

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2021-0081659
(43) 공개일자 2021년07월02일

(51) 국제특허분류(Int. Cl.)

G16H 50/50 (2018.01) G01W 1/02 (2006.01)

G06Q 10/02 (2012.01) G16H 50/30 (2018.01)

(52) CPC특허분류

G16H 50/50 (2018.01)

G01W 1/02 (2013.01)

(21) 출원번호 10-2019-0173838

(22) 출원일자 2019년12월24일

심사청구일자 없음

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

박수진

서울특별시 양천구 목동동로 430 목동신시가지아파트6단지 604동 402호

이승준

서울특별시 강서구 양천로6길 28 도시개발12단지아파트 1204동 203호

(뒷면에 계속)

(74) 대리인

이지연

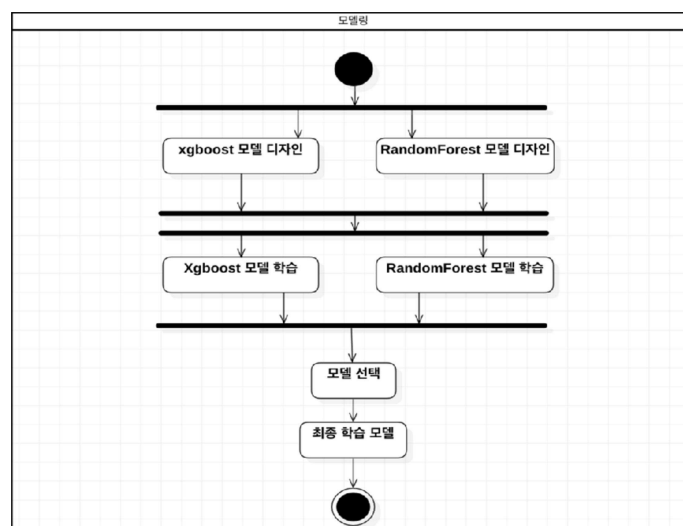
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 날씨 및 미세먼지 데이터를 이용한 질병지수 예측 방법

(57) 요약

본 발명은 머신 러닝 모델을 이용하여 날씨 및 미세먼지 데이터를 활용한 질병지수 예측하는 것을 특징으로 하는 질병 지수 예측 방법에 관한 것이다. 본 발명에 따른 질병 지수 예측 방법은 XGBoost 방법론을 사용해서 기존의 보건기상지수 예측 정확도 향상에 크게 기여하였고, 새로운 질병지수를 추가함으로써 보건기상지수의 다양성을 높였다. 이는 사용자들에게 더 정확한 정보를 제공할 뿐만 아니라 날씨와 관련된 많은 질병에 대해 경각심을 일깨워 미리 건강에 주의할 수 있도록 도움을 줄 것으로 기대된다. 또한, 머신러닝 모델을 사용하는 만큼 지속적인 모델의 업데이트를 통해 추가적인 성능 향상이 있을 것으로 기대된다.

대표도 - 도6



(52) CPC특허분류

G06Q 10/02 (2013.01)

G16H 50/30 (2018.01)

(72) 발명자

이준기

경기도 수원시 영통구 광교마을로 156 광교마을40
단지 4006-1301

김수인

경기도 남양주시 별내면 청학로114번길 34 청학주
공7단지아파트 705동 503호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711093716
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신기술인력양성사업
연구과제명	SW중심대학지원사업
기 여 율	1/1
과제수행기관명	서강대학교산학협력단
연구기간	2019.01.01 ~ 2019.12.31

명세서

청구범위

청구항 1

머신 러닝 모델을 이용하여 날씨 및 미세먼지 데이터를 활용한 질병지수 예측하는 것을 특징으로 하는 질병 지수 예측 방법.

발명의 설명

기술 분야

[0001] 본 발명은 질병 지수 예측 방법 및 시스템에 관한 것으로서, 더욱 구체적으로는 날씨 및 미세먼지 데이터를 활용한 질병지수 예측 방법에 관한 것이다.

배경 기술

[0002] 국내에서는 기상청에서 ‘보건기상지수’ 라는 질병지수를 기상청 사이트에서 제공중이다. 현재 4개의 질병정보 (감기, 천식폐질환, 뇌졸중, 꽃가루농도)를 제공중에 있으며 ‘매우높음’, ‘높음’, ‘주의’, ‘낮음’의 4 단계로 지수화 하여 제공중이다. 기상청의 보건기상지수를 예측하는 모델은 의료계의 자문과 회귀분석을 통해 유의미하다고 판단되는 요인들을 사용한 다중회귀모델을 사용하고 있다. 도 1은 기상청에 의해 제공된 보건기상 지수를 도시한 것이다. 도 1을 참조하면, 질병 지수 등급의 기준은 기상청 확인 결과 특정 질병 환자 수의 누적 백분위 5%, 15%, 60%, 100%를 기준으로 경계값을 형성하였다고 설명하였다.

