

딥러닝 모델을 활용한 부산지역 대기오염 예측 고도화 연구

조완철*, 도우곤, 송경주, 박정옥

대기진단평가팀

A Study on Enhancing Air Pollution Prediction in Busan Using Deep Learning Models

Wan-cheol Cho*, Woogon Do, Gyeongju Song, Jeong-ok Park

Air Quality Monitoring and Assessment Team

Abstract

This study developed and evaluated a Deep Neural Network (DNN) model to predict air quality levels for $PM_{2.5}$, PM_{10} , and ozone in the Busan region for the next day (Day1) and the day after (Day2). Observation data from 2021 to 2023 were used as training data, and the model's performance was verified using data from January to October 2024. The CMAQ model exhibited bias, with predictions skewed toward certain air quality grades. To address this bias and improve prediction accuracy, the DNN model incorporated daily mean meteorological and air quality variables, daily maximum values, and correction factors between previous predictions and observations as input variables. Additionally, class imbalance was mitigated through proportional weight adjustments while maintaining the dataset size. As a result, the GFS-based DNN model achieved prediction accuracies of 80.8% for $PM_{2.5}$, 82.5% for PM_{10} , and 90.4% for ozone on Day1, while the accuracies for Day2 were 79.1% for $PM_{2.5}$, 83.2% for PM_{10} , and 90.4% for ozone. Similarly, the UM-based DNN model recorded accuracies of 80.5% for $PM_{2.5}$, 85.4% for PM_{10} , and 89.4% for ozone on Day1, with Day2 results at 78.4%, 83.1%, and 89.8%, respectively. These results confirmed that the optimized DNN model outperformed the CMAQ model in terms of prediction accuracy while partially addressing its bias issues. However, the model's performance still fell short of forecasts made by human experts, indicating a need for further improvement. To enhance prediction accuracy, securing long-term training data and adopting time-series deep learning models, such as LSTM or CNN, should be considered.

Key words : CMAQ, DNN, PM_{10} , $PM_{2.5}$, Ozone, Air Quality Forecast

1. 서론

현대 사회에서는 도시화와 산업화의 영향으로 다양한 오염물질이 발생하고 있으며, 그중 미세먼지(PM₁₀, particulate matter with a diameter of 10 micrometers or less), 초미세먼지(PM_{2.5}, particulate matter with a diameter of 2.5 micrometers or less), 오존(O₃, ozone)은 인체와 환경에 부정적인 영향을 미치는 주요 대기오염물질로 여겨지고 있다.¹⁾ (초)미세먼지는 교통, 산업활동 등을 통해 1차 발생하고 대기 중의 황산화물(SOX, sulfur oxides), 질소산화물(NOX, nitrogen oxides), 휘발성유기화합물(Volatile Organic Compounds, VOCs) 등과의 화학반응을 통해 2차 생성되어 다양한 화학적 특성을 가지며, 호흡기를 통해 인체로 쉽게 유입되고 입자의 크기에 따라 폐혈관까지 침투하여 호흡기 질환과 심혈관 질환을 유발하기도 한다.^{2),3)} 또한 오존은 자외선이 강하고 기온이 높은 조건에서 질소산화물, 휘발성유기화합물의 광화학 반응에 의해 공기중에서 2차 생성되어 강한 산화력을 가지고 있기 때문에 인체에 노출되어 기관지 질환과 호흡기 질환을 유발하기도 한다.⁴⁾ 이러한 이유로 (초)미세먼지, 오존은 국가에서 실시간 농도 제공과 함께 예·경보제를 운영하여 대기 중 농도에 따라 유연하게 대처할 수 있도록 관리되고 있다.

대기질 예보(예측)은 광화학 수치 모델을 활용해 수치 자료를 제공하거나, 대기오염도를 기준에 따라 분류하여 예측 결과를 제공하는 경우가 많다. 국내에서는 2014년부터 환경부 국립환경과학원 대기질통합예보센터에서 CMAQ(Community Multiscale Air Quality modeling system) 모델을 활용해 시·도별 (초)미세먼지와 오존의 농도별 등급을 기준으로 모레까지의 예보를 제공하고 있으며, 이 정보는 에어코리아 홈페이지를 통해 공개되고 있다.⁵⁾ 우리원에서는 2017년부터 CMAQ 모델을 기반으로 구축된 대기질 진단평가시스템을 활용하여, 부산지역을 동부, 서부, 남부, 중부의 4개 권역으로 세분화하여 (초)미세먼지와 오존 농도별 예보 등급을 기준으로 한 모레까지의 예측 결과를 제공하고 있다.⁶⁾

현재 대기질 예측에 활용되는 모델은 배출량, 기상, 지형 자료, 화학 반응식 등 다양한 입력 자료의 불확실성으로 인해 출력 결과에 오차가 발생할 수 있다.^{7),8),9)} 이러한 이유로, 예보 등급은 모델 자료의 수치를 그대로 활용하는 대신, 당일의 예측 담당자가 외부 기상 자

료와 실측 농도의 추세 등을 종합적으로 고려하여 결정하는 방식으로 운영되고 있다. 그러나 이러한 방식은 담당자의 경험과 주관적 판단에 크게 의존하기 때문에 예측의 정확도가 일정하지 않고 결정 과정이 오래 걸린다는 한계가 있다.

