20km의 범위를 벗어나는 병원에 대해서는 배제하였다.

Fig. 3. An example of selecting duplicated hospital name

4. 대기환경정보 획득시스템 구현 및 데이터셋 구축

본 연구에서는 GMap.NET(GMap.NET, 2015)을 기반으로 구축된 데이터베이스로부터 대기환경정보를 획득할 수 있는 시스템을 개발하였다. 또한, 다양한 조건의 대기환경정보를 제공 및 신뢰도에 대한 비교평가를 위해 지역기반 및 위치기반으로 관측소를 분류하여 개발된 시스템을 통한 조건별 데이터셋을 생성 및 구축하였다.

4.1 데이터베이스 구축

데이터베이스는 대기환경정보 획득시스템 개발을 위해 MySQL을 이용하여 구성하였다. 각 기관의 원시데이터로부터 표준화 및 전처리 과정을 통해 아래 Fig. 4와 같이 기본적인 테이블을 생성하였다. KAMIR 병원의 지리적 정보와 환자가 병원에 내원한 날짜 속성을 이용하여 대기환경관측소로부터 매칭되는 정보를 획득할 수 있다. 또한, 시스템을 통해 생성된 데이터셋은 정보 공유를 위해 데이터베이스에서 새로운 테이블로 생성 및 저장된다.

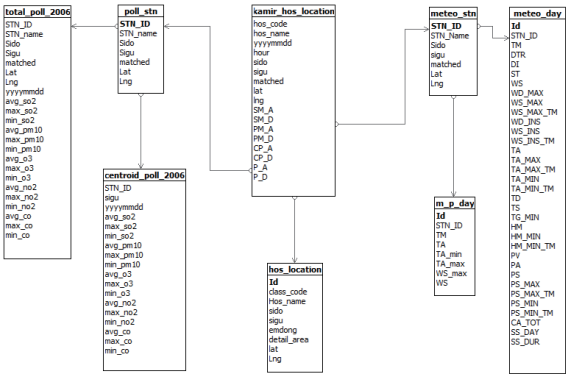
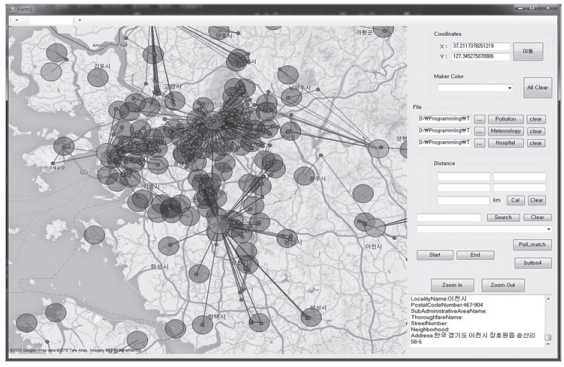


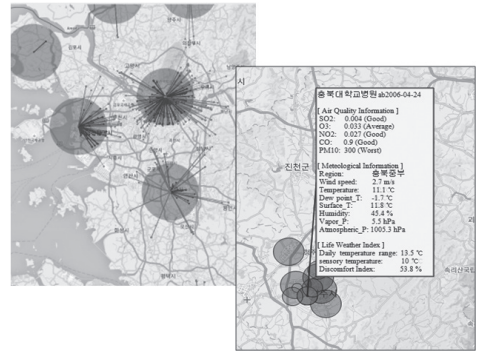
Fig. 4. Database diagram for air quality acquisition system

4.2 시스템 구현 및 시각화

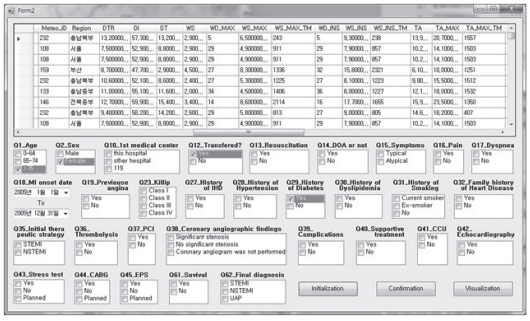
GMap.NET은 닷넷 프레임워크에 기반하며 크로스 플랫폼과 오픈 소스를 지향하는 강력한 지도 서비스 관련 무료 개발 도구이다. 따라서 본 연구에서는 GMap.NET을 이용하여 병원과 각 환경정보 관측지점에 대한 지리적 위치를 통해 조건별로 대기환경관측소와의 매칭을 통해 데이터셋을 생성 및 구축할 수 있는 시스템을 개발하여 실험평가에 이용하였다. 시스템은 C# 언어를 사용하여 Visual Studio 2013 환경에서 개발하였다. 개발된 시스템은 대상 병원의 위치를 기준으로