[0003] 해외의 경우 미국, 호주, 일본등에서 ‘AccuWeather’, ‘Weather Channel’, ‘SunSamrt’ 등에서 관절염, 비염, 편두통, 감기, 꽃가루알러지, UV등을 질병 지수 정보로 제공하고 있으며 보통 기상 예보 사이트의 부가 기능으로써 제공되고 있다. 도 2는 사이트별 질병 지수 제공 현황을 도시한 표이다.

[0004] 먼저, 해외 질병지수 제공 사이트의 경우 국내 맞춤 질병 지수는 제공하지 않아 활용 가능성이 낮다. 뿐만 아니라 기상청의 보건기상지수는 모델과 라벨링 경계값이 모두 공개되어 있는 반면 해외 사이트들의 경우 지수 예측에 사용한 모델이나 등급 경계값들을 공개하지 않아 질병 지수 모델에 대한 개선이나 비교 분석에의 활용이 불가능하다.

[0005] 기상청에서 제공하는 보건기상지수의 경우 현재 시점에서 비추어 볼 때 여러 가지 문제점을 가지고 있다. 우선 기상청 보건기상지수 모델의 경우 회귀분석과 의료계 자문을 이용하여 회차를 선택하였지만 오직 4가지의 회차만을 사용하여 더 많은 회차들로 부터의 정보를 얻을 수 없다. 또한 기상청 보건기상지수 모델은 단순한 회귀분석을 사용하여, 설명력은 높을 수 있으나 등급 경계값을 라벨링하여 측정한 모델 정확도에서 31~39% 정도의 정확도 밖에 나타내지 못한다. 이 뿐만 아니라 회귀분석의 회차들이 회차들간의 교호작용을 전혀 포함하고 있지 않으므로 여러 날씨 정보들의 교호작용을 고려하지 못한다는 단점이 존재한다. 마지막으로 일본의 BioWeather 나 AccuWeather 대비 매우 적은 종류의 질병 지수를 제공한다. 현재 감기, 천식폐질환, 뇌졸중, 꽃가루농도 4개의 보건기상지수만을 제공하기 때문에 사용자들로 하여금 풍부한 정보를 제공하지 못한다는 문제점이 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국등록특허공보 제 10-1828955 호

발명의 내용

해결하려는 과제

[0007] 진술한 문제점을 해결하기 위한 본 발명의 목적은 날씨 및 미세먼지 데이터를 활용한 질병지수 예측 방법을 제공하는 것이다.

과제의 해결 수단

[0008] 진술한 기술적 과제를 달성하기 위한 본 발명의 특징에 따른 질병 지수 예측 방법은 머신 러닝 모델을 이용하여 날씨 및 미세먼지 데이터를 활용한 질병지수 예측하는 것을 특징으로 한다.

발명의 효과

[0009] 본 발명에 따른 질병 지수 예측 방법은 XGBoost 방법론을 사용해서 기존의 보건기상지수 예측 정확도 향상에 크게 기여하였고, 새로운 질병지수를 추가함으로써 보건기상지수의 다양성을 높였다. 이는 사용자들에게 더 정확한 정보를 제공할 뿐만 아니라 날씨와 관련된 많은 질병에 대해 경각심을 일깨워 미리 건강에 주의할 수 있도록 도움을 줄 것으로 기대된다. 또한, 머신러닝 모델을 사용하는 만큼 지속적인 모델의 업데이트를 통해 추가적인 성능 향상이 있을 것으로 기대된다.

