

# 미세먼지와 진료과목의 상관관계 분석을 통한 연관성 예측 방법

(Association Prediction Method Using Correlation Analysis between Fine Dust and Medical Subjects)

임명진\*, 김선미\*\*, 신주현\*\*\*

(Myung Jin Lim, Seon Mi Kim, Ju Hyun Shin)

## 요약

미세먼지 등 다양한 원인으로 한국의 대기 오염 문제가 점점 심해지고 있다. 대기 오염 문제가 심해짐에 따라 많은 사람들이 미세먼지에 대한 관심과 건강에 대한 불안이 높아지고 있다. 따라서 미세먼지가 미치는 영향과 어떤 질환과의 상관관계가 있는지에 대한 다양한 연구가 진행되고 있다. 하지만 단순히 미세먼지가 호흡기 및 심혈관 질환, 고혈압 그리고 당뇨병과 같은 특정 질환과의 관계가 있다고 분석하고 있다. 이에 본 논문에서는 진료내역정보 공공데이터를 활용하여 2016년도 월별 진료횟수가 가장 높은 진료과목 10개를 추출하고 피어슨 상관계수를 사용하여 미세먼지가 어떤 진료과목과 관계가 있는지 분석한다. 그리고 더 세분화하여 미세먼지와 성별과 연령대에 따른 진료과목과의 상관관계를 분석한다. 미세먼지와 진료과목이 가장 강한 양의 상관관계인 여성-중년군은 2011년부터 2015년까지의 상관관계를 분석하고 회귀 분석을 통해 연관성 계수를 추출하여 미세먼지 농도에 따른 진료과목과의 연관성 예측 방법을 제안한다.

■ 중심어 : 미세먼지, 진료과목, 상관관계, 연관성 예측, 연관성 계수

## Abstract

Air pollution problems in Korea are gradually becoming a higher concern due to various reasons such as fine dust, causing anxiety among people with regard to their health. Although various studies have been carried out on the relationship between the influence of fine dust and a certain disease, they are mostly focusing on the analyzation that fine dust is related to specific illnesses such as respiratory and cardiovascular diseases, hypertension and diabetes. In this paper, we utilize the public data of medical history information to extract ten medical care subjects with the highest number of monthly care in 2016, and analyze the relation of fine dust with certain medical subjects using Pearson correlation coefficient. We also subdivide and analyze the correlation between fine dust and the medical subjects according to their gender and age. Middle-aged Female group with the strongest positive correlation between fine dust and the medical subjects is analyzed with the correlation from 2011 to 2015, with its relevance coefficient extracted by regression analysis in order to predict the correlation with the medical subjects according to the fine dust concentration.

■ keywords : Fine Dust, Medical Subjects, Correlation, Association Prediction, Correlation coefficient

## I. 서론

대기 오염은 넓은 지역에 피해를 준다는 점에서 시급히 해결해야 하는 환경 문제 중 하나이다[1]. 대표적인 대기 오염 문제 중 하나인 미세먼지(PM10)는 대기 중 장기간 떠다니는 인위적으로 발생하는 지름 10 $\mu$ m 이하의 먼지를 지칭한다[2]. 세계보건기구에서 2013년에 미세먼지를 1급 발암물질로 지정함에 따라

미세먼지에 대한 사람들의 관심이 급격하게 증가하고 있다[3,4]. 미세먼지는 건강에 치명적인 위협을 줄 뿐만 아니라 국민 생활과 경제활동 전반에 부정적인 영향을 미치고 있다[5]. 미세먼지에 장시간 노출될 경우 호흡기 및 심혈관 질환이 발생할 수 있고, 심하면 사망에 다다를 수 있다[6]. 이렇게 대기 오염은 관련 질환의 발생에 영향을 주고, 질병부담 증가와 같은 부정적인 현상을 야기 하고 있다. 또한 국민들의 삶의 질과 생활패턴의 변화를 불러오게 된다.

\* 학생회원, 조선대학교 컴퓨터공학과

\*\* 학생회원, 조선대학교 소프트웨어융합공학과

\*\*\* 정회원, 조선대학교 ICT융합학부

본 연구는 미래창조과학부 및 한국인터넷진흥원의 "고용계약형 정보보호 석사과정 지원사업"의 연구결과로 수행되었음 (과제번호 H2101-16-1001)