한편, 전산·통신 기술 등의 발전으로 대기오염물질에 대한 관측 기술이 고도화되고 관측 정보의 접근성과 제공 방식이 편리해지면서 대기오염물질 관측 자료와 머신러닝을 활용한 대기질 예측 연구가 활발히 진행되고 있다. 머신러닝은 주로 시간 단위 데이터를 예측하기 위해 사용하였는데, Samal et al.(2021)은 초미세먼지의 시간 관측 누락 데이터 처리와 예측을 통합적으로 수행하기 위해 인공신경망(Artificial Neural Network, ANN) 모델을 활용하였고 Wang et al.(2022)은 ANN 모델의 입력층과 출력층 사이에 여러 개의 은닉층을 추가하여 고도화시킨 심층신경망(Deep Neural Network, DNN) 모델을 활용하여 초미세먼지 농도를 예측하였다.^{10),11)} 여기서 시간 단위 데이터는 일 단위로 집계하였을 때 특정한 패턴을 보이는 경향이 있으며, 이러한 패턴을 분석하고 예측하기 위해 다양한 머신러닝 기법이 활용되었다. Kow et al.(2022)는 이미지 처리에 특화된 합성곱신경망(Convolutional Neural Network, CNN)을 활용해 (초)미세먼지와 대기질 지수(Air Quality Index, AQI)를 예측하였으며 Menares et al.(2021)는 시계열 데이터 처리에 특화된 순환신경망(Recurrent Neural Network, RNN)의 확장 형태인 LSTM(Long Short Term Memory network) 모델을 활용하여 초미세먼지의 일최고 농도를 예측하였다.^{12),13)} 또한 최근에는 인공위성에서 관측되는 대기 자료에 관한 연구도 진행하였는데, Teng et al.(2023)은 인공위성을 통해 관측되는 AOD(Aerosol Optical Depth) 데이터와 DNN 모델을 활용하여 72시간 초미세먼지 농도 예측하기도 하였다.¹⁴⁾

한편, 국립환경과학원은 2018년부터 머신러닝을 활용하여 (초)미세먼지와 오존에 대한 단기(3일) 및 중기(7일) 예보를 수행하는 연구를 진행하였다. 이 과정에서 DNN, CNN, LSTM 등의 모델을 적용해보면서 일 단위 예측에 적합한 모델을 도출하기 위한 다양한 시도가 이루어졌다.^{15),16)} 또한, 우리원에서는 2022년 부산의 4개 권역을 대상으로 초미세먼지 일평균 농도를 예측하는 DNN 모델을 연구하여, 대기오염 예보 등급 예측에 머신러닝 기법을 적용할 가능성을 검증하였으며, 그 결과 80.1%의 예측 적중률을 확인하였다.¹⁷⁾ 그

러나 농도를 예측한 후 이를 기반으로 등급을 산정하는 방식에서 예측된 농도가 실제 기준치 경계에 가까운 경우 수치적으로는 예측 결과가 정확하더라도 등급 산정에서는 오답으로 처리될 가능성이 있다. 이로 인해 농도 예측 결과와 등급 예보 결과 간의 불일치가 발생하고 이는 모델 평가 및 예보의 신뢰성을 저하시킬 수 있다고 판단하였다.

본 연구에서는 2022년에 수행된 선행 연구를 기반으로, 예측 대상 물질을 (초)미세먼지와 오존으로 확장함과 동시에 예측 기간을 기존의 익일(Day1)에서 모레(Day2)까지로 확대하였고 기존의 농도 예측 후 등급을 산정하는 방식에서 발생할 수 있는 불일치를 개선하기 위해, 등급 자체를 직접 예측하는 기법을 적용하였으며, 예측 항목과 모델 구조를 개선하여 DNN 모델의 신뢰성과 효율성을 강화하였다. 이러한 개선사항을 바탕으로 2021년부터 2023년까지의 데이터를 학습에 활용하였고, 이를 통해 2024년의 대기질 예측 결과를 도출하고 적중률과 실무 활용 가능성을 평가하였다.

2. 연구방법

2.1. 대기질 진단평가시스템을 활용한 대기질 예측

우리원에서는 2017년부터 운영 중인 대기질 진단평가시스템을 활용하여, 부산지역의 (초)미세먼지, 오존 예측 등급을 매일 오전과 오후 두 차례 발표하고 있다.

예측은 Fig. 1과 같이 CMAQ 모델 자료, 기상 자료, 과거 자료의 추세 분석을 기반으로 이루어지며, 당일 예측 담당자가 이를 종합적으로 검토하여 등급을 결정한다. 오전에는 당일(day0)과 익일(day1)의 대기질을, 오후에는 익일(day1)과 모레(day2)의 대기질을 예측하며, 오후 예측 결과를 기준으로 예측 적중률을 평가하고 있다. 부산지역은 지역 특성에 따라 동부(기장군), 서부(북구, 사하구, 강서구, 사상구), 남부(중구, 서구, 동구, 영도구, 남구, 해운대구, 수영구), 중부(금정구, 동래구, 부산진구, 연제구)의 4개 권역으로 구분되며, 예측 담당자는 각 권역별로 구분하여 대기질 등급을 발표한다.