도면의 간단한 설명

[0010] 도 1은 기상청에 의해 제공된 보건기상지수를 도시한 것이다.
 도 2는 사이트별 질병 지수 제공 현황을 도시한 표이다.
 도 3은 본 발명에 따른 질병 지수 예측 방법에서 선택된 49개의 휘처를 표시한 도표이다.
 도 4는 본 발명에 따른 질병 지수 예측 방법에 있어서 사용된 주상병 코드를 도시한 도표이다.
 도 5는 기상청 제공 보건기상지수 경계 기준값을 표시한 도표이다.
 도 6은 본 발명에 따른 질병 지수 예측 방법을 구현하기 위한 전반적인 학습 모델 구축 절차를 도시한 것이다.
 도 7은 본 발명에 따른 질병 지수 예측 방법을 이용하여 구축한 웹 페이지에 있어서, 기상 예측 정보 및 질병 지수를 제공하는 화면을 예시적으로 도시한 것이며, 도 8은 질병 지수 예보를 제공하는 화면을 예시적으로 도시한 것이다.
 도 9는 본 발명에 따른 질병 지수 예측 방법에 따라 구현된 질병 지수 예측 모델 시스템을 개괄적으로 도시한 구성도이다.
 도 10은 기존 회귀 분석 모델의 질병 지수 예측 정확도를 도시한 것이며, 도 11은 본 발명에 따른 방법을 구현한 시스템을 이용한 질병 지수 예측 정확도(휘처 49개)를 도시한 것이며, 도 12는 본 발명에 따른 방법을 구현한 시스템을 이용한 질병 지수 예측 정확도(휘처 10개)를 도시한 것이다.
 도 13은 본 발명에 의해 새로 추가한 5개의 질병 지수에 대한 XGBoost 모델을 이용한 본 발명에 따른 방법의 예측 정확도를 도시한 도표이다.

발명을 실시하기 위한 구체적인 내용

[0011] 본 발명에 따른 질병 지수 예측 방법은, 기존의 기상청 제공 보건기상지수 대비하여, 아래와 같이 차별점을 가지고 있다. 첫째, 본 질병지수 예측 모델은 더욱 다양한 휘처를 사용한다. 기존 보건기상지수 예측 모델이 오직 4가지의 휘처를 사용했다면 본 발명에 따른 방법은 미세먼지 데이터를 포함하여 더욱 다양한 기상데이터의 휘처를 사용하기 때문에 더욱 높은 정확도를 보인다. 둘째, 기존 기상청 보건기상지수는 단순한 회귀분석을 사용하였으나 본 발명에 따른 방법은 머신러닝 모델을 사용하여 더욱 높은 정확도를 보이면서도 해석이 용이하다. 마지막으로 본 발명에 따른 방법은 기존에 제공하던 4가지 지수뿐만 아니라 더욱 다양한 질병지수를 제공함으로써 사용자에게 더 풍부한 양의 정보를 제공할 수 있다. 이러한 차별점들을 토대로 본 프로젝트는 사용자들에게 기존 기상청 보건기상지수 보다 질적으로, 양적으로 더욱 향상된 정보를 제공할 수 있을 것으로 기대된다.

[0012] 본 발명에 따른 날씨 및 미세먼지 데이터를 활용한 질병지수 예측 방법은 종래의 기상청 보건기상지수를 기준으로 이를 개선 및 발전시키는 것을 목적으로 접근하였다. 첫째로 본 질병지수 예측 모델은 기존의 기상청 보건기상지수 대비 더 많은 휘처를 사용한다. 기존의 보건기상지수는 최저기온, 일교차, 평균기압, 평균습도 4가지의 휘처만을 사용한 반면, 본 발명에 따른 방법은 기상청 API를 통해 받을 수 있는 53개의 휘처와 미세먼지 데이터에서 받을 수 있는 6개의 휘처 중 지역별로 공통적으로 제공되는 49개의 휘처를 사용한다. 실시간 질병지수의

경우 이 49개의 회차를 모두 사용할 수 있으므로 49개의 회차 중 유의미한 회차를 선택하여 질병지수 예측 모델의 성능을 향상시킬 수 있었다. 도 3은 본 발명에 따른 질병 지수 예측 방법에서 선택된 49개의 회차를 표시한 도표이다.

[0013] 하지만 1주일간의 예상 질병지수의 경우 기상청 날씨 예보 API와 미세먼지 API를 통해 불러올 수 있는 예보데이터가 날씨 데이터 중 최고기온, 최저기온, 평균 상대습도, 평균 풍속, 평균 해면기압 그리고 미세먼지 데이터 중 PM25(초미세먼지)제외한 5개 회차들 뿐이므로 이를 이용해 모델을 구성하였다. 이 10가지의 회차들은 기존의 기상청 보건기상지수의 회차들과 동일하거나 선형 관계를 가지고 있으므로 충분히 유의미한 회차들이며, 이에 따라 유의미한 모델을 만들 수 있을 것이라고 판단하였다.

[0014] 둘째로 더 많은 질병군에 대한 질병지수를 제공할 수 있도록 예측 질병지수 항목을 확대하였다. 질병에 대한 정보는 진료내역정보 데이터에서 주상병코드와 요양 개시일자를 기준으로 환자 수를 추출하였다. 기존의 보건기상지수의 경우 기상청에서 주상병코드를 명시하여 놓아, 해당 질병의 환자 수를 쉽게 추출할 수 있었다. 새로이 추가되는 관절염, 심장질환, 폐렴 등의 새로운 질병지수들은 개발을 위해 관련 주상병코드를 검색하여 각 주상병 코드에 해당하는 하위 주상병코드를 모두 정리 후 이에 맞는 환자 수를 추출하여 모델링에 데이터로써 활용하였다. 이에 따른 주상병코드는 도 4에 도시된 도표와 같다.