접수일자 : 2018년 09월 04일

수정일자 : 2018년 09월 22일

게재확정일 : 2018년 09월 27일

교신저자 : 신주현 e-mail : jhshinkr@chosun.ac.kr

미세먼지로 인한 국민의 불편과 불안은 점차 높아지고 있고, 미세먼지 농도가 높은 날은 외출을 자제하거나 마스크를 착용하는 등 우리 삶에 직접적인 영향을 미치고 있다. 하지만 미세먼지를 포함한 대기 오염에 관한 데이터의 축적, 활용과 이를 이용한 대응방안은 제대로 이루어지지 못하고 있다[2]. 따라서 본 논문에서는 진료내역정보 공공데이터를 활용하여 2016년도 월별 진료횟수가 가장 높은 진료과목 10개를 분류하고 미세먼지가 어떤 진료과목과 관계가 있는지 분석하기 위해 피어슨 상관관계수를 사용했다. 그리고 미세먼지로 인해 노인과 여성의 사망이 증가되었다는 관련연구를 바탕으로 성별과 연령대에 따른 상관관계를 분석한 결과 여성-중년군의 진료과목이 미세먼지와 가장 강한 양의 상관관계인 것을 알 수 있었다. 따라서 2011년부터 2015년까지 미세먼지와 여성-중년군의 진료과목과의 상관관계를 분석하고 회귀 분석을 통해 연관성 계수를 추출하여 미세먼지 농도에 따른 진료과목과의 연관성을 예측한다.

본 논문의 구성은 다음과 같다. 2장에서는 미세먼지의 영향에 관련된 연구들과 미세먼지와 질병과의 상관관계 분석에 관한 기존 연구들을 기술하며, 3장에서는 미세먼지와 진료과목의 상관관계 분석 방법을 기술한다. 4장에서는 성별과 연령에 따라 상관관계를 분석하고 회귀 분석을 통해 연관성 계수로 본 논문에서 제안하는 미세먼지와 진료과목의 연관성 예측 방법을 기술한다. 마지막으로 5장에서는 결론 및 향후 연구에 대해 기술하며 마무리한다.

## II. 관련연구

### 1. 미세먼지의 영향

김형근 외(2015) 연구에서는 실내 환경에 따른 미세먼지 흡입량 평가를 위해 미세먼지가 발생하는 실내 공간에서 실제 인체에 침착되는 미세먼지의 양을 분석하기 위하여 호흡기 침착 모델을 도입하였다. 그리고 호흡기 침착 모델을 적용하여 미세먼지가 많이 발생하는 가스레인지 조리 시 분석을 통해 인체 미세먼지 흡입량을 평가하였다. 미세먼지 흡입량 평가를 통하여 입자별 흡입량과 상세한 미세먼지 분포를 확인할 수 있었다[7].

조용민(2014) 연구에서는 미세먼지 노출에 의한 건강영향에 대해 분석하였다. 미세먼지에 의한 건강영향은 호흡기질환(폐암, 급성하기도 폐질환, 만성 폐쇄성 질환)과 심혈관계 질환(뇌졸중, 허혈성 심질환)으로 크게 구별되고 건강영향의 정도는 먼지의 크기(지름), 하루 중 활동양성, 노출대상의 취약성 등 다양한 요인에 따라 달라졌다. 메타분석결과를 통해 미세먼지 노출에서 노인과 여성 사망이 증가하고 심부전 환자들에서 입원 및 사망이 증가했다. 모든 분석적 연구에서 총사망은 미세먼지의

장기간 노출과 직접적인 연관성이 있었으며 사망증가는 심혈관 및 호흡기에 원인이 있었다[8].

### 2. 미세먼지와 질병의 상관관계

많은 연구들이 미세먼지와 질병과의 연관성을 보여준다. Pope(1999)의 연구에서는 미세먼지와 심장 박동수와의 연관성을 분석하였다. 미세먼지가 100 $\mu$ m 증가할 때 분당 심장 박동수가 평균 0.8회 이상 증가하는 것을 보여주었다[9].

손동욱(2016) 연구에서는 성별에 따른 거주 지역에서 대기 중 미세먼지 오염도와 당뇨병 위험도와의 상관관계를 분석하였다. 대표적인 대기오염물질의 하나인 미세먼지와 당뇨병 위험성 증가 사이의 연관성을 분석함으로써 대기오염의 건강상의 유해성에 대한 검증을 시도하였다. 연구 결과 대기중 미세먼지 오염도와 여성의 당뇨병 유병률 증가 사이에 유의미한 상관성이 있음이 확인되었다. 특히 이러한 상관성은 40세 이상 연령의 여성 인구 집단에서 명확히 확인되고 있었다. 만성질환인 당뇨병의 증가 추세 및 우리나라 대기오염 문제의 심각성을 고려할 때 중년 여성들의 미세먼지 노출로 인한 건강상의 피해를 줄일 수 있는 노력이 상당히 필요할 것으로 판단되었다[10].