대기질 진단평가시스템은 CMAQ 모델을 기반으로 실시간 모델링이 가능하도록 구축된 시스템이다. Fig. 2 및 Table 1과 같이, 시스템은 동아시아 27 km 격자에서 한반도 9 km 격자, 영남권 3 km 격자, 부산 1 km 격자로 부산을 중심으로 중첩된 4단계 도메인으로 구성된다. 여기서 Domain 4의 도시대기 측정소가 포함된 1km 격자의 자료를 대기질 예측에 활용하고 있다.

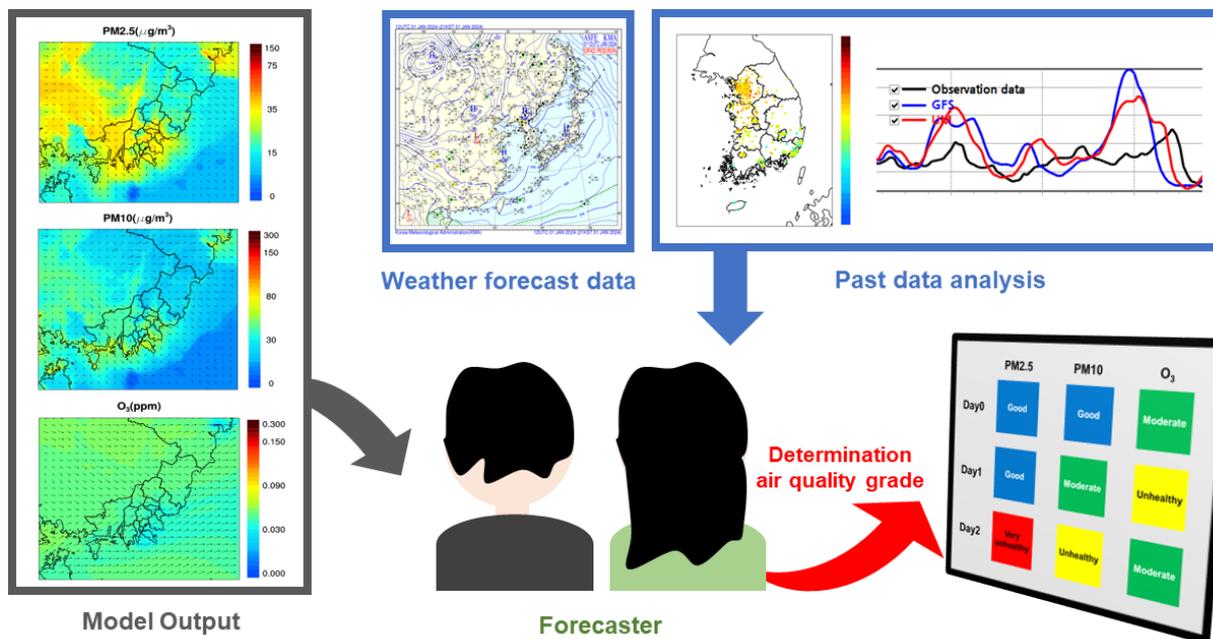


Fig. 1. Air quality forecasting process

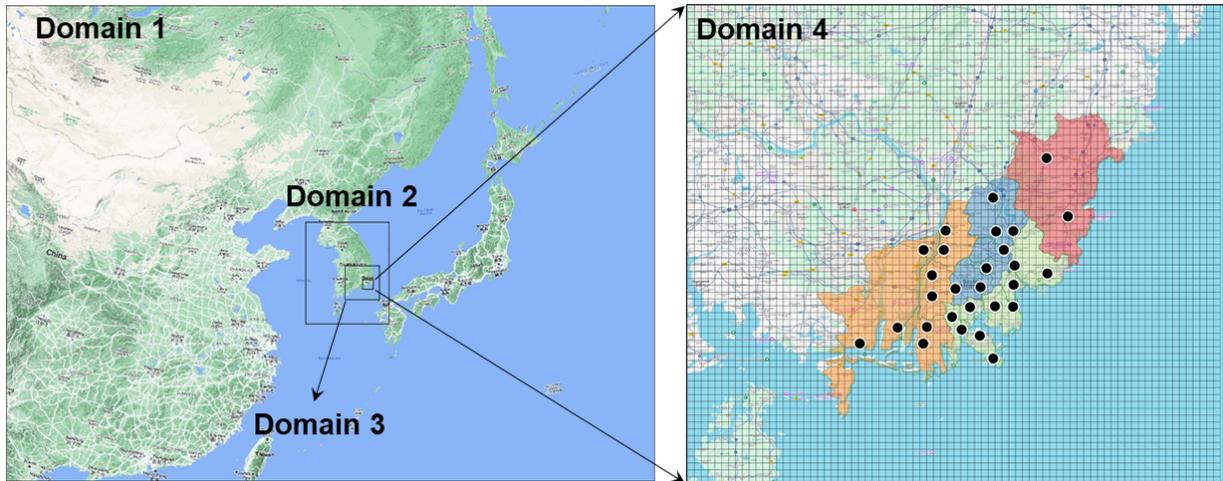


Fig. 2. Domain configuration for air quality modeling. Left: Domains 1 to 3. Right: Domain 4 with monitoring stations and geographical features.

