

ORIGINAL ARTICLE

## DNN을 활용한 부산지역 초미세먼지 예보방안

도우곤\* · 김동영 · 송희진 · 조갑제

부산광역시 보건환경연구원

### A Study on the PM<sub>2.5</sub> forecasting Method in Busan Using Deep Neural Network

Woo-Gon Do\*, Dong-Young Kim, Hee-Jin Song, Gab-Je Cho

Busan Metropolitan City Institute of Health and Environment, Busan 46616, Korea

#### Abstract

The purpose of this study is to improve the daily prediction results of PM<sub>2.5</sub> from the air quality diagnosis and evaluation system operated by the Busan Institute of Health and Environment in real time. The air quality diagnosis and evaluation system is based on the photochemical numerical model, CMAQ (Community multiscale air quality modeling system), and includes a 3-day forecast at the end of the model's calculation. The photochemical numerical model basically has limitations because of the uncertainty of input data and simplification of physical and chemical processes. To overcome these limitations, this study applied DNN (Deep Neural Network), a deep learning technique, to the results of the numerical model. As a result of applying DNN, the *r* of the model was significantly improved. The *r* value for GFS (Global forecast system) and UM (Unified model) increased from 0.77 to 0.87 and 0.70 to 0.83, respectively. The RMSE (Root mean square error), which indicates the model's error rate, was also significantly improved (GFS: 5.01 to 6.52 ug/m<sup>3</sup>, UM: 5.76 to 7.44 ug/m<sup>3</sup>). The prediction results for each concentration grade performed in the field also improved significantly (GFS: 74.4 to 80.1%, UM: 70.0 to 77.9%). In particular, it was confirmed that the improvement effect at the high concentration grade was excellent.

**Key words** : CMAQ, Deep Neural Network, PM<sub>2.5</sub> forecast

#### 1. 서론

세계적으로 급격한 산업화와 도시화가 진행됨에 따라 초미세먼지는(PM<sub>2.5</sub>, particulate matter with an aerodynamic diameter ≤ 2.5 μm) 도시지역에서의 심각한 환경 문제의 원인으로 인식되어 왔다. 초미세먼지는 대중의 건강에 나쁜 영향을 미치며 특히 심혈관이 나 호흡기계 질환의 유발을 또는 이와 관련한 사망자수를 증가시키는 것으로 알려져 있다(Ostro et al., 2007; Pope et al., 2015). 따라서 초미세먼지의 시공간적인

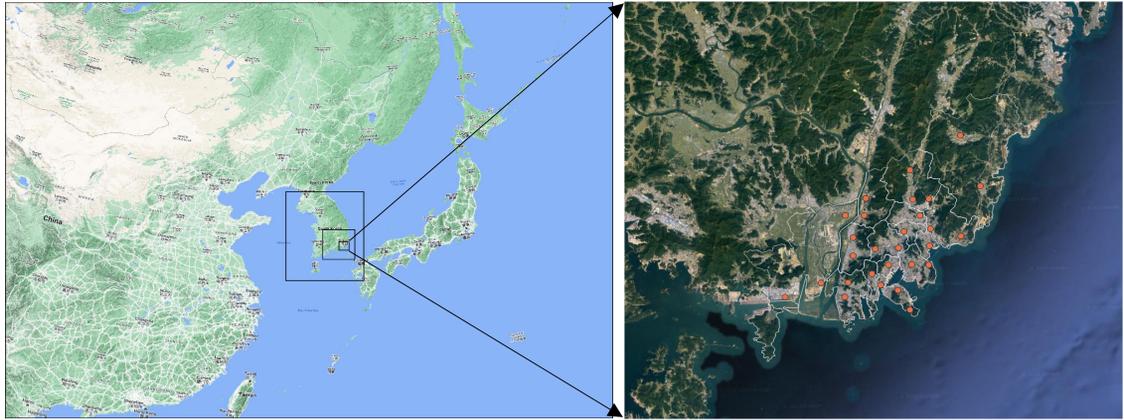
분포와 인위적인 배출원의 영향 또는 기상학적인 원인을 규명하고 농도 예측을 위한 다양한 시도가 이루어져 왔다. 기상인자를 입력하여 전통적인 다중선형회귀(Multiple Linear Regression, MLR)로 시간대별 그리고 일별 미세먼지 농도를 예측하기 위한 노력과 함께 기상인자와 화학반응 또는 기상인자와 위성기반 AOD (Aerosol Optical Depth) 자료 등을 활용한 다양한 연구가 수행되었다(Slini et al., 2006; Diaz-Robles et al., 2008; UI-Saufie et al., 2011; Abdullah et al., 2020). 최근에는 전산기술의 발달로 광화학 수치모델

Received 16 August, 2023; Revised 22 August, 2023;

Accepted 23 August, 2023

\*Corresponding author : Woo-Gon Do, Busan Metropolitan City Institute of Health and Environment, Busan 46616, Korea  
Phone : +82-51-309-2928  
E-mail : dou777@korea.kr

© The Korean Environmental Sciences Society. All rights reserved.  
© 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.



**Fig. 1.** Modeling domain for air quality diagnosis and evaluation system. Left part shows locations of domain1 to domain3, right part shows geographical features of domain4 and locations of air quality monitoring stations.

의 결과를 실시간으로 공개하거나 이를 활용하여 대기 오염도를 사전에 예측하는 경우가 많아지고 있다. 국내의 경우 초미세먼지는 2015년부터 대기환경기준 항목으로 채택되면서 전국의 대기오염측정소에서 상시로 측정되기 시작하였다. 또한 국립환경과학원의 대기질 통합예보센터에서는 (초)미세먼지 및 오존의 농도 등급예보를 2014년부터 시행해오고 있으며 오염물질의 농도 등급과 더불어 한반도 전체 권역에 대한 CMAQ 모델링 결과를 시각화하여 에어코리아를 통하여 공개하고 있다(Korea Environment Corporation, 2022). 한편 지자체 중 최초로 부산광역시에서는 CMAQ 모델을 기반으로 하는 대기질 진단평가시스템을 구축하여 2017년부터 지역내 대기오염물질 농도를 예측하는데 활용하고 있다. 하지만 광화학 수치모델의 결과를 직접적으로 대기오염 예측자료로 사용하기에는 실측되는 대기오염물질의 농도와 차이가 발생하기 때문에 다소 무리가 있으며 이는 배출량 입력자료의 불확실성, 초기 기상장과 기상 모사에 따른 불확실성, 모델에서 사용되는 대기오염물질 농도 초기장의 실제와의 차이 그리고 모델에서 사용되는 물리, 화학과정의 단순화 등 기본적인 문제에서 기인한다(Choi and Koo, 2013). 따라서 이러한 광화학 수치모델의 기본적인 약점을 보완하기 위하여 입력자료 및 초기조건의 현실화, 모델 내 물리, 화학과정의 개선 등 다양한 시도가 지속적으로 이루어져 오고 있다(Choi and Koo, 2013; Kim and Jang, 2014; Jo et al., 2017; Kitayama et al., 2019). 이와 더불어 최근에는 대기오염물질의 시, 공간적인 예측

에 머신러닝 기법을 활용하는 사례가 늘고 있다(Rybarczyk and Zalakeviciute, 2018). 대기오염물질 농도 분석을 위한 전통적인 통계모델은 활용 가능한 변수의 종류나 통계적인 특성에 많은 제약이 있었으나 인공지능망이나 서포트벡터머신 또는 랜덤포레스트 같은 머신러닝 기법들은 변수의 비선형성에 포함된 불확실성을 극복하여 매우 높은 정확성을 보여주는 것으로 나타나고 있다(Joharestani et al., 2019). Dutta and Jinsart (2021)은 인도 구와하티 지역의 미세먼지 농도를 다중선형회귀(Multiple Linear Regression, MLR), 인공 신경망(Artificial Neural Network, ANN), 분류회귀 의사결정나무(Classification and Regression Tree, CART)의 세 가지 방법으로 예측하여 인공 신경망이  $PM_{10}$ 의 예보에 실효적으로 적용할 수 있음을 확인하였고, Shahriar et al.(2020)은 방글라데시 주요 도시의 (초)미세먼지 농도를 대기오염물질 농도와 기상요소를 결합하여 다양한 머신러닝기법을 사용하여 예측하고 정확성을 비교하였다. Madhavi et al.(2014)은 뉴질랜드 오클랜드 지역에서 실시간 측정되는 기상요소를 인공신경망 모델에 적용하여 이산화질소 농도를  $R^2$ 가 최대 0.95가 되도록 예측한 바 있으며 Goulier et al.(2020)은 독일 Münster 지역의 질소산화물과 오존 농도를 실시간 측정자료와 교통량, 시계열 정보를 이용하여 인공신경망으로 예측을 하였다. 이와 더불어 국내에서도 머신러닝을 대기오염 예측에 활용한 다수의 사례가 있다(Cha and Kim, 2018; Cho et al., 2019). 이들 연구결과들에서 예측된 대기오염변수들은 실측치

**Table 1.** Grid configurations of modeling system

Model domain		Grids information
Horizontal grids	Domain1	174 × 128 × 27 km
	Domain2	67 × 82 × 9 km
	Domain3	83 × 83 × 3 km
	Domain4	78 × 70 × 1 km
Vertical levels	sigma levels	1.000, 0.995, 0.990, 0.985, 0.970, 0.950, 0.930, 0.910, 0.880, 0.840, 0.800, 0.740, 0.700, 0.600, 0.450, 0.000