이환희 외(2015) 연구에서는 반복 측정 분석을 통한 미세먼지와 고혈압 유병률의 상관관계를 분석하였다. 2009년부터 2013년도까지 지역사회 건강조사 자료를 바탕으로 고혈압 유병률과 미세먼지에 대한 분석 결과 양의 상관관계가 있음이 확인되었다. 2009년에 측정된 미세먼지를 5년간의 고혈압 유병률 자료와 분석한 결과 1 $\mu$ m의 미세먼지가 증가할 때마다 고혈압 유병률은 1.003(95%) 증가했다. 또한 시간과 미세먼지와의 교호작용을 확인할 수 있었다. 미세먼지가 장기간 효과가 있다는 결과에서 2년이 지난 시점에서부터 영향을 준다는 구체적인 기간적 영향을 보여주었다[11].

기존 연구 검토 결과, 미세먼지가 인체에 어떤 영향을 미치고, 특정 질환과의 상관관계 분석에 관한 많은 연구가 진행되고 있으나, 대부분의 연구가 특정 질환에만 국한되어있으며, 진료과목과의 상관관계를 나타내는 연구는 없는 것으로 분석되었다.

## III. 제안한 방법

### 1. 개요

본 논문에서는 성별과 연령에 따른 미세먼지와 진료과목의 상관관계 분석을 통해 연관성 예측에 대한 방법을 제안한다. 본 논문에서 사용된 진료내역정보 데이터는 공공데이터 포털 사이트의 국민건강정보 데이터 셋을 활용하였다. 국민건강보험가입

자 중 해당 연도에 요양기관으로부터의 진료내역이 1건 이상인 가입자 100만 명을 무작위로 선별하고, 항목 선정 과정을 거쳐 선정된 해당 가입자의 기본정보와 진료 정보를 추출하여 진료내역정보 데이터 셋을 1차적으로 구성하였다. 1차적으로 구성된 진료내역정보 데이터 셋의 최적화를 위해 구간분포비율을 최대한 유지한 상태에서 데이터 정제 작업을 거쳐 최종 진료내역정보 데이터 셋을 구성하였다. 국가중점 모든 개방데이터 셋은 매년 12월에 2년 전 데이터를 추가하는 방식으로 1년 단위로 데이터를 갱신한다[12]. 그중 가장 최근 자료인 2016년도 진료내역정보를 활용하였다. 진료내역정보는 기준년도, 가입자 일련번호, 진료내역 일련번호, 성별 코드, 연령대 코드, 시도 코드, 요양개시일자, 서식코드, 진료과목 코드, 주상병 코드, 부상병 코드, 요양 일수, 입내원 일수 등으로 구성되어 있다. 표 1은 2016년 진료내역정보의 일부이고 총 1,048,576건을 데이터 셋으로 활용하였다.

미세먼지 데이터는 기상자료개방 포털 사이트에서 파일셋 조회-부유분진측정을 활용하였다. 파일셋은 황사를 관측할 수 있는 부유분진측정기(PM10) 장비로부터 관측된 연단위의 파일이다. 부유분진측정기 관측이란 대기중의 부유하는 공기를 흡입하여 직경이 10 $\mu$ m이하인 황사가 포함된 먼지가 필터에 침착되고, 동위원소 C-14에서 방출되는 베타선을 필터 여지에 쏘아 감쇄된 베타선을 검출기로 측정하여 농도를 산출한다[13]. 2016년 자료 중에서 월별 대기오염도(도시별) 파일을 이용하여 도시별로 월별 미세먼지 농도를 합산한 후 평균을 계산하여 데이터 셋으로 활용하였다.

2016년 진료내역정보 데이터와 미세먼지 데이터를 활용하여 두 가지 데이터를 분류한 후 피어슨 상관계수를 이용하여 상관관계를 분석하고, 더 세분화하여 성별과 연령대에 따라 상관관계를 분석한 후 강한 양의 상관관계가 있는 성별의 연령대는 2011년부터 2015년까지 과거 5년의 데이터와 2016년 데이터를 회귀 분석을 통해 연관성을 예측하는 방법을 제안한다.

표 1. 2016년 진료내역정보

기준년도	가입자 일련번호	진료내역 일련번호	성별 코드	연령대 코드	시도 코드	요양개시 일자	서식 코드	진료과목 코드	주상병 코드	부상병 코드	요양 일수	입내원 일수
2016	1	34153116	2	2	45	20160817	3	14	L209	L0888	1	1
2016	2	46374827	2	8	41	20160202	3	1	J209	J304	1	1
2016	2	31151110	2	8	41	20160407	3	1	J209	J304	1	1
2016	2	50348269	2	8	41	20160104	3	1	A099	K291	1	1
2016	2	15822952	2	8	41	20160130	3	13	J40	J304	1	1
2016	2	42990877	2	8	41	20160416	3	13	J40	J304	1	1
2016	2	40156783	2	8	41	20160820	3	13	J40	J304	1	1
2016	2	30696074	2	8	41	20160204	3	13	J40	J304	1	1

## 2. 분류

본 절에서는 2016년 진료내역정보 데이터에서 진료횟수가 가장 많은 진료과목 10개(TOP10)를 분류하고, 월별 미세먼지 농도 데이터와 통합하는 과정을 기술한다. 진료내역정보에서 진료과목별로 진료횟수를 합산하여 순위를 부여했다. 표 2는 진료과목별 진료횟수와 순위이다. 내과의 진료횟수가 365,330회로 다른 진료과목에 비해 진료횟수가 가장 많은 것을 볼 수 있다.