Table 1. Grid configuration of CMAQ model

Configuration	Model domain	Grid information
Horizontal grids	Domain 1	174 × 128 × 27 km
	Domain 2	67 × 82 × 9 km
	Domain 3	83 × 83 × 3 km
	Domain 4	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

대기질 진단평가시스템은 Fig. 3과 같이 기상 입력 자료와 배출량 입력자료를 CMAQ 모델에 입력하여 시·공간적인 대기오염물질 농도를 산출한 뒤, 이를 대기질 예측과 평가에 활용하기 쉽도록 설계되었다. 기상 입력자료는 NCEP(National Center for Environmental Prediction)의 GFS(Global Forecast System)와 기

상청 국가 기상 슈퍼컴퓨터 센터의 지역 예보 모델인 UM(Unified Model)을 기반으로, 이를 WRF(Weather Research and Forecasting) 모델에 입력하고 있다.^{18),19)} GFS는 전구 예보모델이며, UM은 수평 약 12 km, 연직 약 80 km 해상도를 가지는 지역 예보모델로, 두 모델 모두 하루 4회(00UTC, 06UTC, 12UTC,

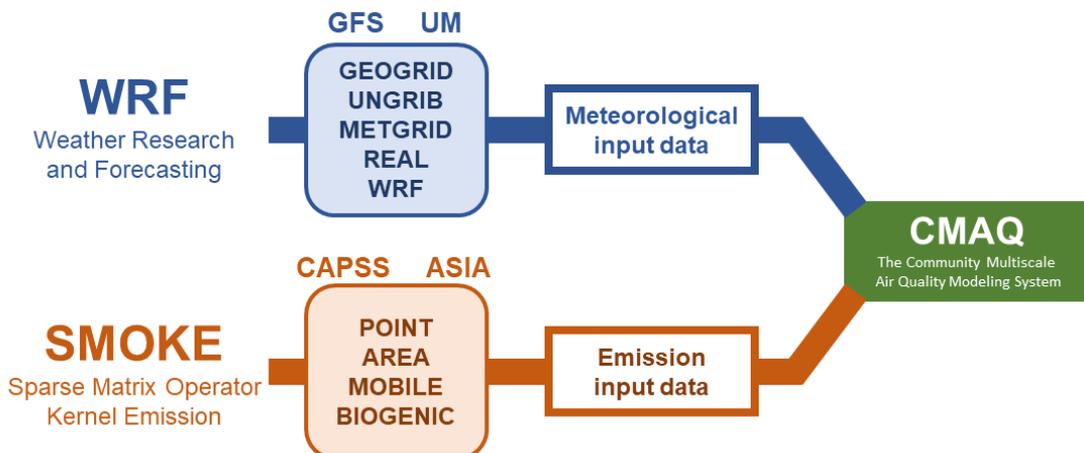


Fig. 3. CMAQ modeling system: meteorological and emission data workflow

18UTC) 수행된다. 이 시스템에서는 06UTC의 모델 자료를 활용하고 있다. WRF 모델은 NCAR(National Center for Atmospheric Research)와 NCEP가 공동으로 개발한 기상 모델로, 전 세계적으로 널리 사용되고 있다.²⁰⁾

배출량 입력자료는 국내 배출량의 경우 환경부 국가미세먼지정보센터에서 제공하는 대기정책지원시스템(Clean Air Policy Support System, CAPSS) 데이터를 사용하며, 배출원별로 1km 격자 단위로 정리된 배출량이 입력된다. 중국의 배출량은 MEIC(Multi-resolution Emission Inventory for China), 중국 외 아시아 지역은 REAS(Regional Emission inventory in Asia) 데이터를 기반으로 하며, 이 자료들은 SMOKE(Sparse Matrix Operator Kernel Emissions) 모델에 입력되어 처리된다. 대기질 진단평가시스템의 세부 모델링 옵션은 Table 2와 같다.

2.2. DNN 모델 구축

심층신경망(Deep Neural Network, DNN)은 인간의 뇌 구조, 특히 뉴런 간의 상호작용에서 영감을 받아 개발된 인공신경망(Artificial Neural Network, ANN)

의 확장된 모델이다. Fig. 2의 왼쪽에서와 같이 ANN의 구조는 입력값(Input, x_1, x_2)과 가중치(w_1, w_2)의 가중 합에 의해 결과가 출력값(Output, y)으로 전달되는 방식으로 작동한다. 이 과정에서 활성화 함수(Activation function)는 입력값의 가중 합을 비선형적으로 변환하여, 뉴런 간의 복잡한 상호작용을 모방하고 출력 노드로 전달되는 값을 결정하는 역할을 한다. 뇌의 뉴런은 시냅스를 통해 신호를 전달하며, 신경망에서도 각 노드가 연결을 통해 데이터를 전달하고 학습을 수행한다. 다음 식은 신경망에서 하나의 노드가 데이터를 처리하는 과정을 수학적으로 나타낸 것이다. 입력 값(x_i)과 가중치(w_i)의 가중 합에 편향(b)을 더한 결과를 활성화 함수(f)에 적용하여 출력값(y)을 계산한다.