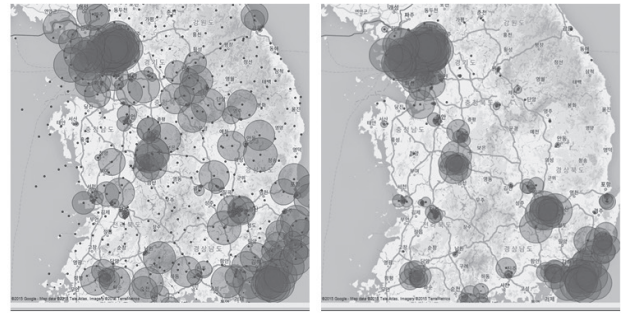


(a) The main page of the system



(b) Visualization about the matching result





(c) Geographical distribution by exploiting clinical information

Fig. 5. A visualized example of the air quality acquisition system

조건에 맞는 대기환경관측소를 매칭하여 이들 관측소로부터 기상정보와 대기오염정보를 획득한다. 또한, 매칭을 통한 데이터셋 생성뿐 아니라, 공간적 연결성이나 인접성과 같은 위치기반의 공간분석을 위해 Fig. 5와 같이 시각화하여 지리적 위치의 상호접근성에 대한 정보를 제공한다.

4.3 데이터셋 생성

지역기반 및 위치기반의 대기환경정보 매칭을 위해, 기상의 경우 종관기상관측소 (79지점)과 종관기상과 방재기상관측을 함께 고려한 종관-방재기상관측소(555지점)을 적용하였으며, 대기오염의 경우 대기오염관측소(257지점)와 동일지역에 대한 중심점을 계산한 중심점-대기오염관측소(78지점)를 적용하였다. 즉, 총 8가지의 분류를 통해 데이터셋을 생성하였으며, 생성된 데이터셋은 실험평가 단계에서 각각의 비교평가에 활용하였다. 관측소 매칭에 필요한 지리적 거리 및 중심좌표 계산은 Haversine(Sinnott, 1984)과 Geographic Midpoint(GeoMidpoint, 2015)를 적용하였다.

4.3.1 지리적 거리 계산

병원의 위치기반 대기환경정보 매칭을 위해, 위경도 좌표를 이용한 거리계산을 통해 가장 인접한 관측소를 매칭하였으며, 두 좌표지점간 거리는 Eq.(1)과 같이 구면기하 삼각함수를 이용한 Haversine 공식을 사용하였다. Eq.(2)는 두 좌표의 거리 d를 산출하는 공식이다.

$$a = \sin^2(\Delta\phi/2) + \cos\phi_1 \cdot \cos\phi_2 \cdot \sin^2(\Delta\lambda/2) \tag{1}$$

$$c = 2 \cdot atan2(\sqrt{a}, \sqrt{(1-a)}) \tag{2}$$

$$d = R \cdot c$$

where, ϕ: Latitude, λ: Longitude, $\Delta\phi$: ϕ_2-ϕ_1, $\Delta\lambda$: λ_2- λ_1, R: Earth's radius

위의 공식을 통해 병원이 위치한 좌표로부터 모든 대기환경관측소의 좌표까지의 거리를 계산하고, 그 중 가장 거리가 가까운 관측소와의 매칭을 통해 대기환경정보를 수집한다. 여기에서 환자 후송 가능 범위를 고려한 허용거리를 약

20km로 정하였으며, 이 거리를 벗어나는 지점에 대해서는 제외하였다.

4.3.2 지리적 중심좌표

대기오염 데이터는 일반적으로 동일 행정구역에 대한 평균값을 활용하고 있다. 이러한 행정구역별 평균값을 가지는 대기오염정보를 구하기 위해, GeoMidpoint (2015)에서 제시하고 있는 Geographic Midpoint Method를 적용하여 지역기반의 대기오염관측소를 분류하였다. 즉, 동일 행정구역에 포함되는 관측소들에 대한 중심점을 계산하고 이를 기준으로 해당하는 지역별 대기오염정보의 평균값을 산출하였다. Fig. 6은 중심점을 계산하는 과정에 대한 예를 보여주며, 전국 257개 대기오염관측지점의 경우 중심점 계산을 통해 78개의 중심점-대기오염 관측지점으로 분류된다.