**Table 2.** Descriptions of model configurations

WRF physics options	CMAQ options
<ul style="list-style-type: none"> <li>• Microphysics option : WSM 6-calss graupel</li> <li>• Long wave radiation : RRTM</li> <li>• Short wave radiation : Goddard</li> <li>• Surface layer scheme : MM5 similarity</li> <li>• Land surface scheme : Noah Land Surface Model</li> <li>• PBL scheme : YSU</li> <li>• Cumulus parameterization : Kain-Fritsch</li> </ul>	<ul style="list-style-type: none"> <li>• Horizontal advection : YAMO</li> <li>• Vertical advection : WRF</li> <li>• Horizontal diffusion : Multi-scale</li> <li>• Vertical diffusion : Eddy</li> <li>• Gas-phase chemistry : CB5</li> <li>• Aerosol chemistry : AE5</li> <li>• Dry deposition : M3Dry</li> </ul>

와 상당한 일치도를 보이는 것으로 나타나 머신러닝이 광화학 수치모델의 약점을 보완할 수 있는 대안이 될 수 있다고 판단된다. 이 중 신경망 모형은 전산 기술의 발달로 최근에 다양한 분야에서 적용되고 있으며 대기오염 연구에도 그 활용성이 높아지고 있다(Cho et al., 2019). 신경망 모형은 단일층으로 구성되는 얇은 신경망 모형(Shallow Neural Network, SNN), 입출력 층과 1개의 은닉층(hidden layer)으로 구성된 인공 신경망(Artificial Neural Network, ANN), 여러 개의 은닉층을 가지는 심층 신경망(Deep Neural Network, DNN)으로 구분된다(Bengio et al., 1994; Schmidhuber, 2015). 은닉층이 많을수록 학습 시간과 연산량이 증가하고 과적합 문제가 발생한다. 하지만 최근에는 이러한 문제점들이 해결되면서 다양한 분야에서 DNN 사용이 급격히 확대되고 있다(Shahraiyini and Sodoudi, 2016). Perez and Reyes(2002)는 칠레의 산티아고에서 1998년부터 2000년까지 8개 측정소의 미세먼지 농도와 기상정보(온도, 상대 습도 및 풍속)를 토대로 SNN을 구축하였는데 이를 통해 대기질 예측 정확도가 개선되는 것을 확인하였다. McKendry(2002)는 캐나다 Chilliwack 지역에서 시간별 대기오염물질 농도(NO, CO, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>)와 관측 기상(기온, 풍속,

풍향)을 활용하여 오존과 미세먼지 그리고 초미세먼지의 일최고 및 일평균 농도를 예측하는 통계적 기법을 연구하였다. Jeon and Son(2018)은 2010년부터 2015년까지 국내 6개 대도시의 일별 미세먼지 관측데이터를 토대로 여러 가지 통계모델을 실험하였는데, 심층 신경망 모형에 의한 등급 예측이 다른 기법(SNN, 다항 로지스틱 회귀모형, SVM, RF)보다 더 정확한 것으로 확인되었다. Shahraiyini and Sodoudi(2016)는 도시지역 미세먼지 예측을 위해서 통계모델이 사용된 기존 연구들을 광범위하게 비교하였다. 또한 최근에 Sayeed et al.(2021)은 CMAQ 모델의 결과에 인공신경망을 적용하여 우리나라 25개 대기오염측정소의 시간별 오존농도를 성공적으로 예측하여 광화학 수치모델의 결과를 개선하는 방법을 제시하였다. 본 연구는 부산광역시 보건환경연구원에서 실시간으로 운영중인 CMAQ 모델링 시스템의 초미세먼지 예측결과를 개선하는 것을 목적으로 한다. 부산광역시 보건환경연구원에서는 매일 수행되는 대기오염진단평가 시스템의 CMAQ 예측모델링 결과를 바탕으로 실시간 대기오염도의 변화 추세와 중규모 기류의 역궤적 등 관련 정보를 담당자들이 분석하여 최대 3일까지의 일평균 농도를 등급으로 예측하고 있다(Public health and environment research

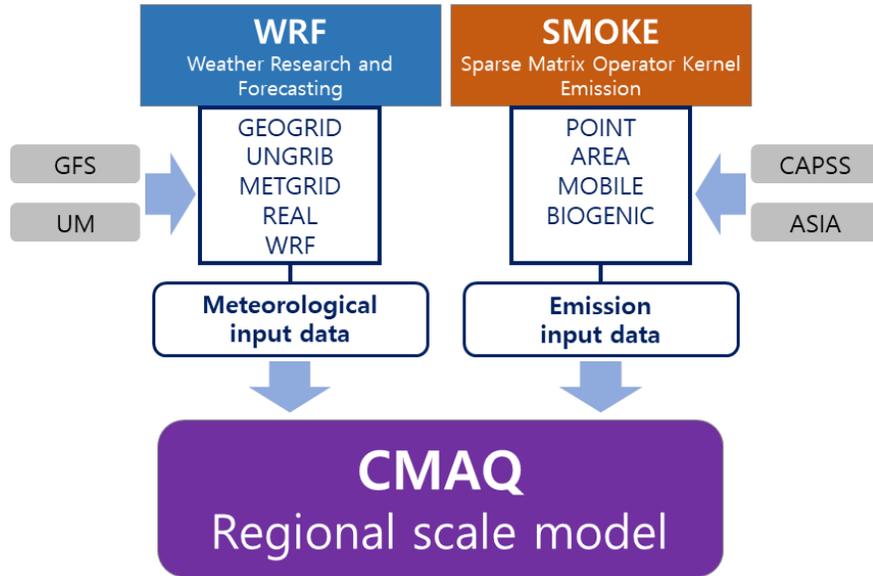


Fig. 2. Schematic diagram of the procedure in air quality diagnosis and evaluation system.

institute, 2021). 대기오염도의 변화 추세와 기상요소 분석 등 담당자가 주관으로 평가하는 부분을 수치화하고 CMAQ 모델결과와 함께 딥러닝 모델에 학습시켜 초미세먼지 농도 등급예측에 자동화 기법을 도입하였다. 본 연구는 수치모델의 근원적인 약점을 극복하여 수치모델의 정확성을 높이는 것을 목적으로 하였으며 향후에는 측정소별로 적용된 CMAQ 모델의 개선방법을 전체 모델격자로 확대하여 정확성이 개선된 부산지역 초미세먼지 예측자료를 생산하여 대기질 개선정책수립에 활용하고자 하였다.

## 2. 자료 및 방법

### 2.1. 진단평가시스템을 활용한 대기질 예측

대기질 진단평가시스템은 CMAQ 모델을 활용하여 실시간 모델링이 수행되는 시스템으로 부산광역시 보건환경연구원에서 2017년에 최초 도입하였다(Public health and environment research institute, 2021). 모델링 영역은 동아시아 27 km 격자 영역에서 한반도 9 km 격자 영역, 영남권 3 km 격자 영역, 최종적으로 부산권의 1 km 격자 영역으로 4단계의 nesting 도메인으로 구성된다(Fig. 1, Table 1).

진단평가시스템의 기상입력자료는 National Center for Environmental Prediction (NCEP)의 Global Forecast System (GFS)와 기상청 국가기상 슈퍼컴퓨터 센터의 지역예보 모델에서 생성된 Unified Model (UM) 자료를 기상모델 Weather Research Forecast (WRF)에 입력하여 생성하고 있다. GFS는 NCEP에서 운영하고 있는 전구 기상예보 수치모델이며 하루에 4 번 6시간 간격으로 실행되고 384시간의 예보자료를 포함하고 있다(NCAR Research Data Archive, 2021). 기상청 국가기상 슈퍼컴퓨터 센터에서 생성된 수치모델 자료는 영국 통합모델(UM)을 기반으로 전지구예보 모델, 지역예보모델, 국지예보모델로 구분되며 진단평가시스템에서는 지역예보모델 자료를 사용하고 있다(Korea Meteorological Administration, 2021). WRF 모델은 2005년부터 미국 NOAA 산하 기관인 NCEP의 현업 모델로 사용되고 있고 세계적으로 널리 보급되어 많은 연구에 활용되고 있는 기상모델이다(WRF MODEL USERS' PAGE, 2021). 배출량 자료를 생성하기 위해서 미국 EPA에서 제공하는 Sparse Matrix Operator Kernel Emissions (SMOKE) 모델을 적용하였다. SMOKE는 미국의 Environmental Modeling Center (EMC)에서 개발된 것으로 모델링에 필요한 배출량을 고효율로 계산할 수 있도록 배출량을 Matrix 구조

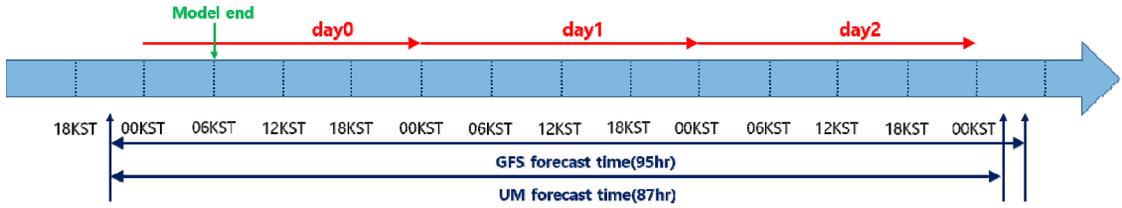


Fig. 3. Schematic diagram of the forecasting time table in the air quality diagnosis and evaluation system.