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right)$$

활성화 함수로는 이진 분류에 적합한 Sigmoid, 회귀 분석에 적합한 ReLU와 Linear, 다중 클래스 분류에 적합한 Softmax 등이 사용된다.²¹⁾ 각 함수는 데이터의 특성과 문제 유형에 따라 선택되며, 신경망의 학

Table 2. WRF and CMAQ model configuration options

WRF option	CMAQ option
Microphysics option : WSM 6-calss graupel	Horizontal advection : YAMO
Long wave radiation : RRTM	Vertical advection : WRF
Short wave radiation : Goddard	Horizontal diffusion : Multi-scale
Surface layer scheme : MM5 similarity	Vertical diffusion : Eddy
Land surface scheme : Noah Land Surface Model	Gas-phase chemistry : CB5
PBL scheme : YSU	Aerosol chemistry : AE5
Cumulus parameterization : Kain-Fritsch	Dry deposition : M3Dry

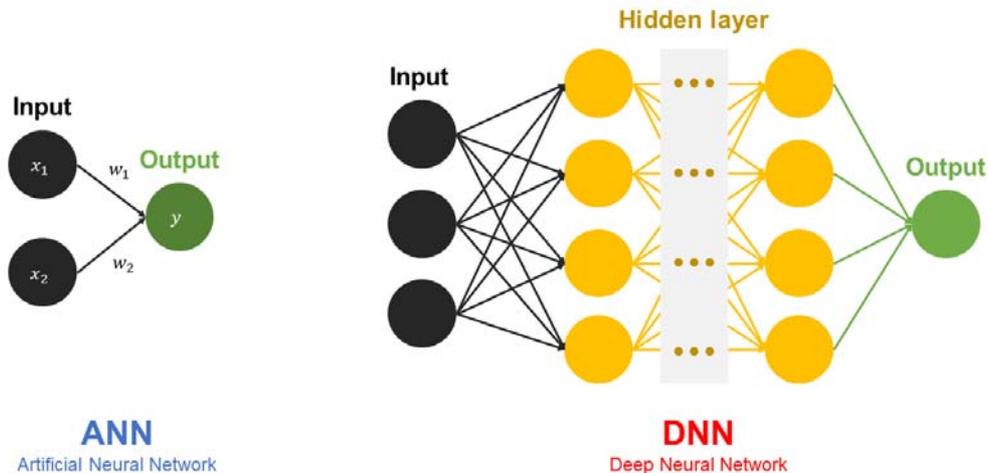


Fig. 4. Structures of Artificial Neural Network and Deep Neural Network

습 성능과 출력 값의 해석에 중요한 영향을 미친다. DNN 모델은 Fig. 2의 오른쪽에서와 같이 ANN 모델을 확장하여 다수의 은닉층(hidden layer)을 포함한 구조로 설계된 모델이다. 이러한 다층 구조를 통해 DNN은 보다 복잡한 데이터 패턴을 학습할 수 있는 능력을 갖추고 있다.

본 연구에서는 2022년 우리원에서 DNN 모델을 기반으로 익일(Day1) PM_{2.5} 농도를 예측한 선행 연구를 바탕으로, 모델의 예측 기간(Day1~Day2)과 예측 항목(PM_{2.5}, PM₁₀, O₃)을 확장하고, DNN 모델의 신뢰성과 효율성을 강화하기 위한 시도를 하였다. 기존과 같이 도시대기 측정소가 위치한 격자의 일평균 데이터를 기반으로 입력 자료를 생성하되, 유사한 농도 자료가 다수 포함될 때 발생할 수 있는 혼란을 줄이기 위해 측정소별 일평균 데이터를 재가공하여 권역별 일평균 자료를 사용하는 방식으로 변경하였다. 여기에 필요에

따라 권역별 일최고 데이터를 활용하여 모델의 신뢰성을 평가하기도 하였다. 모델의 입력 인자는 Table 3과 같으며, 기상요소 관련 변수는 기존과 동일하게 WRF 모델을 활용하였다.

각 권역별로 총 운량(CFRAC), 경계층 높이(PBL), 지상기압(PRSC), 혼합비(Q2), 기온(TEMP2)를 사용하였으며, 기존의 풍속 자료는 풍향 자료와 함께 x, y 성분으로 분리한 형태인 W10X와 W10Y로 재가공하여 활용하였다. 또한, 오존 예측 변수를 보강하기 위해 대기 흡수 태양복사량(GSW) 및 지표 도달 태양복사량(RGRND)를 변수로 추가하였다. 대기오염물질(CO, NO, NO₂, NO_x, O₃, PM_{2.5}, PM₁₀, SO₂, VOCs) 농도 입력 자료는 기존의 전일 실측 평균 자료를 활용하는 방식에서, 예측 시점에 맞는 자료를 활용하기 위해 CMAQ 모델의 출력 예측 자료로 변경되었다. 또한, 실제 데이터와 예측 자료 간의 추세를 반영하기 위해 예

Table 3. Input factors for air quality grade prediction models.