중심좌표에 대한 계산 알고리즘은 Fig. 7과 같다. 여기에서 가중치*(weight)*의 경우 모든 좌표에 대해 동일한 날짜가 적용되므로 모든 가중치($w_1,w_2,...,w_n$)에 대해 1의 값이 적용하여 가중치에 대한 계산을 생략할 수 있다.

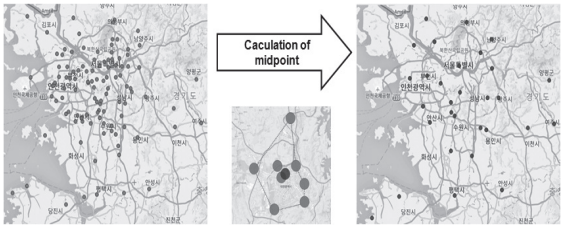


Fig. 6. Midpoint calculation process using geographical midpoint method

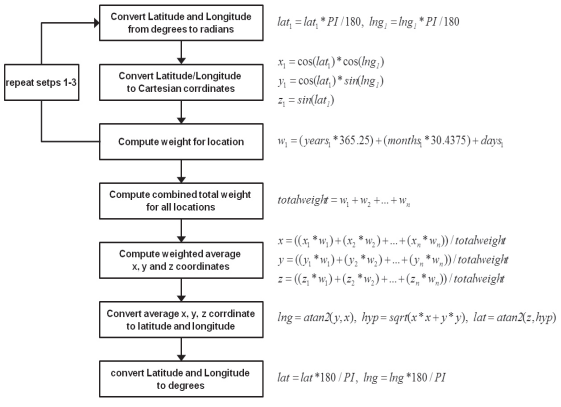


Fig. 7. Algorithm of Geographical Midpoint Method

5. 실험평가 및 결과

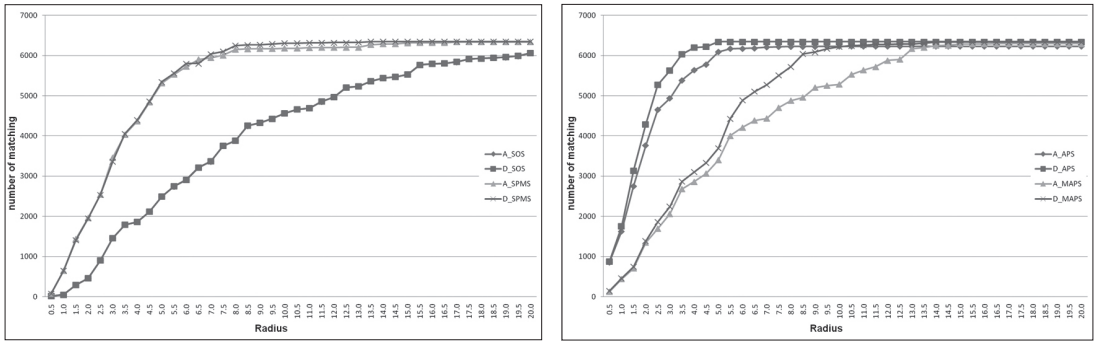
본 연구에서는 조건별로 획득된 데이터셋에 대해 각각 매칭관측소 수, 매칭 거리 및 대기환경정보의 차이에 대해 평가하였다. 기존 연구에 적용되는 데이터셋은 지역기반으로 관측소를 분류하고, 이를 통해 1~3시간 단위로 수집된 대기환경정보의 일별 평균값을 적용하고 있다. 따라서 기존 지역기반의 대기환경정보를 비교대상으로 하여, 본 연구에서 개발한 대기환경정보 획득 시스템을 통해 새롭게 구축되어진 데이터셋을 거리와 매칭률에 대해 조건별 비교를 통한 신뢰성을 평가하였다. 평가는 결측치를 제외한 6,345개 병원을 대상으로 하였으며, Table 2과 같이 기상관측소와 대기오염관측소를 지역기반과 거리기반으로 분류한 총 8개 데이터셋을 적용하여 비교하였다. 기상관측소는 지역기반의 종관기상(A_SOS)과 종관+방재기상(A_SPMS), 거리기반의 종관기상(D_SOS)과 종관+방재기상(D_SPMS)으로 구분하였으며, 대기오염관측소는 지역기반의 대기오염(A_APS)과 중심점-대기오염(A_MAPS), 거리기반의 대기오염(D_APS)과 중심점-대기오염(D_MAPS)으로 각각의 데이터셋을 생성하여 최대 20km의 반경을 적용하여 비교하였다.