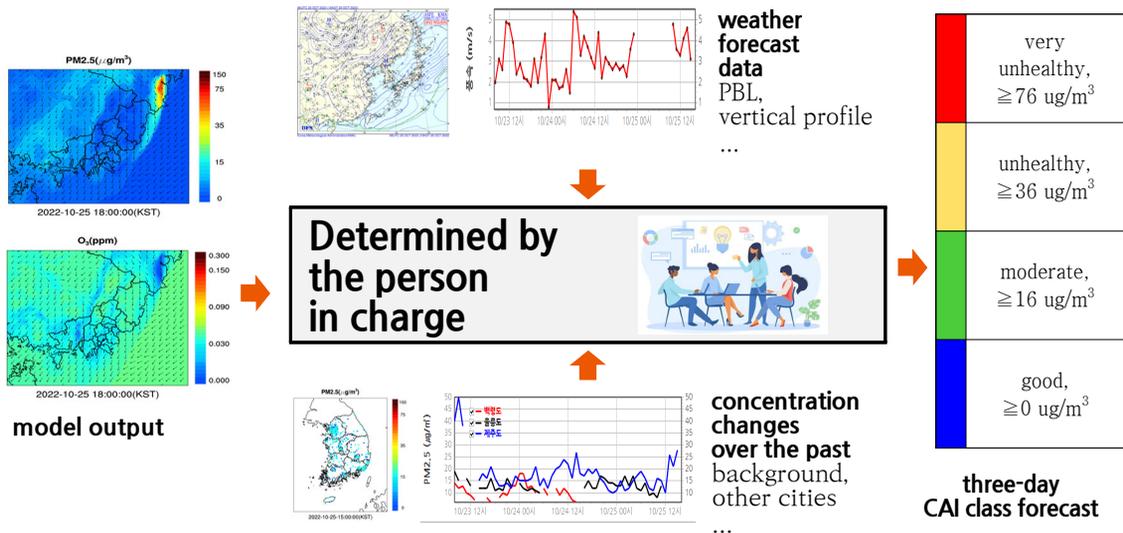


Fig. 4. Air pollution class forecasting process based on CMAQ results including weather forecast data and tendency of concentration over Korea.

체로 생성하는 배출량 모델링 시스템이다(Community Modeling and Analysis System, 2021). SMOKE 모델에 사용하는 동아시아 지역의 배출량자료로 중국의 Multi-resolution emission inventory for China (MEIC) 배출량, 중국을 제외한 아시아 지역은 Regional Emission inventory in ASia (REAS) 배출량 자료를 이용하였다. MEIC 배출량은 중국 Tsinghua University에서 개발되었으며 중국지역을 대상으로 수평해상도 0.25도로 power, industry, residential, transportation, agriculture의 5개 부문에 대하여 배출량을 공개하고 있다(MEICModel, 2021). REAS 배출량은 Ohara et al.(2007)에 의해 최초로 개발되었으며 SO<sub>2</sub>, NO<sub>x</sub>, CO, NMVOC 등 총 10개 물질에 대하여 아시아지역 전체에 대하여 0.25도 해상도로 배출량을 공

개하고 있다 (Ohara et al., 2007; Regional Emission inventory in ASia Data Download Site, 2021). 국내 배출량 자료는 환경부 국가미세먼지정보센터의 대기정책지원시스템(Clean Air Policy Support System, CAPSS)에서 공개되는 2017년 배출원별, 1 km 격자별 배출량을 사용하였다. 생성된 기상 및 배출량 입력자료들은 최종적으로 광화학 수치모델인 CMAQ에 입력된다. CMAQ은 미국 EPA가 정한 규제모형 중에서 가장 많이 이용되는 3차원 광화학 오일리러안 대기질 모델이며, 대기 중 오염농도, 건성침적, 습성침적 등 여러 가지 물리적 과정과 대기 중에서 발생하는 광화학 반응 등 상세한 물리·화학 반응 모듈을 포함하고 있다(Byun and Ching, 1999; EPA, 2021). 대기질 진단평가시스템의 구조와 모델링 옵션은 Table 2, Fig. 2에 제시하였다.

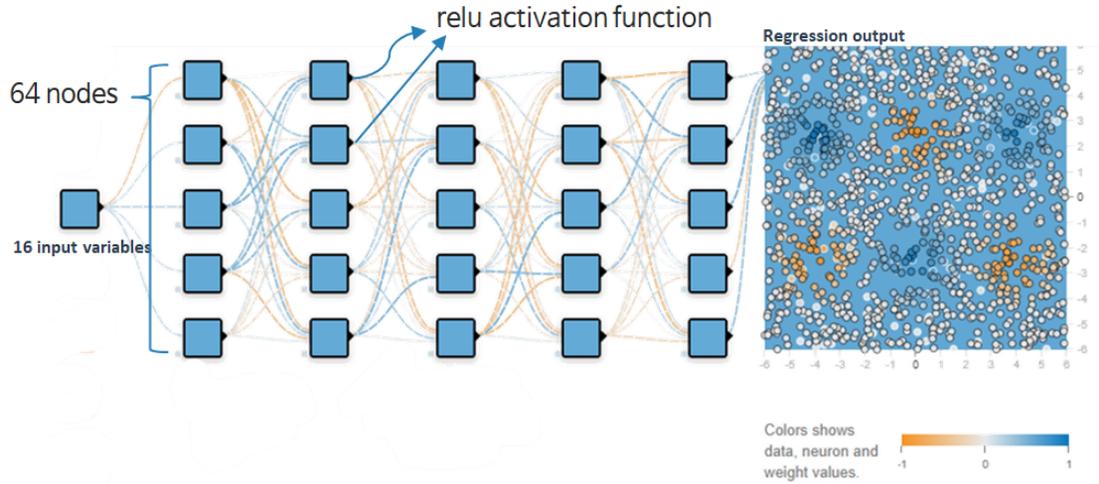


Fig. 5. Conceptual diagram of an artificial neural network including the flow of input and output data in this study.

진단평가시스템의 기상입력자료는 NCEP의 GFS와 기상청의 UM 자료를 사용하고 있으며 두 경우 모두 예측기상장을 포함하고 있다. 이에 따라 예측기상장을 활용한 CMAQ 모델의 결과에도 예측농도장이 포함하게 되는데 입력되는 기상자료와 계산에 소요되는 시간을 고려하면 Fig. 3과 같은 시간 주기성을 가지게 된다. 입력되는 기상자료에 따라 CMAQ 모델은 총 95시간(GFS) 또는 87시간(UM)의 계산치를 가지게 되나 계산에 소요되는 시간 때문에 계산대상 첫날 약 6시간 이후에 모델의 계산이 종료되면서 첫날 6시간의 결과는 예측의 기능을 못하게 된다. 하지만 현업에서는 일평균 농도 범위로 대기질 예측이 수행되고 있으므로 모델의 계산이 종료된 시점에는 당일(day0)과 내일(day1), 모레(day2)의 일평균 농도값을 활용할 수 있게 된다.

예측 기상장을 입력하여 생성된 CMAQ의 예측농도와 더불어 대기오염예보를 위하여 다양한 정보를 활용하고 있다. 대기오염도 예측발표는 매일 오전과 오후 2회 이루어지고 있으며 CMAQ 모델에서 예측되는 시간별 농도변화 자료와 더불어 예측시점까지의 부산과 주변지역 및 우리나라 전체의 대기오염도 변화와 배경농도 측정망에서의 유출입 가능성을 확인하고 있다. 또한 기상청에서 제공되는 풍향, 풍속 등 기상요소들의 예보 자료와 행성경계층(Planetary Boundary Layer, PBL), 상층기상정보 등의 요소를 고려하여 일평균 대기오염도의 등급을 결정하고 있다. 다시 말하면 매일 오전

과 오후 2회, 담당자는 당일 새벽에 계산이 끝난 CMAQ 모델의 대기오염도 예측치와 부산과 주변지역의 대기오염도 변화 추세 그리고 상층과 지상의 예측기상정보를 고려하여 당일(day0)부터 모레(day2)까지의 일최고 오존, 일평균 (초)미세먼지농도의 등급을 결정하게 된다.

## 2.2. 심층 신경망(Deep Neural Network) 모델

인공신경망(Artificial Neural Network, ANN)은 기계학습과 인지과학에서 생물학의 신경망에서 영감을 얻은 통계학적 학습 알고리즘이다. 다른 층의 뉴런(노드)들 사이의 연결 패턴, 연결의 가중치를 갱신하는 학습과정, 마지막으로 뉴런의 가중 입력을 활성화도 출력으로 바꿔주는 활성화 함수, 이 세 가지의 인자를 이용하여 정의된다. 활성화 함수는 입력 신호의 총합이 활성화를 일으키는지를 정의하는 역할을 하게 된다 (Ramsundar and Zadeh, 2018). 심층 신경망의 layer는 크게 input, hidden, output layer로 구성되며, 단일 hidden layer를 포함하는 경우 인공 신경망(ANN), 다수의 hidden layer를 포함하는 경우 심층 신경망(DNN)으로 구분한다. Fig. 5는 CMAQ 모델의 예측결과와, 기상변수, 대기질 농도값을 input layer에 전달한 후, 5개의 hidden layer를 거치고 일평균 농도를 예측하는 인공 신경망의 개념도이다. 각 layer의 뉴런을 연결하는 연결선들은 퍼셉트론에서 설명한 것과