표 2. 진료과목별 진료횟수와 순위

진료과목코드	진료과목	진료횟수	순위
1	내과	365,330	1
5	정형외과	168,496	2
10	산부인과	27,861	8
11	소아청소년과	66,544	5
12	안과	68,576	4
13	이비인후과	112,305	3
14	피부과	45,855	6
15	비뇨기과	24,888	9
23	가정의학과	28,985	7
24	응급의학과	10,098	10

진료내역정보 데이터에서 진료횟수가 많은 TOP10 진료과목을 분류하고, 진료과목별로 월별 진료횟수를 합산하였다. 그리고 미세먼지 데이터에서 월별 미세먼지농도의 평균을 계산하여 두 데이터를 통합하였다. 표 3은 TOP10 진료과목의 월별 진료횟수와 미세먼지농도를 통합한 데이터의 일부이다. 이 데이터를 이용하여 진료과목과 미세먼지와의 상관관계를 분석한다.

표 3. 진료과목의 월별진료횟수와 미세먼지농도

월	내과	정형외과	산부인과	소아청소년과	안과	이비인후과	피부과	PM10	월	내과	정형외과	산부인과	소아청소년과	안과	이비인후과	피부과	PM10
1	30902	12675	2207	5247	5244	9633	3244	51.35	7	25473	14508	2349	4540	5538	6595	4431	29.45
2	32547	12641	2127	5570	5088	9580	3166	48.31	8	26196	14911	2377	4179	6198	6797	4904	32.9
3	33841	14671	2372	6512	6153	10547	3404	61.41	9	27580	13714	2251	4651	5749	8367	4102	37
4	30613	14259	2252	5940	6078	10401	3531	67.5	10	29794	13962	2375	5464	5535	9456	3769	39.55
5	29633	14461	2281	5568	5919	9262	4008	53.92	11	31844	13648	2337	6531	5541	10723	3633	54.53
6	26964	15056	2283	4596	5682	7260	4074	41.94	12	39943	13990	2650	7746	5851	13684	3589	49

### 3. 상관 분석

본 절에서는 TOP10 진료과목의 월별 진료횟수와 월별 미세먼지 농도와의 상관 분석을 기술한다. 진료과목과 미세먼지농도가 서로 얼마나 밀접하게 선형적 관계를 맺고 있는지를 상관이라고 하며, 상관관계가 어떤 관련성을 가지고 변화하는지 연관성을 분석하는 것을 상관 분석이라고 한다. 두 변수의 연관 정도를 나타내는 상관계수를 파악하려면 우선, 산점도(Scatter Plot)로부터 두 변수 간에 관련성을 그림을 이용하여 상관의 정도와 방향으로 시각적으로 파악한다[14]. 그림 1은 미세먼지농도와 이비인후과의 상관관계를 산점도로 시각화한 것이다. 산점도에 따르면 미세먼지 농도가 높아짐에 따라 이비인후과의 진료횟수도 많아지는 경향을 보인다. 따라서 둘의 관계는 양의 상관관계라고 볼 수 있다. 분석방법으로는 상관 분석으로 가장 많이 사용하는 피어슨 상관계수(Pearson Correlation Coefficient)를 이용하였다. 상관계수는 -1~1 사이의 값을 가지며 0일 경우에는 두 변수 간의 상관관계가 전혀 없다는 것을 뜻한다. 보통 0.3~0.7 사이의 값이면 뚜렷한 양의 상관관계로 간주하고, 0.7~1.0 사이의 값이면 강한 양의 상관관계로 간주한다[15].

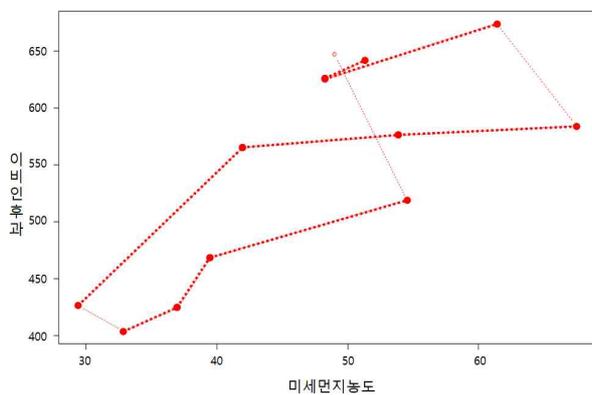


그림 1. 미세먼지농도와 이비인후과의 상관관계

그림 2는 TOP10 진료과목의 진료횟수와 미세먼지농도와의 상관관계를 분석한 결과이다. 1~24는 진료과목 코드이고, PM10은 미세먼지 농도이다. 텍스트는 상관계수를 나타내고 있고, 파란색이 진해질수록 양의 상관관계, 빨간색이 진해질수록 음의 상관관계를 나타내고 있다. 미세먼지농도와 뚜렷한 양의 상관관계인 진료과목은 내과(1), 소아청소년과(11), 이비인후과(13), 가정의학과(23)라는 결과를 볼 수 있다.