Table 2. Summary of the used dataset

Dataset	amount of station	Description
A_SOS, D_SOS	79	Synoptic weather Observation Station (SOS)
A_SPMS, D_SPMS	555	SOS with Prevention Meteological Station (SPMS)
A_MAPS, D_MAPS	78	Midpoint Air Pollution Station (MAPS)
A_APS, D_APS	257	Air Pollution Station (APS)

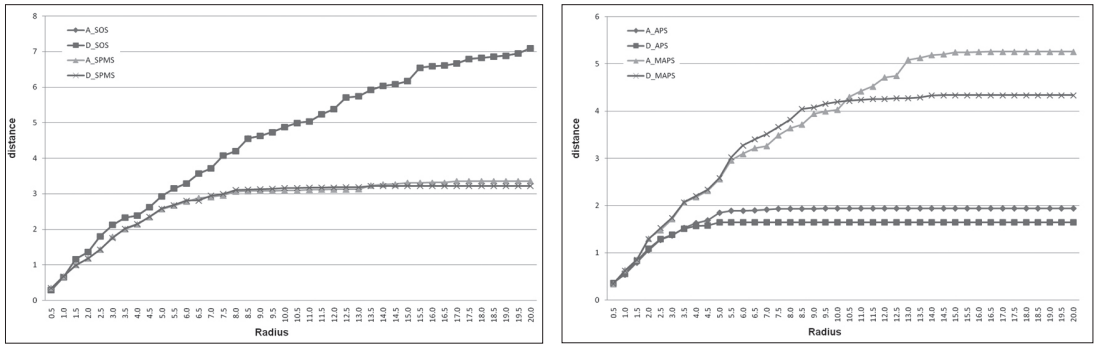
* A_ : Area-based, D_ : Distance-based

Fig. 8은 대상병원과 기상 및 대기오염 관측소의 매칭수를 비교한 결과를 보여준다. 기존 연구들에서 활용하고 있는 종관기상의 경우, 지역과 거리기반에 대한 큰 차이가 없었으며 10km 범위에서 약 72%의 매칭률을 보였다. 하지만, 종관+방재기상의 경우 10km 범위에서 약 90% 이상의 매칭률을 보였다(Fig. 8(a)). 대기오염관측소는 동일지역에 대한 중심점-



(a) The number of matched meteorological station (b) The number of matched air pollution station

Fig. 8. Matching rate of the air quality observation station



(a) Average distance of matched meteorological station (b) Average distance of matched air pollution station

Fig. 9. Average distance rate of matched air quality observation station

대기오염관측소(78개)의 경우 10km 범위에서 지역기반이 76.93%, 거리기반이 90.15%의 매칭률을 보였으며, 257개 대기오염관측소에 대한 매칭률의 경우 98.20%, 99.97%를 각각 나타내었다(Fig. 8(b)). 또한, 각 조건별 기준범위에 대한 평균매칭 거리를 Fig. 9와 같이 비교하였다. 기존 관측정보에 비해 새로 구축된 정보가 기상정보의 경우 약 3km, 대기오염정보의 경우 약 2km 이내에서 대부분의 병원이 대기환경정보를 적용할 수 있음을 알 수 있다. Table 3은 조건별 데이터셋에 대한 매칭률과 평균거리에 대한 비교결과를 요약하여 보여준다.

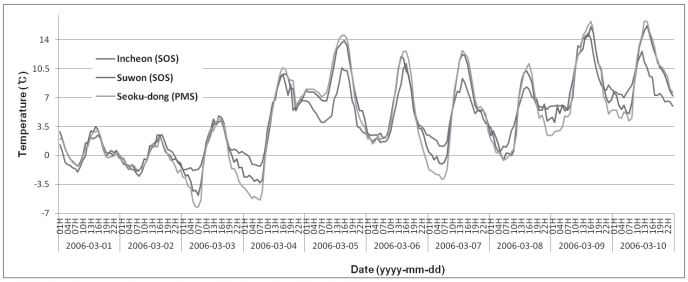
Table 3. A description of summary results