**Table 3.** List of input parameters used to train DNN model

Variable type	Name	Description
CMAQ model output	cmaq	Daily mean CMAQ output at PM <sub>2.5</sub> monitoring site(predicted value)
Meteorological elements	CFRAC	Total cloud fraction at PM <sub>2.5</sub> monitoring site(predicted value)
	PBL	Planetary boundary level height at PM <sub>2.5</sub> monitoring site(predicted value)
	PRSC	Surface pressure at PM <sub>2.5</sub> monitoring site(predicted value)
	Q2	Mixing ratio at PM <sub>2.5</sub> monitoring site(predicted value)
	TEMP2	Temperature at PM <sub>2.5</sub> monitoring site(predicted value)
	WSPD10	Wind speed at PM <sub>2.5</sub> monitoring site(predicted value)
Air pollutant	NOx	Nitrogen oxide concentration at PM <sub>2.5</sub> monitoring site(historical value)
	O3	Ozone concentration at PM <sub>2.5</sub> monitoring site(historical value)
	CO	Carbon monoxide concentration at PM <sub>2.5</sub> monitoring site(historical value)
	SO2	Sulfur dioxide concentration at PM <sub>2.5</sub> monitoring site(historical value)
	PM10	PM <sub>10</sub> concentration at PM <sub>2.5</sub> monitoring site(historical value)
	PM25	PM <sub>2.5</sub> concentration at PM <sub>2.5</sub> monitoring site(historical value)
	PM25bkg	Spatial mean PM <sub>2.5</sub> concentration over national background monitoring site(historical value)
	ulsan	Spatial mean PM <sub>2.5</sub> concentration throughout Ulsan Metropolitan City(historical value)
gsnd	Spatial mean PM <sub>2.5</sub> concentration throughout Gyeongsangnam-do Province(historical value)	

같이 각각의 가중치를 가지고 있다. DNN은 먼저 이들을 일정 수치로 초기화한 후, 최종 output layer의 손실(loss)을 줄이는 방향으로 가중치를 갱신하게 된다.

DNN 기반의 미세먼지 예측과 관련하여 Dedovic et al.(2016)은 3년간의 사라예보 기상변수 및 미세먼지 농도를 사용하여 1개의 hidden layer를 가진 인공 신경망을 통해 미세먼지 농도를 예측하였다. 여기서 이전 년도의 미세먼지 농도 데이터를 가진 확장된 input 데이터 셋을 활용하여 미세먼지 농도 예측 성능을 높일 수 있음을 확인하였다. Cha and Kim(2018)은 2014에서 2017년간 기상 및 대기질 데이터를 이용하여 미세먼지 농도를 예측하였다. Jeon and Son(2018)은 일일의 미세먼지 농도를 ‘ 좋음’, ‘보통’, ‘나쁨’, ‘매우나쁨’의 범주로 예측하였는데 이를 위해 일별 대기오염물질, 기상자료, 전일 및 이틀 전 중국의 미세먼지 농도, 계절 변수를 사용하였으며 DNN 모델과 인공 신경망, Support Vector Machine(SVM), 다항 로지스틱 회귀 모형, random forest 모델과 예측성능을 비교하여, 200개의 node를 가진 3개의 hidden layer로 구성된 DNN 모델이 다른 모델보다 고농도 예측의 적중률이 높음을 확인하였다. 본 연구에서는 보건환경연구원의 대기질 예측항목 중 초미세먼지의 일평균 농도 예측과

정에 DNN 모형을 도입하였다. 진단평가시스템 CMAQ 모델의 결과를 기본으로 대기오염물질 농도와 기상 예측변수를 고려하여 내일의 농도 등급을 예측하는 과정에서 담당자가 경험적으로 고려하는 변수들을 수치화된 변수로 전환하여 DNN에 입력하였다. DNN 모델의 학습을 위하여 사용되는 변수는 진단평가시스템 CMAQ 모델의 지점별 초미세먼지 예측결과와, 기상요소 관련 변수, 대기오염도 관련 변수, 총 3가지 그룹으로 구분하였다(Table 3). CMAQ 모델의 예측결과는 대기오염측정소와 같은 위치의 초미세먼지 농도의 일평균값을 각각의 예측대상일에 대하여 산정하였다. 기상요소 관련 변수도 진단평가시스템의 기상모델(WRF)의 일평균 예측결과를 지점별로 산정하여 구축하였다. 기상요소 관련 변수는 각 지점별 총운량(CFRAC), 경계층 높이(PBL), 지상기압(PRSC), 혼합비(Q2), 기온(TEMP2), 풍속(WSPD10), 총 6개 항목으로 구성되며 대기오염 예측시 담당자가 가장 많이 고려하는 기상예보 자료와 관련 있는 변수들로 구성하였다. CMAQ의 예측결과와 WRF의 예측기상 변수들은 자동화 스크립터를 활용하여 모델의 결과가 생성되면 자동적으로 추출하여 변수로 저장하였다. 대기오염도 관련 변수도 예측 시 가장 많이 고려하는 변수들로 구성하였

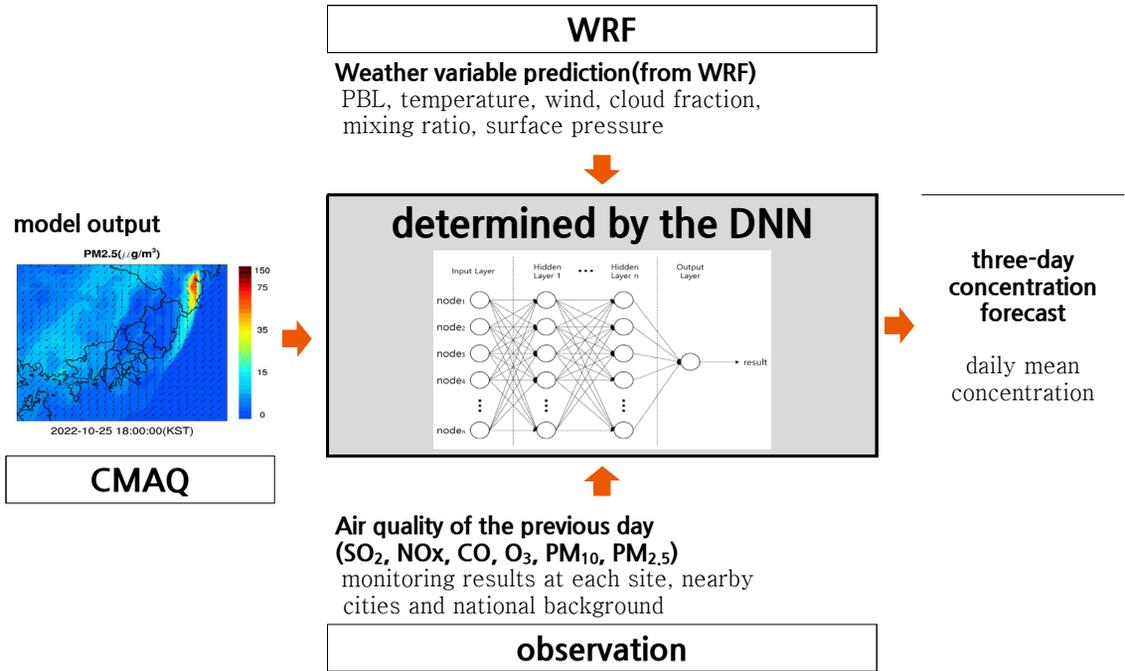


Fig. 6. Schematic diagram of the air pollution forecasting process with DNN.

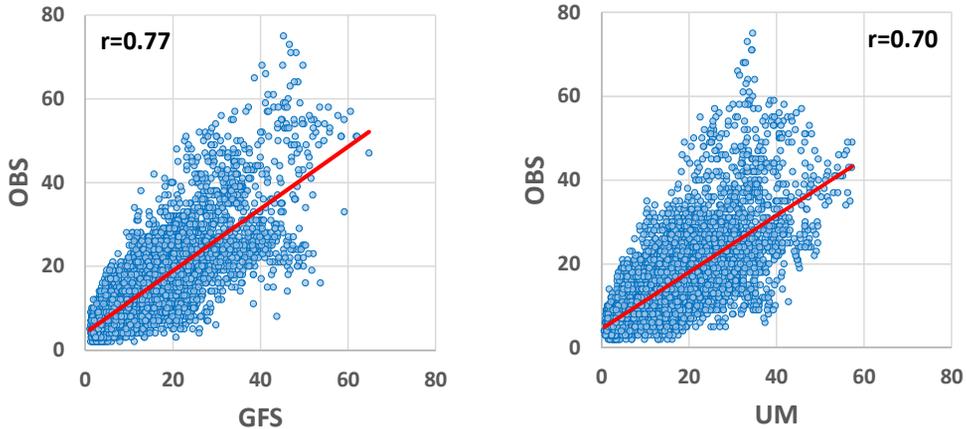
는데 측정지점별 대기오염 측정항목들(NO<sub>x</sub>, O<sub>3</sub>, CO, SO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>)의 전일 일평균 자료와 국외 유입의 영향을 파악하기 위하여 국가 배경농도의 측정망에서의 전일 일평균 초미세먼지 농도, 그리고 부산 주변지역에서의 영향을 고려하기 위하여 울산과 경상남도의 전일 일평균 초미세먼지 농도, 총 9개 항목을 변수로 선정하였다. DNN 훈련을 위하여 선정된 총 16개 변수 중 CMAQ 모델결과와 기상요소 관련 변수들은 각각 진단 평가시스템의 예측모델링 결과 생성되는 예측대상일(day0)의 일평균을 사용하였고 대기오염도 관련 변수들은 측정지점에서 측정된 예측 대상 전일의 평균값을 사용하였다. 또한 입력변수들의 단위 차이에 의한 영향을 배제하기 위하여 모든 변수들은 표준화를 하여 사용하였다.