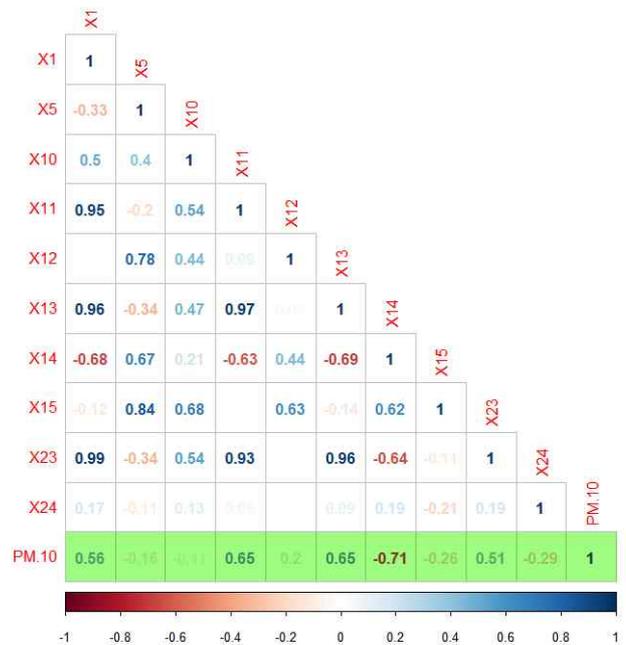


그림 2. 진료과목의 진료횟수와 미세먼지농도의 상관관계

표 4는 미세먼지농도와 뚜렷한 양의 상관관계인 진료과목의 상관계수를 나타내고 있다. 소아청소년과(11)와 이비인후과(13), 내과(1), 가정의학과(23) 순으로 상관관계가 높다고 볼 수 있다[16].

표 4. 미세먼지농도와 진료과목의 상관계수

진료과목코드	진료과목	상관계수
1	내과	0.56
11	소아청소년과	0.65
13	이비인후과	0.65
23	가정의학과	0.51

### IV. 실험 결과 및 고찰

#### 1. 성별과 연령에 따른 상관 분석

본 절에서는 앞에서 제안한 방법을 더 세분화하여 성별과 연령에 따른 TOP10 진료과목의 월별 진료횟수와 월별 미세먼지 농도와의 상관 분석을 기술한다. 미세먼지로 인해 노인과 여성의 사망이 증가되었다는 관련연구를 바탕으로 성별과 연령별로 세분화하여 상관 분석을 하였다. 성별은 남성을 M, 여성을 W 로 표시하였다. 연령은 2016년 UN이 재정립한 평생 연령 기준에 0~17세는 미성년자, 18~65세는 청년, 66세~79세는 중년, 80세~99세는 노인, 100세 이후는 장수노인으로 분류된 것을 참고하였다. 진료내역정보에 0세부터 85세 이상까지 5세단위로 그룹화된 연령대 코드와 UN의 평생 연령 기준과 대조하여 4개의 군 즉, 0~19세는 미성년자(minor), 20~64세는 청년(youth), 65~79세는 중년(middle), 80세 이상은 노인(old)으로 분류하였다.