Matching condition			Matching rate(%)					Average distance(km)		
			2km	4km	6km	8km	10km	70%	80%	90%
Meteorological station	SOS	A	7.27	29.35	45.85	61.13	71.87	4.78	5.53	6.52
		D	7.27	29.38	45.88	61.17	71.90	4.76	5.52	6.50
	PMS	A	30.91	68.78	90.21	96.93	97.34	2.19	2.42	**2.76**
		D	30.76	69.27	91.39	98.44	99.45	2.19	2.40	**2.71**
Air pollution station	MAPS	A	21.23	45.11	66.38	76.93	83.23	3.31	3.84	4.51
		D	21.83	48.95	77.01	90.15	98.05	3.11	3.39	3.80
	APS	A	59.32	88.81	97.29	98.09	98.20	1.23	1.48	**1.66**
		D	67.63	97.68	99.97	99.97	99.97	1.17	1.24	**1.43**

Table 4. Observed information from Dongtan sungsim hospital

Station name	Distance	Average_TA(℃)	Average_DoT(℃)	Max_DoT(℃)	Min_DoT(℃)
Incheon	49.55km	4.05	1.96	5.9	0.4
Seoku-dong(***nearest***)	3.7km	4.52	-	-	-
Suwon	10.38km	4.65	0.77	2.9	0

* DoT : Difference of Temperature



* SOS : Synoptic weather Observation Station, PMS : Prevention Meteological Station

Fig. 10. A comparison of temperature between area-based and distance-based matching

Fig. 10 ~ Fig. 14는 2006년 3월 1일부터 10일까지 관측소로부터 측정된 기온과 대기오염정보를 1시간 단위로 적용하여 비교한 그래프이다. Fig. 10은 화성시에 위치한 '동탄성심병원'에 대해 가장 인접한 방재기상관측소의 기상정보를 기준으로 두 종관기상관측소인 인천관측소(지역기반), 수원관측소(거리기반)와의 기온차를 비교한 그래프이며, Table 4는 Fig. 10의 결과를 요약하여 보여준다. Table 4를 통해 인접한 수원관측소에서 관측된 정보와의 인천관측소에서 관측한 정보보다 기온의 차이가 적은 것을 알 수 있다. 또한 방재기상을 함께 고려하였을 경우 매칭거리와 기상정보의 신뢰성에 있어 효율적인 것으로 나타났다.

Fig. 11 ~ Fig. 14는 포항시 남구에 위치한 '대송의원'으로부터 2.3km 거리의 대송면지점(지역기반)과 1.1km 거리의 장흥동지점(거리기반)에서 관측한 4가지 대기오염정보(SO_2, O_3, NO_2, CO)를 적용하여 비교한 결과이다. 결과를 통해 '대송의원'의 경우 철강산업단지와 입접해 있기 때문에 장흥동지점의 대기오염수치에 영향을 더 받는 것으로 나타났다. 해당관측소에서 관측되지 않은 날짜를 제외하고, 6,345개 대상병원에 대해 대기환경정보를 적용하여 비교하였을 때, 각각 평균, 최대, 최소값에 대해서 지역별로 서로 차이가 있음을 알 수 있었다. Fig. 15는 기존 지역기반 종관기상정보와 가장 신뢰성이 높은 거리기반 종관+방재기상정보를 36개 지역으로 구분하여 평균, 최대, 최저기온에 대한 기온차를 비교한 결과를 보여준다. 일부 지역은 기존 연구결과에서 영향을 미칠 수 있는 범위보다 크게 나왔다. 결론적으로, 기존 정보와의 비교결과를 통해 기존의 지역기반의 정보를 적용하여 분석하기 보다는 지리적 위치를 고려한 정보를 적용하여 분석을 한다면 오차를 최소화 하고 보다 신뢰성을 바탕으로 의미 있는 결과를 얻을 수 있을 것으로 기대할 수 있다.