[0015] 도 4는 본 발명에 따른 질병 지수 예측 방법에 있어서 사용된 주상병 코드를 도시한 도표이다.

[0016] 마지막으로 본 발명에 따른 질병 지수 예측 방법에서는 기존의 회귀분석이 아닌 머신러닝 모델(XGBoost)을 활용함으로써 모델의 정확도를 기존 기상청 보건기상지수 회귀분석 모델 대비 크게 향상 시켰다. 우선 기상청 보건기상지수 회귀모델과 새로운 질병지수 모델에 대한 정확한 평가를 위해 라벨링 전략을 구성하였다. 이를 위해 기상청에 직접 보건기상지수 4등급 기준을 확인한 결과 상이하였듯 누적백분위 5%, 15%, 60%, 100% 기준으로 경계 값을 구성하여 등급을 구분하였다. 예를 들어 특정 질병의 환자 수를 예측하였을 때 해당 수치가 해당 연도 환자 수 분포의 상위 5%에 해당하는 수치일 경우 매우 높음에 해당하는 1등급을 부여하는 방식이다.

[0017] 도 5는 기상청 제공 보건기상지수 경계 기준값을 표시한 도표이다.

[0018] 본 발명에 따른 질병 지수 예측 방법에서는 이러한 라벨링 전략에 따라 해당 질병의 연도별 발생 환자수를 누적 백분위로 나누어 4개의 등급 값을 부여하였다. 하지만 특정 질병의 경우 한 해에 발생하는(기록된) 환자 수가 매우 적어 등급 값을 정확히 4개로 구분할 수 없다. 하지만 이러한 질병들에 대해서도 똑같은 경계 값을 부여하여 라벨링 전략에 일관성을 부여하였으며 대신 등급 경계값을 4단계가 아닌 2~3단계 정도로 자동으로 축소시켜 '위험', '보통' 혹은 '매우위험', '주의', '보통' 과 같이 예측하도록 하였다.

[0019] 이러한 라벨링 전략을 토대로 모델 학습을 진행하였다. 학습에 사용한 머신러닝 모델은 XGBoost이다. 단순 분류 모형에 있어서는 복잡한 신경망 모형(딥러닝) 보다 트리기반의 머신러닝 모델들이 우수한 성능을 보이는 경우가 많으며, 모델 최적화가 간단하고 해석이 용이하므로 딥러닝 모델이 아닌 머신러닝 모델을 채택하여 사용하였다. 또한 기존의 모델이 단순 회귀분석 모델인 것을 감안하였을 때, 머신러닝 모델의 사용으로도 충분한 정확도의 향상을 가져올 수 있을 것으로 판단하여 XGBoost 모델을 사용하였다. 실제 모델링과 비교 분석에서는 또 다른 트리기반의 머신러닝인 RandomForest 역시 함께 구현하여 비교하였으나 XGBoost의 경우가 성능, 사용, 속도 면에서 더 나은 기록을 보여주었기 때문에 최종 프로젝트의 모델로 XGBoost 모델을 선택하였으며, 이후 이를 이용해 하이퍼 파라미터의 튜닝 및 모델 최적화를 진행하였다. 이러한 전반적인 학습모델 구축 절차는 도 6의 액티비티 다이어그램과 같다. 도 6은 본 발명에 따른 질병 지수 예측 방법을 구현하기 위한 전반적인 학습 모델 구축 절차를 도시한 것이다.

[0020] 본 발명에 따른 질병 지수 예측 방법은, 도 6에 도시된 방식으로 학습모델을 구축하여 기존 15년치 데이터를 학습시켜 피팅된 예측 모델을 만든 후, 해당 모델의 가중치를 저장하여 '학습된 예측 모델(Pre-trained Prediction Model)'을 생성한다. 이후 학습모델은 1년 혹은 2년 주기로 데이터가 쌓일 때 마다 다시금 새로운 데이터를 학습시킴으로써 계절성 데이터나 특이 기후에 치우치지 않게 지속적으로 모델을 업데이트 시켜 줄 수 있다.