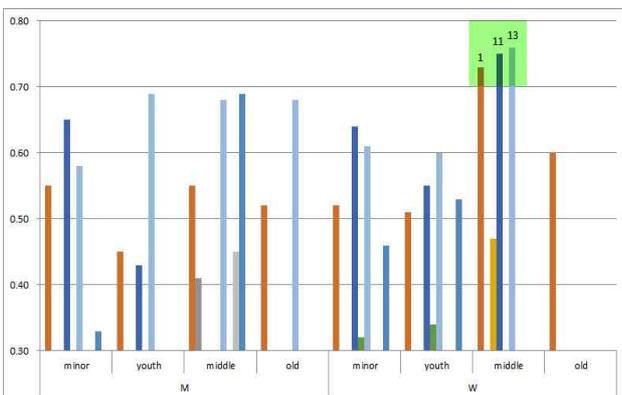
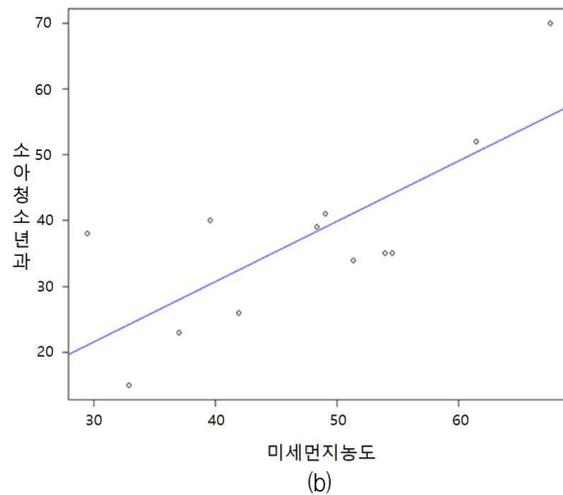
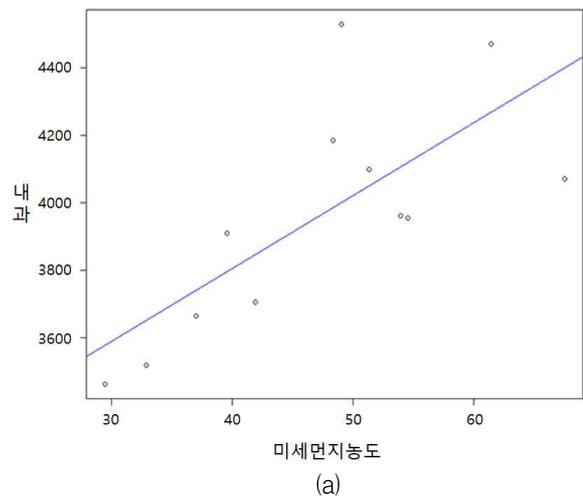


그림 3. 성별과 연령에 따른 상관 분석

그림 3은 성별과 연령에 따른 미세먼지와 진료과목의 상관 분석 결과이다. 미세먼지와 뚜렷한 양의 상관관계가 있는 상관계수 0.3 이상인 진료과목만 차트에 나타냈다. 다른 군에 비해 여성(W)-중년(middle) 군에서 내과(1), 소아청소년과(11), 이비인후과(13)가 미세먼지와 강한 양의 상관관계인 것을 볼 수 있다.

#### 2. 회귀 분석

본 절에서는 다른 군에 비해 미세먼지와 가장 강한 상관관계를 가진 여성(W)-중년(middle) 군을 대상으로 미세먼지와 진료과목과의 회귀 분석을 수행한다. 회귀 분석은 한 개 또는 그 이상의 독립변수들에 대하여 다른 종속변수 사이의 관계를 수학적 모형에 이용하여 설명하고 예측하는 통계적 분석기법이다. 변수 간의 관계에서 다른 변수에 의해 영향을 받아 그 값이 결정되는 변수를 종속변수, 영향을 미치는 변수를 독립변수라고 한다. 여기서 미세먼지는 독립변수로 진료과목의 진료횟수는 종속변수로 설정하였다. 미세먼지농도와 뚜렷한 양의 상관관계인 내과, 소아청소년과, 이비인후과와 회귀 분석을 하였다. 그림 4는 미세먼지 농도와 진료과목의 회귀 분석 결과를 시각화하였고, 파란색 그래프로 회귀 직선을 표현하였다. 위쪽부터 (a)내과, (b)소아청소년과, (c)이비인후과를 나타내고 회귀 직선을 통하여 세 진료과목의 진료횟수는 미세먼지 농도와 비례함을 알 수 있었다.