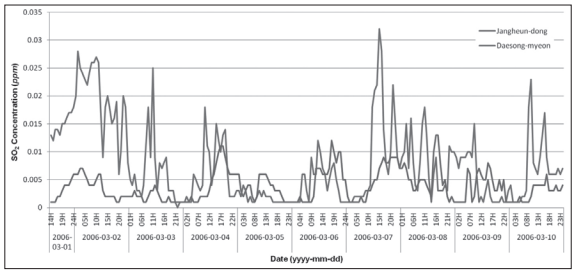


Fig. 11. Concentration difference of SO_2

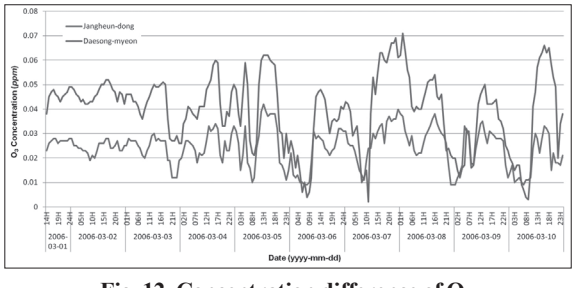


Fig. 12. Concentration difference of O_3

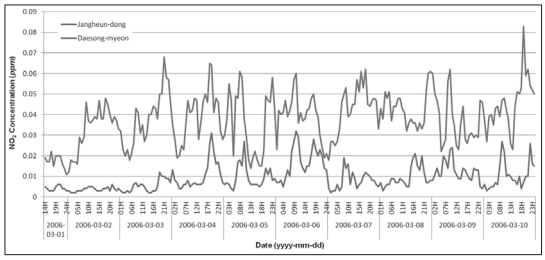


Fig. 13. Concentration difference of NO_2

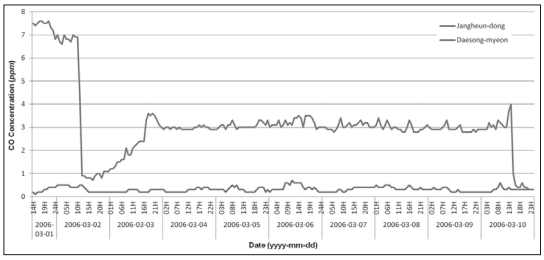
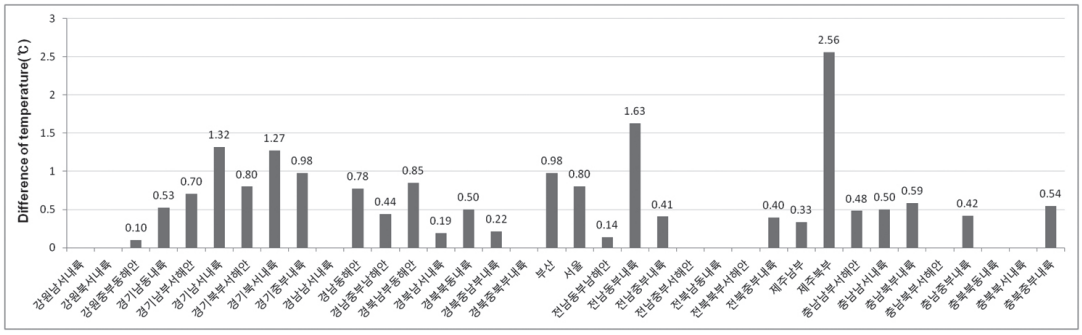
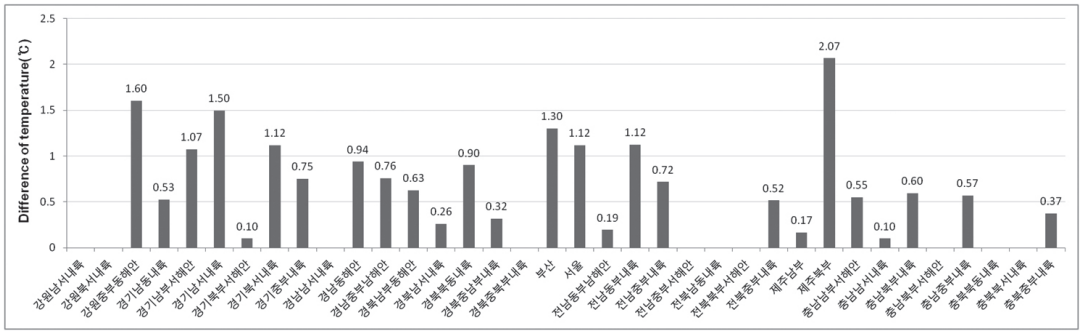