Fig. 6은 본 연구에서 설계된 DNN 모델을 도입한 대기질 예측과정을 요약한 그림이다. 지점별 CMAQ 예측결과와 담당자에 의한 대기오염 예측에서 가장 고려가 많이 되는 기상요소와 대기오염 요소들을 변수화하여 입력하였다. 여기서 기상요소들은 진단평가시스템의 지점별 예측값을 사용하였고 대기오염 변수들은 부산

과 인접지역, 국가 배경농도측정망의 전일 평균값을 사용하였다. DNN 훈련을 위한 타겟은 예측당일의 초미세먼지 일평균 농도이며 따라서 각 지점별로 DNN을 구축하고 예측된 값과 실측치를 비교하였다. DNN 학습을 위하여 2021년 연간 데이터를 사용하였으며 최적의 DNN 파라미터를 적용한 지점별 예측모델을 결정하고 2022년 1월에서 9월까지 측정된 자료를 적용하여 DNN이 적용된 예측시스템의 재현성을 평가하였다. DNN 모델의 세부 파라미터 선정을 위한 모델 튜닝과정을 거쳐, 최종적으로 5개의 hidden layer를 가지는 DNN 모델을 구축하였다. 모델 훈련시 훈련데이터의 20%를 검증데이터로 추출하였으며 epoch는 100, batch size는 64로 한 후 최적의 결과를 모델값으로 선정하였다. DNN 학습시 손실함수(Loss function)를 최소화하는 방법으로는 RMSProp (Root Mean Square Propagation) 옵션을 사용하였고 손실함수는 MAE (Mean Absolute Error)를 적용하였으며 input layer와 hidden layer의 활성화 함수는 Relu(Rectified Linear Unit)함수를 output layer에는 농도값 출력력을 위하여 선형 함수를 적용하였다. Relu함수는 은닉층에서 가장

**Table 4.** Performance of daily mean predicted PM<sub>2.5</sub> from CMAQ model using GFS and UM meteorological input data.

MB		RMSE	
GFS	UM	GFS	UM
-0.20	0.58	6.52	7.44

**Fig. 7.** Scatter plots of the daily mean PM<sub>2.5</sub> concentration between CMAQ results and observation values using GFS(left) and UM(right) meteorological input data.

많이 사용되는 함수로 이전에 사용되던 sigmoid 함수의 기울기 소실 현상을 해결한 함수이다. DNN 모델의 구현은 Python 기반의 Keras (2.1.1) 프레임워크를 사용하였고, 인공지능 엔진은 Google의 오픈소스 라이브러리 TensorFlow(1.4)를 적용하였다(Ramsundar and Zadeh, 2018).

### 2.3. 모델의 적합성 평가 및 분석방법

진단평가시스템의 CMAQ 모델의 결과와 DNN으로 개선된 모델결과의 적합성은 관측자료와의 비교를 통하여 수행하였다. 모델치와 관측치의 적합성을 확인하는 변수들은 매우 다양하며 본 연구에서는 Mean Bias (MB), Root Mean Square Error (RMSE) 그리고 R 값을 계산하여 모델의 적합성을 평가하였다(Yanga et al., 2019). MB는 모델값과 측정값의 차이를 전 기간에 대하여 평균한 것으로 +값이면 측정값에 비해 모델값이 과대평가, -값이면 과소평가하는 것으로 판단할 수 있다. RMSE는 모델값과 측정값의 평균 제곱근 오차로 측정값에 비하여 모델값이 어느 정도의 오차를 가지는지를 판단하는 변수이다. R은 상관계수로 모델값이 측정값에 어느 정도 일치하는지를 설명하는 변수이다.

$$MB = \frac{1}{n} \sum_{i=1}^n (\text{model}(i) - \text{Obs}(i)) \quad (1)$$

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (\text{model}(i) - \text{Obs}(i))^2 \right]^{\frac{1}{2}} \quad (2)$$

$$R = \frac{\sum_{i=1}^n [(\text{model}(i) - \overline{\text{model}})(\text{Obs}(i) - \overline{\text{Obs}})]}{\sqrt{\sum_{i=1}^n (\text{model}(i) - \overline{\text{model}})^2 \sum_{i=1}^n (\text{Obs}(i) - \overline{\text{Obs}})^2}} \quad (3)$$

식에서 Model(i)는 CMAQ 또는 DNN으로 개선된 모델의 결과를, Obs(i)는 도시대기측정소의 관측값을, 변수위의 바는 해당 변수의 평균을 의미한다.

## 3. 결과 및 고찰

### 3.1. 진단평가시스템 적합성 검증

진단평가시스템 CMAQ 모델의 적합성을 확인하기 위하여 2022년 1월에서 9월까지 부산지역 도시대기측정소 28개소의 일평균 초미세먼지 농도를 모델의 당일

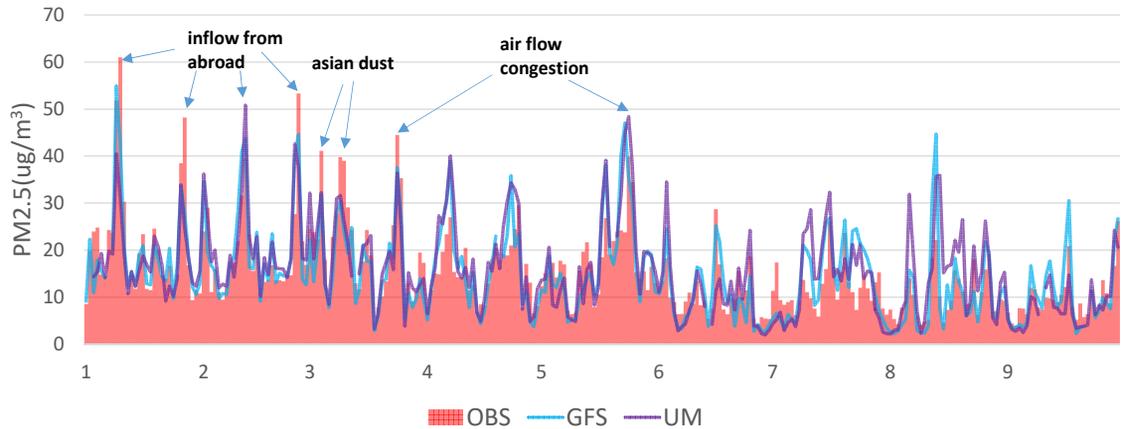


Fig. 8. Time series of the daily mean PM<sub>2.5</sub> concentration between CMAQ results using GFS and UM meteorological input data and observation values.

예측결과와(day0) 비교하였다. CMAQ 모델에 입력되는 초기 기장장의 영향을 파악하기 위하여 GFS와 UM 기상장을 사용하는 경우로 구분하여 모델 적합성 변수인 MB, RMSE, R를 계산하였다(Table 4, Fig. 7). 모델의 편향성을 나타내는 MB의 경우 GFS는  $-0.20$ 으로 실측치를 과소 모의 하는 것으로 나타났으며 UM은  $0.58$ 로 실측치를 과대모의 하는 것으로 나타났다. UM의 절대값이 더 크기 때문에 UM이 GFS보다 다소 편향성이 높은 것으로 판단된다. 모델치와 실측치의 오차율을 나타내는 RMSE의 경우도 GFS는  $6.52$  UM은  $7.44$ 로 GFS가 UM보다 관측치와의 오차가 낮음을 알 수 있다. 관측치에 모델결과의 설명력인 R도 GFS가  $0.77$ , UM이  $0.70$ 로 나타나 GFS가 다소 높았다. 입력기상자료에 따른 CMAQ 예측모델 결과를 살펴보면 GFS는 관측치를 과소모의 하고 UM은 과대모의 하는 경향이 있으며 정확성은 GFS가 다소 높은 것을 알 수 있다. 다시말하면 GFS 기상입력자료가 UM보다 개선된 결과를 보이는 것으로 판단되며 이는 입력되는 기상자료 품질의 차이에 의한 것으로 판단된다. CMAQ 모델의 다른 입력자료나 옵션들은 모두 같기 때문에 GFS와 UM을 사용한 WRF 기상모델링 결과를 기상관측자료와 비교해 보면 그 원인을 파악할 수 있을 것이나 본 연구의 범위를 벗어나므로 향후 시스템 개선 등의 과정에서 고려해 볼 계획이다. 다만 이러한 기상입력자료에 대한 차이는 CMAQ 모델결과를 바탕으로 최종 예보를 수행하는 과정에서 필수적으로 고려하여야 할 사항으로 판단된다.

Fig. 8은 분석대상기간 부산지역 28개 대기환경측정소의 전체 일평균 초미세먼지농도의 시간변화를 CMAQ 모델결과와 비교하여 나타낸 그림이다. 부산지역 전체 일평균이 '나쁨' 이상인  $35 \text{ ug/m}^3$ 을 초과한 날은 총 12일 발생하였으며 CMAQ 모델의 결과와 기상청의 분석자료를 바탕으로 총 세 가지의 주요 원인을 구분하였다. 분석대상 기간중 일평균 미세먼지가 가장 높은 날은 2022년 1월 10일,  $61 \text{ ug/m}^3$ 이었으며 국외유입의 영향으로 전일부터 일평균 농도가 증가하였다. 국외유입에 의한 영향은 이후 2월까지 일평균 농도를 증가시키는 주요 원인이었다. 3월의 고농도는 황사에 의한 영향이었으며 3월 말과 5월에는 국내 대기정체로 인하여 일평균 미세먼지가 증가하였다. CMAQ 모델의 예측치는 전반적으로 일변화 패턴을 잘 따라가는 것으로 판단이 된다. 하지만 국외유입과 황사발생에 따른 고농도 발생시 모델 예측치가 실측치보다 낮게 예측되는 경향이 있으며 국내 대기 정체에 따른 고농도에서는 모델치의 특별한 경향이 없이 높거나 낮게 나타나고 있다. 한편 장마나 강수에 의해 낮은 농도가 나타나는 하절기에는 모델치가 실측치보다 높은 결과가 나타나기도 하였다. 일평균농도 변화에 대한 CMAQ 모델의 결과는 전반적인 추세를 잘 반영하지만 고농도나 저농도에서 차이가 나는 경우가 있으며 이는 광화학모델의 근본적인 문제에 기인한 것으로 최종적인 대기질 예측시 필수적으로 고려하여야 할 사항으로 판단된다.