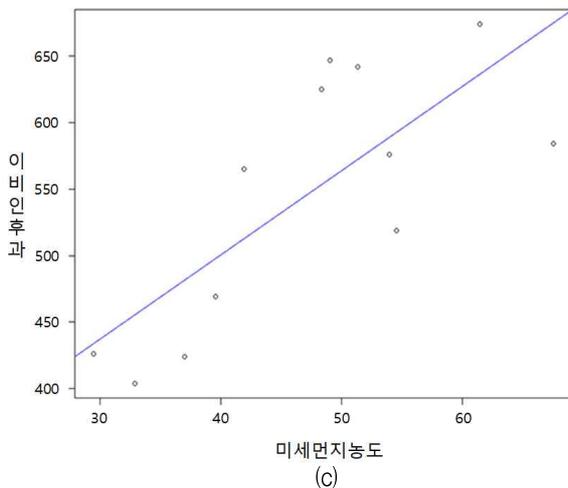


그림 4. 미세먼지 농도와 진료과목의 회귀 분석

표 5는 미세먼지농도와 진료과목의 회귀 분석을 통한 연관성 계수를 나타내고 있다. 연관성 계수는 독립변수가 종속변수를 설명하는 정도로서 회귀모형의 설명력을 표현하는 것이다. 0과 1사이의 값으로 나타내는데, 0에 가까울수록 설명력이 낮고, 1에 가까울수록 높다고 해석할 수 있다[17]. 회귀 분석은 그림 4의 산점도를 가장 잘 설명하는 하나의 선을 구하는 방법인데 그 선이 얼마나 데이터를 잘 설명할 수 있는가에 대한 점수가 연관성 계수인 것이다. 이비인후과의 연관성 계수가 가장 큰 것을 확인할 수 있고, 연관성 계수를 백분율로 환산하여 미세먼지농도와 이비인후과의 진료횟수는 53%의 관련이 있다는 것을 알 수 있다. 연관성 계수를 이용하면 미세먼지농도와 진료과목과의 연관성 예측이 가능하다.

표 5. 미세먼지농도와 진료과목의 연관성 계수

번호	진료과목	연관성 계수
그림 4. (a)	내과	0.4862
그림 4. (b)	소아청소년과	0.5201
그림 4. (c)	이비인후과	0.533

### 3. 연관성 예측

본 절에서는 연관성 예측을 위해 2011년부터 2015년까지의 미세먼지 데이터와 여성(W)-중년(middle) 군의 진료과목을 상관 분석하고 회귀 분석을 통한 연관성 계수를 2016년 데이터와 비교하여 예측 값을 추정한다.

먼저 2011년부터 2015년까지의 진료내역정보 데이터에서 여성(W)-중년(middle) 군의 내과, 소아청소년과, 이비인후과의 진료횟수를 추출하고, 각 연도별 미세먼지농도가 가장 높은 3~5월 데이터와 상관관계를 분석하고 회귀 분석을 진행하였다. 표

6은 2011년부터 2015년까지 과거 5년 데이터의 회귀 분석을 통한 연관성 계수와 2016년도의 연관성 계수를 비교하고 있다. 이비인후과의 2011년부터 2015년까지의 연관성 계수는 0.4758이고, 2016년도의 연관성 계수인 0.533과 비교한 결과 89.21%의 예측 값이 나왔고, 평균적으로 86.61%의 통계적으로 유의한 예측 값이 나오는 것을 볼 수 있었다.

표 6. 과거 5년과 2016년의 연관성 계수 비교

진료과목	연관성 계수		Ratio(%)
	2011~2015년	2016년	
내과	0.4012	0.4862	82.51
소아청소년과	0.4583	0.5201	88.11
이비인후과	0.4758	0.533	89.21
평균			86.61

## V. 결론

본 논문에서는 미세먼지와 진료과목의 상관관계 분석을 통한 연관성 예측을 하였다. 2016년도 진료내역정보에서 진료횟수가 가장 많은 TOP10 진료과목을 분류하고, 월별 진료과목의 진료 횟수와 월별 미세먼지농도의 평균값을 통합하였다. 통합한 데이터로 진료과목별 진료횟수와 미세먼지농도와의 상관관계를 분석하였고, 뚜렷한 양의 상관관계가 있는 내과, 소아청소년과, 이비인후과, 가정의학과에 대하여 성별과 연령으로 세분화하여 상관관계를 분석하였다. 다른 군에 비해 여성-중년 군에서 내과, 소아청소년과, 이비인후과가 미세먼지와 강한 양의 상관관계인 것을 볼 수 있었다. 또한 회귀 분석을 하고 연관성 계수를 이용하여 이비인후과의 2011년부터 2015년까지 과거 5년의 연관성 계수와 2016년 연관성 계수를 비교한 결과 89.21%의 통계적으로 유의한 예측 값이 나왔다. 연관성 계수를 이용하면 미세먼지농도와 진료과목과의 연관성 예측이 가능하다. 본 논문에서는 2011년부터 2016년까지의 자료만 실험하였으나, 2008년부터 미세먼지농도 데이터가 존재하므로 연도별 진료내역정보와 미세먼지농도 데이터를 빅데이터 화하여 연관성을 분석한다면 더욱 정확한 예측이 가능할 것이다. 그리고 예측을 통해 각 진료과목별 성별과 연령대를 구분하여 환자에 따른 미세먼지 맞춤형 예방서비스도 가능해질 것이다.