Variable type	Variable	Description
Meteorological elements	CFRAC	Total cloud fraction at monitoring site (predicted value)
	PBL	Planetary boundary level height at monitoring site (predicted value)
	PRSC	Surface pressure at monitoring site (predicted value)
	Q2	Mixing ratio at monitoring site (predicted value)
	TEMP2	Temperature at monitoring site (predicted value)
	W10X	Wind x-component at monitoring site(predicted value)
	W10Y	Wind y-component at monitoring site(predicted value)
	GSW (O ₃ only)	Solar radiation absorbed at monitoring site(predicted value)
	RGRND (O ₃ only)	Solar radiation reaching the surface at monitoring site(predicted value)
Air pollutant concentration	CO	Carbon monoxide(CO) concentration at monitoring site(predicted value)
	NO	Nitric oxide(NO) concentration at monitoring site(predicted value)
	NO ₂	Nitrogen dioxide(NO ₂) concentration at monitoring site(predicted value)
	NO _x	Nitrogen Oxides(NO _x) concentration at monitoring site(predicted value)
	O ₃	Ozone(O ₃) concentration at monitoring site(predicted value)
	PM _{2.5}	PM _{2.5} concentration at monitoring site(predicted value)
	PM ₁₀	PM ₁₀ concentration at monitoring site(predicted value)
	SO ₂	Sulfur dioxide(SO ₂) concentration at monitoring site(predicted value)
Adjustment factor	VOC	Volatile organic compounds(VOCs) concentration at monitoring site(predicted value)
	PM _{2.5} diff (PM _{2.5} only)	Difference between the modled and observed PM _{2.5} value
	PM ₁₀ diff (PM ₁₀ only)	Difference between the modled and observed PM ₁₀ value
	O ₃ diff (O ₃ only)	Difference between the modled and observed O ₃ value
	PM _{2.5} MODEL (PM _{2.5} only)	PM _{2.5} grade derived from CMAQ model predictions
PM ₁₀ MODEL (PM ₁₀ only)	PM ₁₀ grade derived from CMAQ model predictions	
O ₃ MODEL (O ₃ only)	O ₃ grade derived from CMAQ model predictions	
Temporal information	Month	Calendar month (1 to 12) indicating the monitoring time.
	Day of week	Day of the week (0 to 6), reflecting weekly variations in air quality.

측 항목에 대해서는 전일 관측 자료와 예측 자료 간의 수치 차이를 변수화하여 보정 자료로 활용하였다.

추가로, CMAQ 결과를 예보 등급에 맞게 변환하여 보정 자료로 추가적으로 반영하였다. 변수 선정에 관한 예비 분석 과정에서, 인접 지역의 농도를 입력 변수로 포함할 경우 간섭 현상이 발생하여 예측 정확도가 저하되는 문제가 확인되었다. 이에 따라 울산, 경상남도 등 인접 지역의 농도 자료는 제외하였다. 또한, 각 권역별 대기질 등급 예측의 정확성을 높이기 위해, 권역별로 입력 변수를 개별적으로 분리하여 사용하였다. 국가 배경농도 자료는 측정소별 데이터 누락이 잦아 일관성 있는 자료 확보가 어려웠으며, 예측의 신뢰성을 저하시킬 우려가 있어 제외하였다. 추가로, 월(Month)과 요일(Day of Week) 변수를 도입하여 계절적 요인과 주간 주기의 영향을 모델에 반영함으로써 예측의 정밀도를 향상시켰다. 마지막으로, 기존의 농도 예측 후 등급을 산정하는 방식에서 발생할 수 있는 불일치를 보완하기 위해, 출력 변수를 대기질 예측 등급으로 변경하였다.

DNN 모델의 학습 자료는 대기질 진단평가시스템에서 활용하는 자료와 동일하게, 전일 06UTC(21KST) 기준으로 생성된 GFS와 UM 자료의 기상 인자와 이를 이용해 생성된 CMAQ 모델 결과를 기반으로 구성되었다. Fig. 5에 나타난 바와 같이, 매일 06KST에 GFS 95시간 자료와 UM 87시간 자료가 생성되며, 이를 바탕으로 익일(Day1)과 모레(Day2)의 입력 인자를 추출하였다. DNN 학습을 위하여 2021년~2023년의 데이터를 활용하였고 학습 데이터의 20%는 모델 검증에 사용하였으며 2024년 1월~10월 자료를 이용하여 대기질 등급 예측 적중률을 평가하였다.