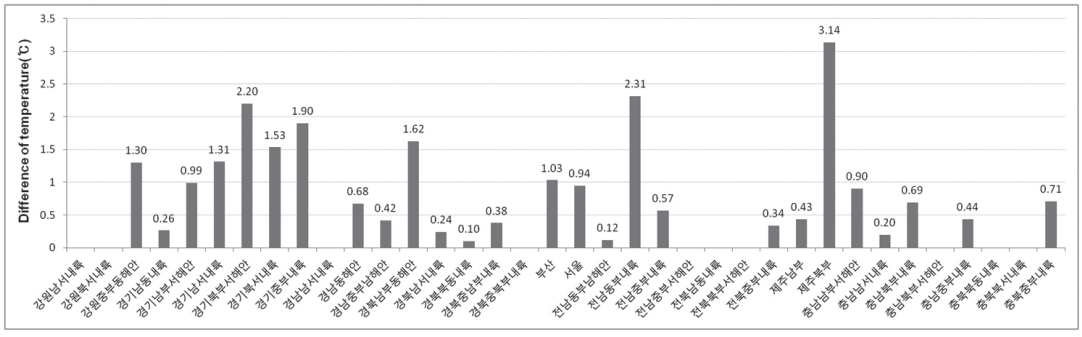
Fig. 14. Concentration difference of CO



(a) Average temperature



(b) Maximum temperature



(c) Minimum temperature

Fig. 15. Regional temperature differences of average, maximum and minimum temperature

6. 결 론

본 연구는 보다 신뢰성 있는 대기환경정보를 수집하기 위한 목적으로 GMap.Net 기반의 대기환경정보 획득 시스템을 개발하였으며, 개발된 시스템을 통해 전국 병원의 지리적 위치를 고려한 거리기반의 인접한 관측소로부터 신뢰성 높은 정보를 수집할 수 있다. 또한, 이렇게 획득 및 생성된 데이터셋은 신뢰성을 검증하기 위해 기존 데이터셋과 거리, 매칭율, 정보오차에 대한 비교평가를 수행하였다. 비교평가 결과에 따르면, 기존에 적용되어진 행정구역 단위로 구분된 지역기반 보다 거리기반일 경우 관측소와의 매칭률 및 평균거리가 효율적인 것으로 나타났으며, 대기환경의 시간별 속성 값의 경우에도 수치적인 차이를 보였다. 이러한 지리적 특성을 고려한 높은 신뢰성의 정보를 적용하여 대기환경정보에 대한 데이터셋을 구축한다면, 국내 심혈관계질환 뿐 아니라 다양한 발병에 영향을 미칠 수 있는 환경적 위험요소에 대한 다각적 연구에 기초자료로써 유용하게 사용되어질 것이다. 이 논문의 한계점은 환자들의 거주지 또는 발병위치에 대한 정보가 아닌 내원한 병원의 위치정보를 활용하였으며, 지리학에서 얻을 수 있는 다양한 지형정보를 제외한 위치좌표만을 사용한 것이다. 따라서, 데이터셋의 높은 신뢰성을 확보하기 위해 특정병원에 내원하기 직전 환자들의 위치정보에 대한 데이터 축적이 요구되며, 다양한 지리학적 정보를 활용한다면 보다 신뢰성 있는 세부적인 분석이 가능할 것이다.

감사의 글

이 논문은 2013년도 미래창조과학부의 재원으로 한국연구재단(No.2013R1A2A2A01068923)과 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업(IITP-2016-H8501-16-1013)의 지원을 받아 수행된 연구임.

References

Amiya, S., Nuruki, N., Tanaka, Y., Tofuku, K., Fukuoka, Y., Sata, N., Kashima, K., and Tsubouchi, H. (2009), Relationship between weather and onset of acute myocardial infarction: can days of frequent onset be predicted?, *Journal of cardiology,* Vol. 54, No. 2, pp. 231-237.

An, S.G. (2006), *Recommendation for Application of Emergency Medical Information Center : Case by Patient Pre-hospital and Inter-hospital Transportation*, Master's thesis, Public Health Yonsei University, Seoul, Korea, 56p. (in Korean with English abstract)

Bang, J.S. (2012), *Study on Analysis of Obstacles to EMT-paramedic's Pre-hospital Paramedic Emergency care for Cardioplegic Patients*, Master's thesis, Dongshin University, Naju, Korea, 84p. (in Korean with English abstract)

Cao, J., Cheng, Y., Zhao, N., Song, W., Jiang, C., Chen, R., and Kan, H. (2009), Diurnal temperature range is a risk factor for coronary heart disease death, *Journal of Epidemiology*, Vol. 19, No. 6, pp. 328-332.