## REFERENCES

[1] 강태천, 강행봉, "Domain Adaptation 방법을 이용한 기계학습 기반의 미세먼지 농도 예측," *멀티미디어학*

- 회논문지, 제20권, 제8호, 1208-1215쪽, 2017년 8월
- [2] 손대성, 윤영선, 박용덕, 김성현, “빅데이터 활용 기반을 이용한 미세먼지 대응 방안 연구,” *한국지능정보시스템학회 학술대회논문집*, 23-24쪽, 한국과학기술회관, 대한민국, 2017년 11월.
- [3] 서영화, “서울에서 채취한 고농도 미세먼지(PM<sub>10</sub>)에서 황산이온과 유기성 탄소물질과의 상관관계 분석,” *환경관리학회지*, 제20권, 제3호, 71-80쪽, 2014년 9월
- [4] 엄세정, 김보미, 권용재, 윤성범, 권혁준, “미세먼지의 농도와 호흡기질환의 상관관계,” *한국환경교육학회 학술대회 자료집*, 183-186쪽, 서울에너지드림센터, 대한민국, 2017년 11월
- [5] 권혁준, 안은규, 서민영, 김혜곤, “인천시를 기준으로 한 교통량과 미세먼지량의 상관관계 분석,” *한국환경교육학회 학술대회 자료집*, 서울대학교, 대한민국, 164-167쪽, 2016년 6월
- [6] 이석철, 황현숙, 김창수, “쾌적한 지하철 이용을 위한 실시간 환경 모니터링 시스템의 설계 및 구현,” *한국멀티미디어학회 학술발표논문집*, 79-82쪽, 동서대학교, 대한민국, 2006년 5월
- [7] 김형근, 김대연, “실내 환경에 따른 미세먼지 흡입량 평가를 위한 호흡기 침착모델 적용,” *대한건축학회 추계학술발표대회 논문집*, 323-324쪽, 명지대학교, 대한민국, 2015년 10월
- [8] 조용민, 이지호, “미세먼지 노출에 의한 건강영향과 공기청정기의 효율적 사용,” *의료정책포럼*, 제12권, 제2호, 45-54쪽, 2014년 11월
- [9] Pope CA, Dockery DW, Kanner RE, Villegas GM, Schwartz J. “Oxygen saturation, pulse rate, and particulate air pollution : A daily time-series panel study,” *Am J Respir Crit Care Med*, vol. 159, no. 2, pp. 365-372, Feb. 1999.
- [10] 손동욱, “성별에 따른 거주지역 대기 중 미세먼지(PM<sub>10</sub>) 오염도와 당뇨병 유병 위험도와의 상관관계 분석,” *대한 국토·도시계획학회지*, 제51권, 제4호, 211-223쪽, 2016년 8월
- [11] 이환희, 김호, “반복측정 분석을 통한 미세먼지(PM<sub>10</sub>)와 고혈압 유병률의 상관관계 분석: 2009-2013 지역사회 건강조사 자료를 이용하여,” *보건학논집*, 제52권, 제1호, 43-48쪽, 2015년 3월
- [12] 공공데이터포털(2011),  
<https://www.data.go.kr/dataset/15007115/fileData.do>(accessed Jun, 1, 2018).
- [13] 기상자료개방포털(2015),  
<https://data.kma.go.kr/data/climate/selectDustList.do?pgmNo=67>(accessed Jun, 1, 2018).
- [14] 여광수, 김철중, 이재현, 김순석, “상관도를 이용한 국내 의료기관용 개인정보 비식별화 방안 연구,” *스마트미디어저널*, 제5권, 제4호, 83-89쪽, 2016년 12월
- [15] 김선홍, 신호섭, 손석현, “대용량 교통정보와 R을 이용한 교통정보 상관관계 분석,” *정보과학회논문지 : 시스템 및 이론*, 제41권, 제4호, 151-157쪽, 2014년 8월

- [16] 임명진, 김선미, 신주현, “진료내역정보를 활용한 미세먼지와 진료과목의 상관관계 분석,” *멀티미디어학회 학술발표대회논문집*, 제21권, 제1호, 95-98쪽, 2018년 5월
- [17] 주영지, 홍택은, 신주현, “교통사고 데이터의 패턴 분석과 Hybrid Model을 이용한 피해자 상해 심각도 예측,” *스마트미디어저널*, 제5권, 제4호, 75-82쪽, 2016년 12월

## 저자 소개



### 임명진(학생회원)

2000년 군산대학교 컴퓨터과학과 학사 졸업.

2018년 조선대학교 소프트웨어융합공학과 석사 졸업.

2018년~현재 조선대학교 컴퓨터공학과 박사 과정.

<주관심분야 : 빅데이터 처리, 데이터마이닝, 머신러닝>



### 김선미(학생회원)

2016년 조선대학교 컴퓨터공학부 학사 졸업.

2016년~현재 조선대학교 소프트웨어융합공학과 석사 과정.

<주관심분야 : 빅 데이터 처리, 데이터마이닝, 머신러닝>



### 신주현(정회원)

1986년~2011년 (주)청진정보 팀장, (주)투루텍 기술이사

2007년 조선대학교 전자계산학과 이학박사.

2011년~2017년 조선대학교 제어계측로봇공학과 조교수.

2018년~현재 조선대학교 ICT융합학부 부교수.

<주관심분야 : 멀티미디어 데이터베이스, 빅데이터 처리, 데이터마이닝, 머신러닝, 감성정보 처리 등>