광화학수치모형을 활용하여 시간 또는 일평균 대기

**Table 5.** Performance of prediction by concentration grade using GFS meteorological input data during training periods (2022. 1.-2022. 9.)

total : 74.4%		GFS				total
		very unhealthy	unhealthy	moderate	good	
OBS	≥76	very unhealthy				
	≥36	unhealthy	146(46.2%)	169	1	316
	≥16	moderate	219	1,676(66.56%)	623	2,518
	≥0	good	6	891	3,735(80.63%)	4,632
total			371	2,736	4,359	7,466

**Table 6.** Performance of prediction by concentration grade using UM meteorological input data during training periods (2022. 1.-2022. 9.)

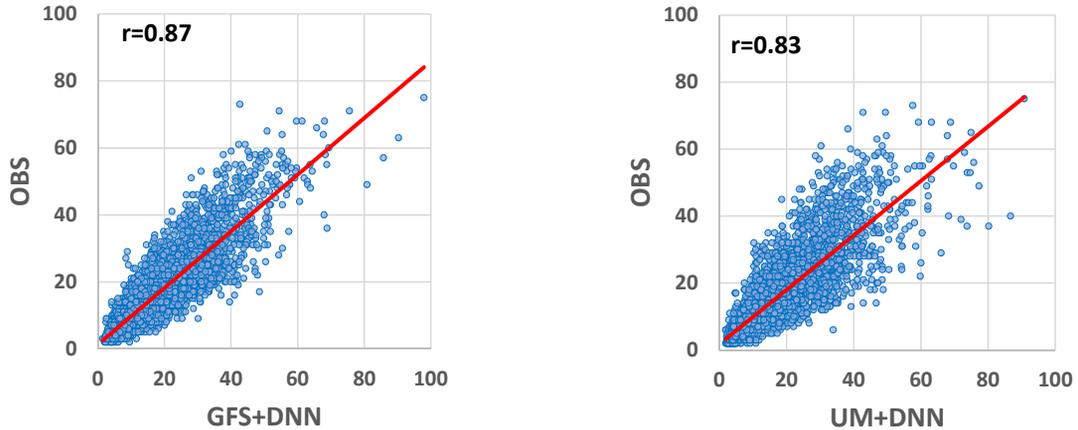
total : 70.0%		UM				total
		very unhealthy	unhealthy	moderate	good	
OBS	≥76	very unhealthy				
	≥36	unhealthy	138(43.53%)	179		317
	≥16	moderate	214	1,731(68.96%)	565	2,510
	≥0	good	34	1,231	3,313(72.37%)	4,578
total			386	3,141	3,878	7,405

질을 예측할 경우 앞에서 살펴본 것처럼 모델의 적합성 문제로 실효성에 의문이 들게 된다. 따라서 대기오염지수나 등급을 활용한 예측방법이 국 내외에서 활용되고 있다(Department for Environment Food & Rural Affairs, 2021; Ministry of Earth Science, Govt. of India, Indian Institute of Tropical Meteorology, 2021). 국내의 경우 통합대기환경지수에 따른 농도 등급을 활용하고 있는데 농도 수준에 따라 ‘ 좋음’, ‘보통’, ‘나쁨’, ‘매우나쁨’의 4단계로 구분하고 있다(Table 4, 5). 부산광역시 보건환경연구원에서도 CMAQ 모델결과를 바탕으로 향후 3일간의 일평균 초미세먼지 농도의 등급을 최종적으로 예측발표하고 있다. Table 5, 6은 GFS와 기상청 UM 기상입력 자료를 적용한 28개 대기환경측정소 전체에 대하여 예보당일(day0)의 CMAQ 초미세먼지 일평균 예측결과를 농도 범위로 환산하고 실제 관측값의 농도 범위와 일치 여부를 나타낸 것이다. Table에서 숫자는 농도등급의 발생횟수를 의미하며 괄호는 실제로 발생한 농도등급의 횟수에 대한 모델에서 계산한 농도등급의 적중률을 의미한다. GFS와 UM 모두 ‘ 좋음’ 등급에서 각각 80.63%, 72.37%로 가장 높은

적중률을 보였다. ‘보통’ 등급에서 각각 66.56%, 68.96%, ‘나쁨’ 등급에서 각각 46.5%, 43.53%로 전체 적중률은 GFS가 74.4%, UM이 70.0%로 GFS 기상입력자료가 UM보다 높은 것을 알 수 있다. 농도등급에 의한 적중률 계산은 DNN의 테스트 데이터로 사용되는 2022년 1월부터 9월까지를 대상으로 계산하였으며 같은 기간에 ‘매우나쁨’ 등급은 발생하지 않았다. 부산광역시 보건환경연구원에서는 부산지역을 4개 권역으로 구분하고 CMAQ 모델결과를 권역별로 구분하여 평균 농도 계산하고 추가적인 정보를 활용하여 농도등급 예측을 실시하고 있으나 본 연구에서는 CMAQ 모델의 측정소별 예측결과 전체를 분석하여 차이가 있다. 하지만 대기오염피해 예방을 위한 농도등급 예측에 있어 중요한 부분인 고농도 등급의 CMAQ 모델 자체만의 적중률이 약 43%에서 46%인 것은 다소 아쉬움이 있으며 이러한 차이를 극복하기 위하여 추가적인 정보와 담당자들의 주관을 활용이 필수적인 것으로 판단된다. 그러나 담당자들의 전문성 정도나 주관에 따라 예측결과에 차이가 있을 수 있어 본 연구와 같이 객관화된 방법을 도입할 필요가 있음을 다시 한번 확인할 수 있었다.

**Table 7.** Performance of daily mean predicted PM<sub>2.5</sub> from CMAQ+DNN.

MB		RMSE	
GFS+DNN	UM+DNN	GFS+DNN	UM+DNN
1.34	1.32	5.01	5.76

**Fig. 9.** Scatter plots of the daily mean PM<sub>2.5</sub> concentration between CMAQ+DNN results and observation values using GFS(left) and UM(right) meteorological input data.

### 3.2. CMAQ 결과 개선을 위한 DNN의 적용

GFS와 UM을 사용한 진단평가시스템 CMAQ모델의 일평균 초미세먼지 예측결과를 개선하기 위하여 CMAQ 모델의 결과에 DNN을 적용하였다. 전술한 바와 같이 DNN의 입력자료는 CMAQ 모델 결과와 더불어 담당자들이 예보과정에서 가장 많이 확인하는 사항을 기상과 대기질 농도변화 관련 변수로 구분하여 입력하였고, 2021년 1월부터 12월까지 CMAQ 모델의 결과와 측정지점별 관측값을 사용하여 모델을 훈련하였다. 기상관련 변수는 담당자들이 기상청 홈페이지의 그래프나 예보자료에서 확인하며 본 연구에서는 담당자들이 확인하는 요소들과 관련된 변수들을 WRF 기상모델에서 추출하여 데이터베이스화하였다. 훈련과정에서 측정지점별로 최적의 DNN을 결정하고 모델성능을 확인하기 위하여 2022년 1월부터 9월까지 실제 사례에 적용하였다. Table 7은 CMAQ 예측결과에 DNN 적용하고 모델의 적합성 변수를 계산한 결과이다. CMAQ 모델만 사용하였을 경우 MB는 GFS는 관측값을 과소 모의하고 UM은 관측값을 과대모의 하였으며 DNN이 적용되면 두 경우 모두 관측값을 과대 모의하는 것으로

나타났다. 현실적으로 미래의 값을 정확하게 예측하는 것이 불가능한 점을 고려하면, 예측치의 과대모의가 대기질 예보를 통한 피해예방의 측면에서 나은 것으로 판단된다. 관측값과의 오차율은 GFS와 UM 각각 5.01, 5.76으로 CMAQ의 결과만을 계산한 경우, 6.52, 7.44보다 상당히 개선되었음을 알 수 있다. CMAQ 결과를 관측치와 일치시키기 위한 추가적인 과정을 고려할 때 모델치의 개선은 당연한 결과로 판단되나 약 5-6 ug/m<sup>3</sup>의 오차가 결과에 여전히 포함되는 것을 알 수 있으며 이는 DNN의 입력자료 구성과 학습과정에서 결정되는 파라미터에서 기인하는 것으로 판단된다. 관측값과의 상관성도 CMAQ만을 사용한 경우 0.77, 0.70에서 0.87, 0.83으로 상당히 개선되는 것으로 나타났다. 따라서 CMAQ의 결과에 DNN을 적용할 경우 모델의 오차가 감소하고 설명력이 증가하여 대기질 예보의 활용성이 증가할 것으로 판단된다. 다만, 본 연구의 경우 DNN의 입력변수로 기상과 대기질 분야 총 16개를 사용하였고 훈련기간도 2021년 1년의 자료만을 사용한 결과로 입력자료와 훈련기간을 증가시킬 경우 획기적인 개선효과가 있을 것으로 예상된다.



Fig. 10. Comparison of daily predicted  $PM_{2.5}$  time series between CMAQ and CMAQ+DNN.