[0021] 이렇게 예측된 질병지수 모델은 본 발명에 의해 구성된 웹사이트를 통해 제공된다. 웹 페이지는 크게 사용자에게 지역코드와 보건기상지수 선호도를 반영한 실시간 예보 데이터와 개별 지수 목록을 제공하는 메인페이지, 개별 질병지수에 대한 예보를 그래프로 제공하는 페이지, 회원관리를 위한 로그인 및 인포메이션 페이지 세 파트로 구성되었다.

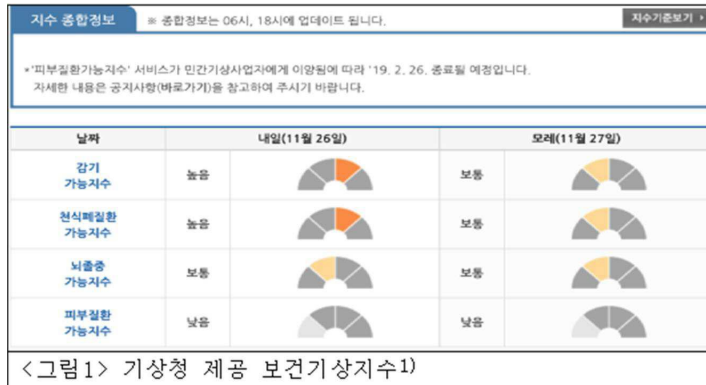
[0022] 도 7은 본 발명에 따른 질병 지수 예측 방법을 이용하여 구축한 웹 페이지에 있어서, 기상 예측 정보 및 질병

지수를 제공하는 화면을 예시적으로 도시한 것이며, 도 8은 질병 지수 예보를 제공하는 화면을 예시적으로 도시한 것이다.

- [0023] 도 7 및 도 9에서 볼 수 있다시피, 메인페이지에서는 실시간 기상데이터를 활용한 날씨정보와 더불어 각 질병지수별 심각도를 픽토그램의 색상 차이로 확인할 수 있다. 사용자가 추가적인 예보 정보를 얻고자 한다면 개별지수 페이지로 접근하여 주간 질병지수의 예측 변화를 확인할 수 있으며 이에 대한 대처 요령을 제공한다. 이러한 모든 정보는 사용자가 가입시 입력한 정보에 따라 사용자가 관심있는 8가지 질병정보와 6가지 지역 정보에 맞추어 정보를 제공할 수 있도록 하였다. 이러한 전체적인 「날씨 및 미세먼지 데이터를 활용한 질병지수 XGBoost 모델」 시스템 콘셉트 도식은 도 9와 같다.
- [0024] 도 9는 본 발명에 따른 질병 지수 예측 방법에 따라 구현된 질병 지수 예측 모델 시스템을 개괄적으로 도시한 구성도이다.
- [0025] 이하, 본 발명에 따른 방법에 의해 구현된 시스템의 성능을 평가한다. 우선 기상청에서 기존에 제공되고 있는 세 가지 보건 기상 지수에 대하여 기존의 모델과 새로운 모델의 예측 정확도를 비교 평가한 표는 도 10, 도 11, 도 12와 같다. (* : 환자수 부족으로 인해 라벨이 3개로 줄어든 경우. ** : 라벨이 2개로 줄어든 경우.)
- [0026] 도 10은 기존 회귀 분석 모델의 질병 지수 예측 정확도를 도시한 것이며, 도 11은 본 발명에 따른 방법을 구현한 시스템을 이용한 질병 지수 예측 정확도(회처 49개)를 도시한 것이며, 도 12는 본 발명에 따른 방법을 구현한 시스템을 이용한 질병 지수 예측 정확도(회처 10개)를 도시한 것이다.
- [0027] 라벨의 개수가 바뀌는 경우를 제외한 질병지수 예측 정확도 14건을 기준으로 비교해보면, 기존의 회귀 분석 모델이 평균 39.48%의 정확도를 보인 것에 비해 XGBoost를 이용한 새로운 질병지수 예측 모델은 49개의 회처를 이용한 경우와 10개의 회처를 이용한 경우 각각 57.24%, 56.15%의 정확도를 보여 약 1.4배의 정확도 상승을 보였다. 라벨의 개수가 3개 혹은 2개로 줄어든 경우는 70%에서 90%에 달하는 정확도를 보이기도 했다.
- [0028] 더욱이 XGBoost를 이용한 두 모델의 성능 비교를 위해 도 10과 도 11을 살펴보면 49개의 회처를 모두 사용할 경우와 예보데이터로 얻을 수 있는 10개의 회처만을 이용할 경우의 정확도 차이가 근소할 뿐만 아니라, 회처가 많다고 정확도가 항상 높지도 않음을 알 수 있었다. 따라서 모든 회처를 사용한 실시간 모델을 따로 만들 필요 없이, 실시간 질병 지수 역시 10개의 회처를 이용한 모델로 충분히 예측 가능 판단하여 두 개의 모델 중 회처 10개를 이용한 모델만을 사용하기로 했다.
- [0029] 최종적으로 선택된 10개의 회처만을 이용하는 XGBoost 모델로 기존에 기상청에서 제공하지 않는 5개의 질병(심근경색, 관절염, 동맥경화, 폐렴, 부정맥)에 대한 질병지수 예측 모델을 만들었고 그 결과는 도 13과 같다. 도 13은 본 발명에 의해 새로 추가한 5개의 질병 지수에 대한 XGBoost 모델을 이용한 본 발명에 따른 방법의 예측 정확도를 도시한 도표이다.
- [0030] 본 발명에 따른 질병 지수 예측 방법은 XGBoost 방법론을 사용해서 기존의 보건기상지수 예측 정확도 향상에 크게 기여하였고, 새로운 질병지수를 추가함으로써 보건기상지수의 다양성을 높였다. 이는 사용자들에게 더 정확한 정보를 제공할 뿐만 아니라 날씨와 관련된 많은 질병에 대해 경각심을 일깨워 미리 건강에 주의할 수 있도록 도움을 줄 것으로 기대된다. 또한, 머신러닝 모델을 사용하는 만큼 지속적인 모델의 업데이트를 통해 추가적인 성능 향상이 있을 것으로 기대된다.
- [0031] 이상에서 본 발명에 대하여 그 바람직한 실시예를 중심으로 설명하였으나, 이는 단지 예시일 뿐 본 발명을 한정하는 것이 아니며, 본 발명이 속하는 분야의 통상의 지식을 가진 자라면 본 발명의 본질적인 특성을 벗어나지 않는 범위에서 이상에 예시되지 않은 여러 가지의 변형과 응용이 가능함을 알 수 있을 것이다. 그리고, 이러한 변형과 응용에 관계된 차이점들은 첨부된 청구 범위에서 규정하는 본 발명의 범위에 포함되는 것으로 해석되어야 할 것이다.