GeoMidpoint (2015), Geographic midpoint calculation methods, *GeoMidpoint*, http://www.geomidpoint.com/calculation.html (last date accessed: 22 October 2015).

GMap.NET (2015), Great maps for windows forms & presentation, *CodePlex*, http://greatmaps.codeplex.com (last date accessed: 13 October 2015).

Goggins, W.B., Chan, E.Y., and Yang, C.Y. (2013), Weather, pollution, and acute myocardial infarction in Hong Kong and Taiwan, *International Journal of Cardiology*, Vol. 168, No. 1, pp. 243-249.

Joo, Y.K. (2014), *Relationship among the Mortality of Cardiovascular, Respiratory and Nation Industrial Complex Air Pollution Using Meta Analysis*, Master's thesis, Hanyang University, Seoul, Korea, 47p. (in Korean with English abstract)

Lee, J.H., Chae, S.C., Yang, D.H., Park, H.S., Cho, Y., Jun, J.E., Park, W.H., Kam, S., Lee, W.K., Kim, Y.J., Kim, K.S., Hur, S.H., and Jeong, M.H. (2010), Influence of weather on daily hospital admissions for acute myocardial infarction (from the korea acute myocardial infarction registry), *International Journal of Cardiology*, Vol. 144, No. 1, pp. 16-21.

Lee, H.Y. and Park, M.Y. (2004), Analysis of the emergency medical service area using GIS: the case of Seoul, *The Journal of GIS Association of Korea*, Vol. 12, No. 2, pp. 193-209. (in Korean with English abstract)

Park, H.J., Woo, K.S., Chung, E.K., Kang, T.S., Kim, G.B., Yu, S.D., and Son, B.S. (2015), A time-series study of ambient air pollution in relation to daily mortality count

in Yeosu, *Journal of Environmental Impact Assessment*, Vol. 24, No. 1, pp. 66-77. (in Korean with English abstract)

Radišauskas, R., Bernotienė, G., Bacevičienė, M., Ustinavičienė, R., Kirvaitienė, J., and Krančiukaitė, D. (2014), Trends of myocardial infarction morbidity and its associations with weather conditions, *Medicina*, Vol. 50, No. 3, pp. 182-189.

Ryu, K.S., Park, H.W., Park, S.H., Ishag, I.M., Bae, J.H., and Ryu, K.H. (2015), The discovery of prognosis factors using association rule mining in acute myocardial infarction with ST-Segment elevation, In: Renda, M.E., Bursa, M., Holzinger, A., and Khuri, S. (eds.), *Information Technology in Bio-and Medical Informatics*, Springer International Publishing, Lecture Notes in Computer Science, pp. 49-55.

Shon, H.S., Hwang, K.K., Bae, J.W., Kim, K.A., Lee, J.Y., and Ryu, K.H. (2013), N-terminal pro-B-type natriuretic peptide as prognostic marker for patients of non ST-segment elevation myocardial infarction, *Journal of Central South University*, Vol. 20, No. 8, pp. 2226-2232.

Sinnott, R.W. (1984), Virtues of the Haversine, *Sky and Telescope*, Vol. 68, No. 2, p. 159.

Vanos, J.K., Hebbern, C., and Cakmak, S. (2014), Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities, *Environmental Pollution*, Vol. 185, pp. 322-332.

WHO reports (2015), The top 10 causes of death, *World Health Organization*, Switzerland, http://who.int/en/ (last date accessed: 16 October 2015).

Wilson, P.W., D'Agostino, R.B., Levy, D., Belanger, A.M., Silbershatz, H., and Kannel, W.B. (1998), Prediction of coronary heart disease using risk factor categories, *Circulation*, Vol. 97, No. 18, pp. 1837-1847.

Yang, B.Y. (2004a), *The Application of GIS for Effective Distribution*, Master's thesis, Kyung Hee University, Seoul, Korea, 98p. (in Korean with English abstract)

Yang, H.E. (2004b), *Generalized Additive Model of Air Pollution to Daily Mortality*, Master's thesis, Duk-Sung Women's University, Seoul, Korea, 60p. (in Korean with English abstract)

Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., Varigus, J., and Lisheng, L. (2004), Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study, *The Lancet*, Vol. 364, No. 9438, pp. 937-952.