DNN을 CMAQ 모델의 결과에 적용하면 CMAQ의 결과를 직접 사용하는 경우보다 모델의 적합성 변수들이 개선되는 것을 확인하였다. Fig. 10은 각각의 기상 입력자료에 대한 CMAQ의 결과와 DNN이 적용된 결과의 시계열 그림이다. 위쪽은 GFS 입력자료이며, 아래쪽은 UM입력자료를 나타낸다. 1월에서 3월간의 고농도 발생기간 DNN을 적용하면 CMAQ의 결과보다 실측치의 고농도 패턴을 더 잘 따라가는 것을 알 수 있다. 농도 수준이 상대적으로 낮은 하절기에는 DNN을 적용하면 CMAQ의 과대모의가 개선되는 것을 확인할 수 있었다. 즉, CMAQ의 결과에 DNN을 적용하면 고농도기간 CMAQ의 과소모의와, 저농도 기간 과대모의가 개선되면서 실측치의 변화 패턴을 더 잘 모사하게 되며 따라서 모델의 적합성 변수들이 개선되는 것으로 판단된다.

### 3.3. 농도등급별 개선결과

Table 8, 9는 GFS와 UM 기상입력 자료를 사용한 CMAQ 결과에 DNN을 적용하여 개선된 모델결과의 28개 대기환경 측정소 전체에 대하여 예보당일(day0)의 초미세먼지 일평균 예측결과를 농도 범위로 환산하고 실제 관측값의 범위와 일치 여부를 나타낸 것이다. GFS 기상입력자료의 경우 CMAQ 단일결과와 비교하여 ' 좋음' 등급의 적중률은 80.63에서 79.40%로 다소 감소하였지만 '보통' 등급, 66.56에서 82.85%, '나쁨' 등급은 46.5에서 67.63%로 개선되어 전체 적중률은 74.4에서 80.1%로 개선되었음을 확인할 수 있었다. 특히 대기오염에 의한 사전피해 예방의 관점에서 중요성이 높은 '나쁨' 등급의 적중률이 크게 증가한 것은 중요한 개선점이라 할 수 있다. DNN 결합 모델의 정확성 확

**Table 8.** Performance of prediction by concentration grade using GFS meteorological input data and DNN during training periods(2022. 1.-2022. 9.)

total : 80.1%		GFS+DNN				
		very unhealthy	unhealthy	moderate	good	total
OBS	≥76 very unhealthy					
	≥36 unhealthy	5	211(67.63%)	96		312
	≥16 moderate		164	2,049(82.85%)	260	2,473
	≥0 good		1	918	3,542(79.40%)	4,461
total		5	376	3,063	3,802	7,246

**Table 9.** Performance of prediction by concentration grade using UM meteorological input data and DNN during training periods (2022. 1.-2022. 9.)

total : 77.9%		UM+DNN				
		very unhealthy	unhealthy	moderate	good	total
OBS	≥76 very unhealthy					
	≥36 unhealthy	5	184(58.79%)	124		313
	≥16 moderate		156	1,951(79.15%)	358	2,465
	≥0 good		6	945	3,480(78.54%)	4,431
total		5	346	3,020	3,838	7,209

안을 위한 테스트 기간인 2022년 1월부터 9월간에는 ‘매우나쁨’ 등급의 발생이 없었으나 DNN 적용시 5개의 사례일을 ‘매우나쁨’ 으로 예측하였고 실제로는 ‘나쁨’ 등급이 발생하였다. UM 기상입력자료의 경우 CMAQ 단일의 결과와 비교하여 ‘ 좋음’ 등급, 72.37에서 78.54%로 ‘보통’ 등급, 68.96에서 79.15%, ‘나쁨’ 등급은 43.53에서 58.79%로 개선되어 전체 적중률은 70.0에서 77.9%로 개선되었다. GFS의 경우와 동일하게 ‘나쁨’ 등급의 적중률은 43.53에서 58.79%로 크게 증가하였으며 이는 사전 피해예방의 측면에서 중요한 개선이라 할 수 있다. UM 기상입력자료의 경우도 CMAQ 단일의 경우 ‘매우나쁨’ 등급은 발생하지 않았지만 DNN 결합 모델은 5개의 사례일을 ‘매우나쁨’ 으로 예측하였으며 고농도에 대한 과대모의 경향은 예보에 대한 관점에서는 어느 정도 필요한 부분으로 판단된다. 대기오염도 예보를 위하여 현재 보건환경연구원에서는 진단평가시스템 CMAQ 모델의 결과에 지상 및 상층의 기상예보자료와 대기오염도 변화추세에 대한 담당자의 주관을 추가하여 초미세먼지 일평균 농도등급을 예측하고 있다. 숙련된 담당자일 경우 이는 광화학수치모델의 단점을 극복하기 위한 합리적인 방법일 수 있

으나 현업에서는 담당자의 잦은 변경과 개인별 전문성의 차이 등으로 매우 부정확한 결과를 만들 수 있다. 따라서 주관적으로 고려되는 기상예보자료와 부산의 인접지역의 대기오염도 변화추세를 변수화하고 CMAQ 결과 개선을 위한 심층 신경망 모형(DNN)을 구축할 경우 체계적인 기준으로 개선된 일평균 농도예측을 수행할 수 있을 것으로 판단된다. 본 연구의 경우 DNN의 학습기간을 2021년으로 한정하였고 입력변수도 현재 진단평가시스템에서 수집되는 변수들 중에서 추출하여 다소 아쉬운 점이 있으며 향후 학습기간을 늘리고 입력변수를 다양화 할 경우 훨씬 개선된 결과를 도출할 수 있을 것으로 기대된다.

#### 4. 결론

부산광역시 보건환경연구원에서는 대기질진단평가 시스템을 활용하여 초미세먼지에 대한 3일 후 까지의 대기오염예보를 수행하고 있다. 진단평가시스템 CMAQ 모델의 실시간 모델링 결과를 바탕으로 기상예보자료와 대기오염도 변화 추이를 담당자가 고려하여 일평균 초미세먼지 농도등급을 발표하고 있다. 하지만

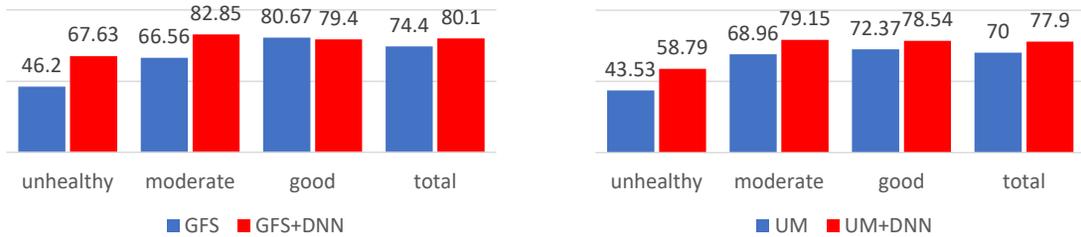


Fig. 11. The change of accuracy rate at each class after DNN application.

광화학수치모델의 기본적인 한계로 CMAQ 결과의 정확성이 낮으며 담당자의 숙련도에 따라 예보결과가 달라지는 문제점이 있다. 본 연구는 이러한 문제점을 개선하고자 최근에 다양한 분야에서 그 활용도가 매우 높아지고 있는 심층 신경망 모델을 CMAQ의 결과에 적용하였다. 현재 담당자가 주관적으로 고려하고 있는 기상예보자료와 대기오염도 변화 추세를 수치화하여 입력하고 2021년 사례를 학습자료로 최적의 모델파라미터를 결정한 후 2022년 1월에서 9월까지의 사례에 적용하여 CMAQ 단일모델의 결과와 비교하였다.

- 1) 진단평가시스템 CMAQ 모델의 예측치는 전반적으로 실측치의 일변화 패턴을 잘 모사하고 있으나 국외 유입과 황사에 따른 고농도 발생 시 모델 예측치가 실측치보다 낮게 예측되는 경향이 있으며 국내 대기 정체에 따른 고농도에서는 모델치의 특별한 경향이 없이 높거나 낮게 나타나고 있다. 모델과 관측치의 상관계수는 GFS 입력자료를 사용할 경우 0.77, UM은 0.70로 나타났으며 이는 입력되는 기상자료의 특성이 반영된 결과이며 최종적인 대기질 예측시 필수적으로 고려하여야 할 사항으로 판단된다.
- 2) 본 연구에서 구축한 DNN 시스템을 CMAQ 모델의 결과에 적용할 경우 1월에서 3월간의 고농도 발생기간 실측치의 고농도 패턴을 CMAQ 단일의 경우보다 더 잘 모사하며 농도 수준이 상대적으로 낮은 하절기 CMAQ의 과대모의가 개선되는 것을 확인할 수 있었다. 모델의 오차율(RMSE)은 GFS와 UM 각각 5.01, 5.76으로 CMAQ의 결과만을 계산한 경우, 6.52, 7.44보다 개선되었으며 관측값과의 상관성도 CMAQ만을 사용한 경우 0.77, 0.70에서 0.87, 0.83으로 상당히 개선되었다.
- 3) 실제적으로 현업에서 활용되고 있는 농도 등급별 정

확도를 살펴보면 GFS 기상입력자료의 경우 CMAQ 단일결과와 비교하여 전체 적중률은 74.4에서 80.1%로 개선되었으며 UM 기상입력자료도 CMAQ 단일 결과와 비교하여 전체 적중률은 70.0에서 77.9%로 개선되었다. CMAQ 단일의 경우 '매우나쁨' 등급은 발생하지 않았지만 DNN 결합 모델은 5개의 사례일을 '매우나쁨'으로 예측하였으며 이러한 고농도에 대한 과대모의 경향은 예보에 대한 관점에서 어느 정도 필요한 부분으로 판단된다.