도면

도면1



도면2

국가	사이트	인플루엔자	알러지	두통	관절염	UV	기타
미국	AccuWeather	✓	✓	✓	✓		호흡기건강, 부비강두통
미국	WeatherChannel		✓				-
호주	SunSmart					✓	-
일본	BioWeather⁵⁾	✓		✓	✓	✓	뇌출혈, 뇌경색, 천식, 열사병, 우울증, 수족구병, 건성피부, 헤루판기나

도면3

항목명	항목설명	항목명	항목설명	항목명	항목설명
AVG_RHM	평균 상대습도	AVG_PA	평균 현지기압	AVG_TD	평균 이슬점온도
MAX_INS_WS_WD	최대 순간풍속 풍향	SUM_FOG_DUR	안개계속시간	HR1_MAX_RN	1시간 최다 강수량
MIN_TG	최소 초상온도	SUM_DPTH_FH_SC	합계 3시간 신적설	MIN_PS	최저 해면기압
SUM_LSS_HR	합계 일조시간	AVG_TS	평균 지면온도	MAX_WS	최대 풍속
MAX_TA	최고 기온	MAX_PS	최고 해면기압	MIN_PS_HRMT	최저 해면기압 시각
AVG_TA	평균 기온	MIN_RHM_HRM_T	평균 상대습도 시각	MAX_TA_HRMT	최고 기온시각
DD_MES	일 최심적설	MIN_TA_HRMT	최저 기온시각	AVG_TCA	평균 전운량
SS_DUR	가조시간	AVG_LMAC	평균 중하층운량	HR24_SUM_RW_S	풍정합
DD_MES_HRMT	일 최심적설시각	HR1_MAX_RN_HRMT	1시간 최다 강수량시각	MAX_WD	최대 풍향
MAX_WS_HRMT	최대 풍속시각	MAX_INT_WS_HRMT	최대 순간풍속 시각	DD_MEFS_HRM_T	일 최심심 적설 시각
AVG_WS	평균 풍속	SUM_RN_DUR	강수계속시간	PM10	미세먼지
N9_9_RN	9-9 강수	MI10_MAX_RN	10분 최다 강수량	O3	오존
SUM_RN	일 강수량	MAX_INS_WS	최대 순간풍속	NO2	이산화질소
MAX_WS_WD	최대 풍속풍향	AVG_PS	평균 해면기압	SO2	아황산가스
MAX_PS_HRMT	최고 해면기압 시각	AVG_PV	평균 중기압	CO	일산화탄소
MIN_RHM	최저 상대습도	DD_MEFS	일 최심심적설		
MIN_TA	최저 기온	MI10_MAX_RN_HRMT	10분 최다 강수량시각		