모델의 세부 파라미터는 3~7개의 은닉층을 탐색하며 최적 구성을 도출하였고, 각 은닉층의 노드 수는 32에서 2048 사이를 32 단위로 설정하여 하이퍼파라미터 튜닝을 수행하였다. 활성화 함수로는 ReLU를 사용하

였으며, 출력층에서는 다중 클래스 분류에 적합한 Softmax 함수를 적용하였다. 손실 함수로는 다중 클래스 분류에 적합한 Categorical Crossentropy를 사용하였으며, 최적화 알고리즘으로는 Adam을 채택하였다. 학습 과정은 Batch Size를 32로 고정하고, 최대 50회 Epoch으로 설정하여 진행되었으며, Keras Tuner의 Hyperband 알고리즘을 이용해 최적의 모델 구성을 도출하였다. 황사 일자는 CMAQ 모델 데이터에 포함되지 않아 학습 데이터에서 제외하였으며, 학습 데이터가 대기오염물질에 따라 특정 등급으로 편향되어 있는 측면을 보완하여 학습하기 위해 등급별 가중치를 적용하는 방법을 병행하였다. 모델 구현 및 학습에는 TensorFlow 2.17.0, Keras 3.5.0, NumPy 1.23.5, Pandas 1.4.2, Scikit-learn 1.0.2, 그리고 Keras Tuner 1.4.7 등의 라이브러리를 활용하였다.

3. 연구결과

3.1. DNN 모델 학습 데이터셋 분석

훈련 데이터로 사용한 2021~2023년의 대기질 관측 등급은 Table 4와 같다. (초)미세먼지는 전체적으로 “좋음”>“보통”>“나쁨”>“매우나쁨” 순으로 관측 빈도가 높았으며, 오존의 경우 “보통”>“나쁨”>“좋음”>“매우나쁨” 순으로 관측 빈도가 높게 나타났다. 이 중 황사 일자와 데이터 수집 누락 일자를 포함하여, GFS 기반 훈련 데이터는 각 권역에서 총 105일(2021년 25일, 2022년 71일, 2023년 9일)의 데이터를 제외하였다. 또한, UM 기반 훈련 데이터는 각 권역에서 총 106일(2021년 24일, 2022년 73일, 2023년 9일)의 데이터를 제외하였다.

한편, (초)미세먼지는 “좋음”등급으로, 오존은 “보통”등급으로 CMAQ 모델의 과소 모의 등의 영향으로 데이터가 편향되어 있어, 학습이 데이터가 많은 등급으로 치우칠 가능성이 있다. 그러나 Fig. 6에서 제시된 월별 관측 자료를 비교한 결과, 관측 빈도가 적은 “나

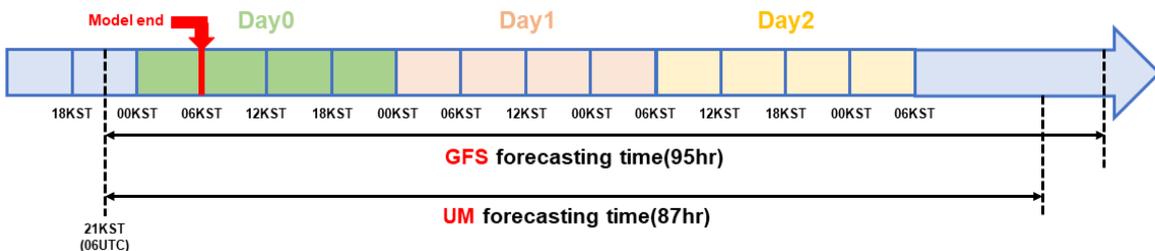


Fig. 5. Forecasting periods and timescales of GFS and UM for Day1 and Day2 Prediction

Table 4. Observation frequency of PM₁₀, PM_{2.5}, and O₃ quality Grades (2021-2023)

Air pollutant	Year	Observation frequency (East / West / South / Central)			
		Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5}	Total	2641 (671 / 558 / 745 / 667)	1581 (389 / 479 / 322 / 391)	154 (34 / 57 / 27 / 36)	4 (1 / 1 / 1 / 1)
	2021	911 (232 / 190 / 252 / 237)	513 (127 / 158 / 108 / 120)	32 (5 / 16 / 4 / 7)	4 (1 / 1 / 1 / 1)
	2022	886 (219 / 184 / 257 / 226)	520 (134 / 162 / 99 / 125)	54 (12 / 19 / 9 / 14)	0
	2023	844 (220 / 184 / 236 / 204)	548 (128 / 159 / 115 / 146)	68 (17 / 22 / 14 / 15)	0
PM ₁₀	Total	2944 (803 / 657 / 748 / 736)	1336 (268 / 411 / 322 / 335)	76 (18 / 21 / 19 / 18)	24 (6 / 6 / 6 / 6)
	2021	972 (261 / 219 / 248 / 244)	460 (97 / 139 / 110 / 114)	16 (4 / 4 / 4 / 4)	12 (3 / 3 / 3 / 3)
	2022	1025 (282 / 227 / 258 / 258)	423 (80 / 134 / 104 / 105)	12 (3 / 4 / 3 / 2)	0
	2023	947 (260 / 211 / 242 / 234)	453 (91 / 138 / 108 / 116)	48 (11 / 13 / 12 / 12)	12 (3 / 3 / 3 / 3)
O ₃	Total	46 (5 / 14 / 6 / 21)	4022 (1003 / 1004 / 1013 / 1002)	312 (87 / 77 / 76 / 72)	0
	2021	8 (0 / 2 / 1 / 5)	1367 (344 / 342 / 346 / 335)	85 (21 / 21 / 18 / 25)	0
	2022	26 (4 / 9 / 4 / 9)	1308 (327 / 324 / 331 / 326)	126 (34 / 32 / 30 / 30)	0
	2023	12 (1 / 3 / 1 / 7)	1347 (332 / 338 / 336 / 341)	101 (32 / 24 / 28 / 17)	0