본 연구의 경우 DNN의 학습기간을 2021년으로 한정하였고 입력변수도 현재 진단평가시스템에서 수집되는 변수들 중에서 선정하여 다소 아쉬운 점이 있으며 향후 학습기간을 늘리고 입력변수를 다양화할 경우 훨씬 개선된 결과를 도출할 수 있을 것으로 기대된다.

## REFERENCES

- Abdullah, A. M., Ramly, Z. T. A., 2020, Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia, *Atmosphere*, 11(289), 1-14.
- Bengio, Y., Simard, P., Frasconi, P., 1994, Learning long-term dependencies with gradient descent is difficult, *IEEE trans. neural netw.*, 5(2), 157-166.
- Byun, D. W., Ching, J. K. S., 1999, Science algorithms of the EPA Models-3 Community Multiscale Air Quality(CMAQ) modeling system, U.S. Environmental Protection Agency(US EPA), EPA/600/R-99/030.
- Cha, J. W., Kim, J. Y., 2018, Development of data mining algorithm for implementation of fine dust numerical prediction model, *J. Korea Inst. Inf. Commun. Eng.*, 22(4), 595-601.
- Cho, K. H., Lee, B. Y., Kwon, M. H., Kim, S. C., 2019, Air quality prediction using a deep neural network

- model, *J. Korean Soc. Atmos. Environ.*, 35(2), 214-225.
- Choi, D. R., Koo, Y. S., 2013, An Evaluation of the influence of boundary conditions from GEOS-Chem on CMAQ Simulations over east Asia, *J. Korean Soc. Atmos. Environ.*, 29(2), 186-198.
- Community Modeling and Analysis System, 2021, Sparse Matrix Operator Kerner Emissions Modeling System, <https://www.cmascenter.org/smoke/>.
- Dedovic, M. M., Avdakovic, S., Turkovic, I., Dautbasic, N., Konjic, T., 2016, Forecasting PM<sub>10</sub> concentrations using neural networks and system for improving air quality, 2016 XI International Symposium on Telecommunications(BIHTEL), 1-6.
- Department for Environment Food & Rural Affairs, 2021, UK AIR, [https://uk-air.defra.gov.uk/forecasting/?day=2#forecast\\_map](https://uk-air.defra.gov.uk/forecasting/?day=2#forecast_map).
- Diaz-Robles, J. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., Moncada-Herrera, J. A., 2008, A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile, *Atmospheric Environ.*, 42(35), 8331-8340.
- Dutta, A., Jinsart, W., 2021, Air Pollution in Indian cities and comparison of MLR, ANN and CART Models for Predicting PM<sub>10</sub> concentrations in Guwahati, India, *Asian J. Atmospheric Environ.*, 15(1), 1-26.
- EPA, 2021, CMAQ: The Community Multiscale Air Quality Modeling System, <https://www.epa.gov/cmaq>.
- Goulier, L., Paas, B., Ehrnsperger, L., Klemm, O., 2020, Modelling of urban air pollutant concentrations with artificial neural networks using novel input variables, *Int. J. Environ. Res. Public Health*, 17(6), 2025.
- Jeon, S. H., Son, Y. S., 2018, Prediction of fine dust PM10 using a deep neural network model, *Korean J. Appl. Stat.*, 31(2), 265-285.
- Jo, Y. J., Lee, H. J., Chang, L. S., Kim, C. H., 2017, Sensitivity study of the initial meteorological fields on the PM10 concentration predictions using CMAQ modeling, *J. Korean Soc. Atmos. Environ.*, 33(6), 554-569.
- Joharestani, M. Z., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S., 2019, PM<sub>2.5</sub> prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data, *Atmosphere*, 10(7), 373.
- Kim, J., Jang, Y. K., 2014, Uncertainty assessment for CAPSS emission inventory by DARS, *J. Korean Soc. Atmos. Environ.*, 30(1), 26-36.
- Kitayama, K., Morinoa, Y., Yamajib, K., Chatania, S., 2019, Uncertainties in O<sub>3</sub> concentrations simulated by CMAQ over Japan using four chemical mechanisms, *Atmospheric Environ.*, 198, 448-462.
- Korea Environment Corporation, 2022, Airkorea, <https://www.airkorea.or.kr/eng/>.
- Korea Meteorological Administration, 2021, <https://www.kma.go.kr/aboutkma/intro/supercom/index.jsp>.
- Madhavi, A. E., Naresh, S., Kim, N. D., Jennifer, A. S., 2014, Development of an ANN-based air pollution forecasting system with explicit knowledge through sensitivity analysis, *Atmospheric Pollut. Res.*, 5, 696-708.
- McKendry, I. G., 2002, Evaluation of artificial neural networks for fine particulate pollution (PM<sub>10</sub> and PM<sub>2.5</sub>) forecasting, *J. Air Waste Manag. Assoc.*, 52(9), 1096-1101.
- MEICModel, 2021, Tracking anthropogenic emissions in china, <http://meicmodel.org/>.
- NCAR Research Data Archive, 2021, NCEP GFS 0.25 degree global forecast auxiliary grids historical archive, <https://rda.ucar.edu/datasets/ds084.3/#description>.
- Ministry of Earth Science, Govt. of India, Indian Institute of Tropical Meteorology, 2021, System of air quality and weather forecasting and research, [http://safar.tropmet.res.in/map\\_data.php?for=current&city\\_id=1](http://safar.tropmet.res.in/map_data.php?for=current&city_id=1).
- Ohara, T., Akimoto, H., Kurokawa, J., Horii, N., Yamaji, K., Yan, X., Hayasaka, T., 2007, An Asian emission inventory of anthropogenic emission sources for the period 1980-2020, *Atmos. Chem. Phys.*, 7, 4419-4444.
- Ostro, B., Feng, W. Y., Broadwin, R., Green, S., Lipsett, M., 2007, The effects of components of fine particulate air pollution on mortality in California: results from CALFINE, *Environ. Health Perspect.*, 115(1), 13-19.
- Perez, P., Reyes, J., 2002, Prediction of maximum of 24-h average of PM<sub>10</sub> concentrations 30h in advance in Santiago, Chile, *Atmospheric Environ.*, 36(28), 4555-4561.
- Pope, C. A., Turner, M. C., Burnett, R. T., Jerrett, M., Gapstur, S. M., Diver, W. R., Krewski, D., Brook, R. D., 2015, Relationships between fine particulate air pollution, cardiometabolic disorders, and cardiovascular mortality, *Circ. Res.*, 116(1), 108-115.
- Public health and environment research institute, 2021, <https://heis.busan.go.kr/environmental/air006.aspx>.
- Ramsundar, B., Zadeh, R. B., 2018, Tensorflow for deep learning: From linear regression to reinforcement

- learning, O'Reilly Media.
- Regional Emission inventory in ASia Data Download Site, 2021, <https://www.nies.go.jp/REAS/index.html>.
- Rybarczyk, Y., Zalakeviciute, R., 2018, Machine learning approaches for outdoor air quality modelling: a systematic review, *Appl. Sci.*, 8(12), 2570.
- Sayeed, A., Choi, Y., Eslami, E., Jung, J., Lops, Y., Salman, A. K., Lee, J. B., Park, H. J., Choi, M. H., 2021, A Novel CMAQ-CNN hybrid model to forecast hourly surface-ozone concentrations 14 days in advance, *Sci. Rep.*, 11, 10891.
- Schmidhuber, J., 2015, Deep learning in neural networks: An overview, *neural networks*, 61, 85-117.
- Shahraiyni, H. T., Sodoudi, S., 2016, Statistical Modeling Approaches for PM10 prediction in urban areas: A review of 21st-century studies, *Atmosphere*, 7(2), 15.
- Shahriar, S. A., Kayes, I., Hasan, K., Salam, M. A., Chowdhury, S., 2020, Applicability of machine learning in modeling of atmospheric particle pollution in Bangladesh, *Air Qual. Atmos. Health*, 13, 1247-1256.
- Slini, T., Kaprara, A., Karatzas, K., Moussiopoulos, N., 2006, PM10 forecasting for Thessaloniki, Greece, *Environ. Model. Softw.*, 21(4), 559-565.
- UI-Saufie, A., Yahya, A., Ramli, N., Hamid, H., 2011, Comparison between multiple linear regression and feed forward back propagation neural network models for predicting PM10 concentration level based on gaseous and meteorological parameters, *Int. J. Appl. Sci. Technol.*, 1(4), 42-49.
- WRF MODEL USERS' PAGE, 2021, <https://www2.mmm.ucar.edu/wrf/users/>.
- Yanga, X., Wua, Q., Zhaob, R., Cheng, H., He, H., Ma, Q., Wang, L., Luo, H., 2019, New method for evaluating winter air quality: PM<sub>2.5</sub> assessment using Community Multi-Scale Air Quality Modeling (CMAQ) in Xi'an, *Atmospheric Environ.*, 211, 18-28.
- 
- Researcher. Woo-Gon Do  
Busan Metropolitan City Institute of Health and Environment  
[dou777@korea.kr](mailto:dou777@korea.kr)
  - Researcher. Dong-Young Kim  
Busan Metropolitan City Institute of Health and Environment  
[dykim0406@korea.kr](mailto:dykim0406@korea.kr)
  - Researcher. Hee-Jin Song  
Busan Metropolitan City Institute of Health and Environment  
[shj0853@korea.kr](mailto:shj0853@korea.kr)
  - Researcher. Gab-Je Cho  
Busan Metropolitan City Institute of Health and Environment  
[chogj64@korea.kr](mailto:chogj64@korea.kr)