도면4

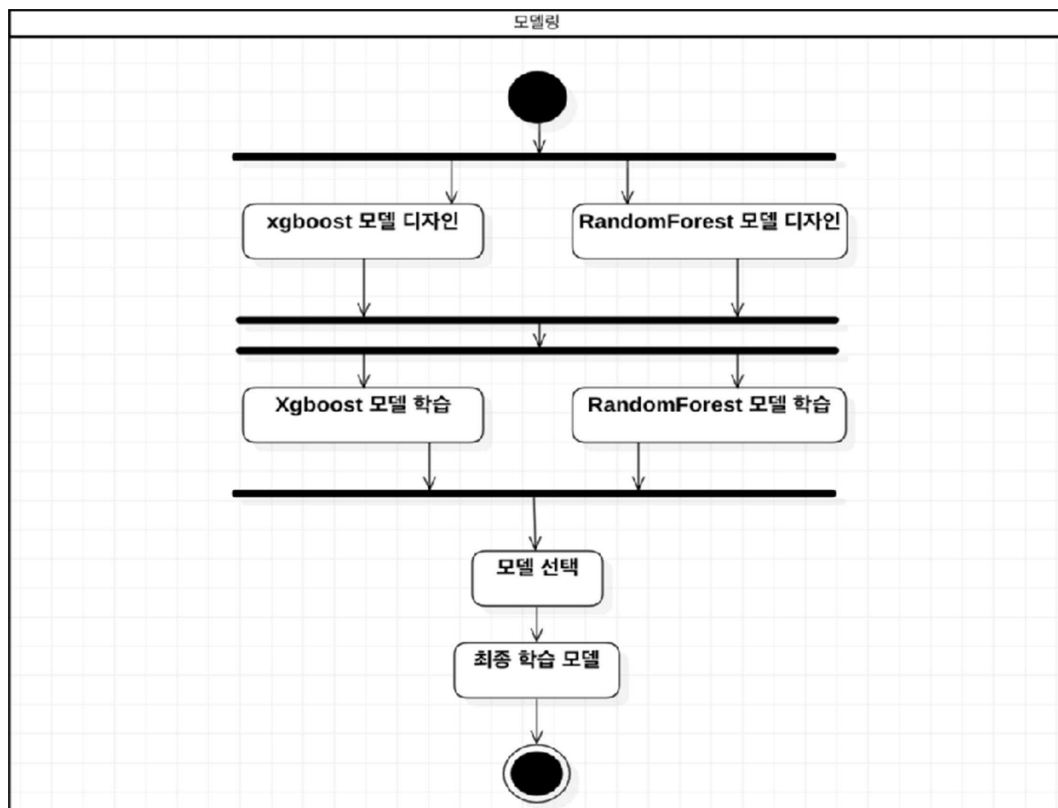
질병군	질병 표준 주상병코드
부정맥	I47, I48, I49(I49.8, I49.9)
폐렴	J12 ~ J18, J67 ~ J68
동맥경화	I70
관절염	M00 ~ M19(M10제외), M45
심근경색	I21 ~ I24
천식, 폐질환	J44, J45, J46
뇌졸중	I60, I61, I62, I63
감기	J00, J03, J06, J20

도면5

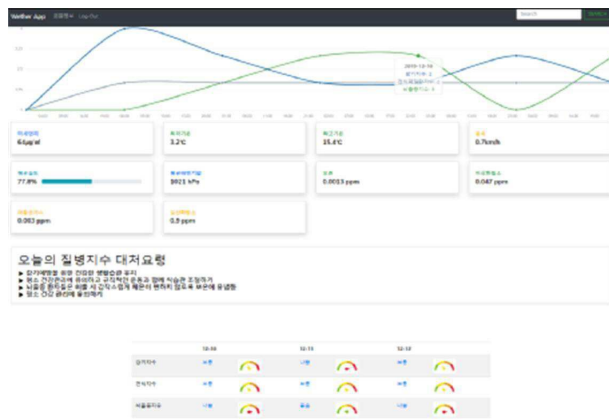
< 천식·폐질환가능지수 산출을 위한 지수값 분류기준 >			
Level	등급	백분율	천식·폐질환가능지수(ALI)
1등급	매우 높음	0-5	$ALS \geq 3.0525$
2등급	높음	6-15	$2.6452 \leq ALS < 3.0525$
3등급	보통	16-60	$1.5354 \leq ALS < 2.6452$
4등급	낮음	61-100	$1 \leq ALS < 1.5354$

<그림2> 기상청 제공 보건기상지수 경계 기준 값

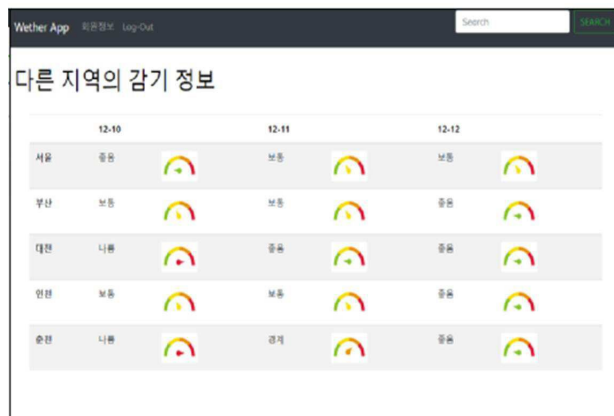
도면6



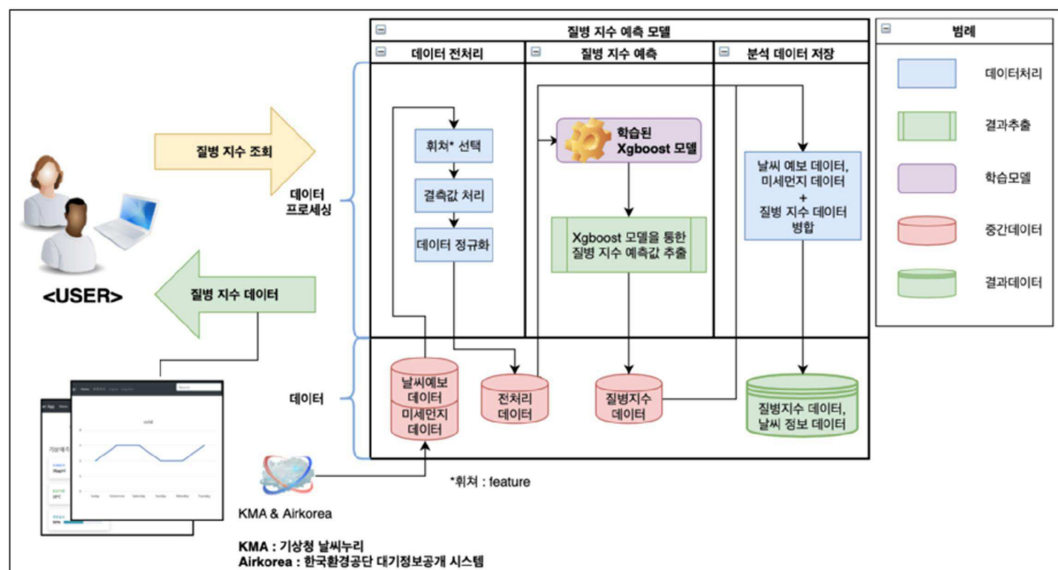
도면7



도면8



도면9



도면10

모델	지수	정확도 (%)					
		서울	부산	인천	광주	대전	울산
기존 (회차4개)	천식	38.69	39.62	39.16	34.78	29.90	30.75
	폐질환						
	뇌졸중	34.91	27.98	33.38	27.16	26.73	18.40
	감기	45.51	49.42	47.23	44.44	43.78	41.18

도면11

모델	지수	정확도 (%)					
		서울	부산	인천	광주	대전	울산
XGBoost (회차49개)	천식	53.76	61.42	53.08	51.44	55.95	58.82
	폐질환						
	뇌졸중	54.58	73.87*	59.51	80.85**	77.56*	89.47**
	감기	55.81	62.79	61.56	57.18	58.69	56.77

도면12

모델	지수	정확도 (%)					
		서울	부산	인천	광주	대전	울산
XGBoost (회차10개)	천식	55.27	57.32	52.12	54.45	56.77	58.28
	폐질환						
	뇌졸중	51.03	73.19*	58.82	83.58**	76.33*	90.56**
	감기	53.76	60.88	60.60	54.45	56.36	55.95

도면13

모델	지수	정확도 (%)					
		서울	부산	인천	광주	대전	울산
XGBoost (회차10개)	심근경색	87.14**	94.53**	92.34**	93.02**	96.85**	98.91**
	관절염	77.02	73.05	76.06	80.30	77.84	82.22*
	동맥경화	89.19**	95.90**	93.84**	99.32**	98.91**	99.59**
	폐렴	63.75	79.21*	75.10*	75.10*	82.90**	87.82**
	부정맥	57.60	54.31	65.12*	79.48*	76.47*	89.74**