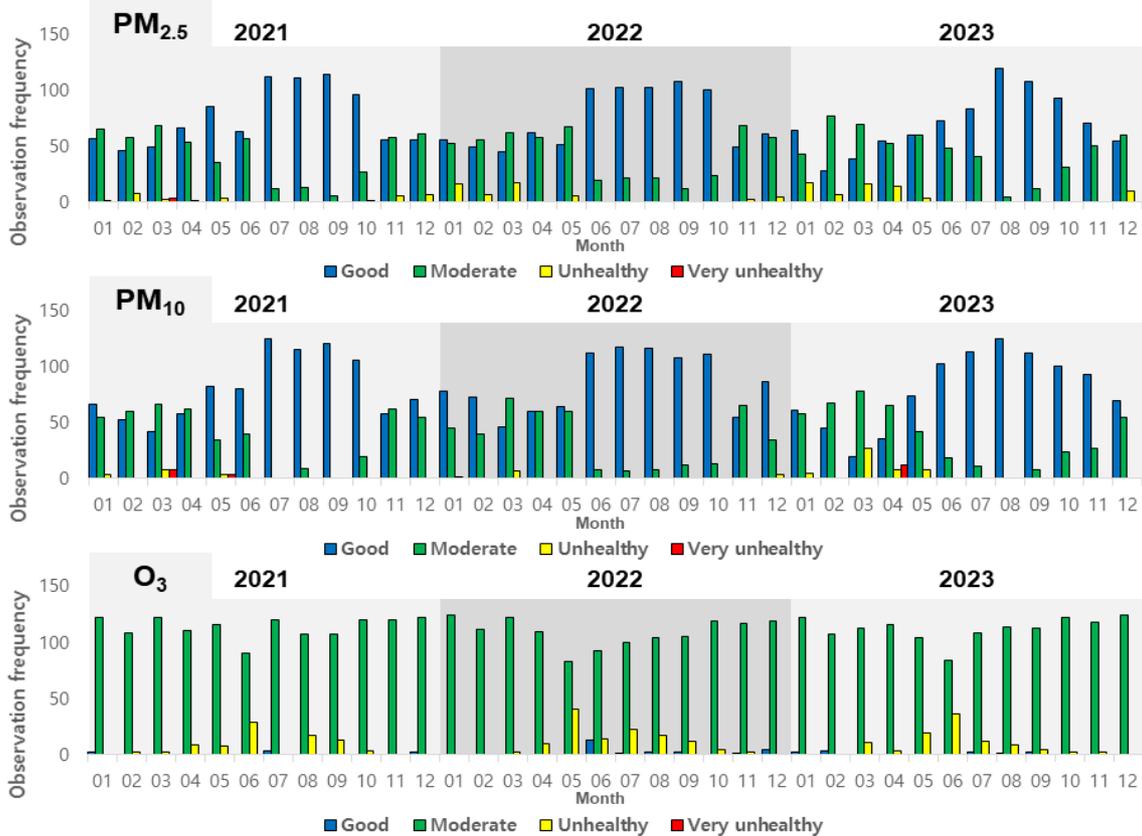


Fig. 6. Monthly observation frequency of PM₁₀, PM_{2.5}, and O₃ quality Grades (2021-2023)

뽐” 등급의 데이터가 특정 월에 집중되어 나타나는 경 월 데이터를 통해 학습 편향성을 일부 완화할 수 있을
향을 확인할 수 있었다. 이에 따라 모델 학습에 포함된 것으로 기대하였다.

Table 5. Observation frequency of PM₁₀, PM_{2.5}, and O₃ Quality Grades (January–October 2024)

Air pollutant	Year	Observation frequency (East / West / South / Central)			
		Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5}	2024.	811	385	24	0
	(JAN-OCT)	(213 / 185 / 218 / 195)	(87 / 112 / 82 / 104)	(5 / 8 / 5 / 6)	
PM ₁₀	2024.	931	273	4	12
	(JAN-OCT)	(243 / 214 / 242 / 232)	(58 / 88 / 58 / 69)	(1 / 0 / 2 / 1)	(3 / 3 / 3 / 3)
O ₃	2024.	8	1083	128	1
	(JAN-OCT)	(3 / 3 / 0 / 2)	(265 / 266 / 271 / 281)	(37 / 36 / 33 / 22)	(0 / 0 / 1 / 0)

Table 6. Prediction Accuracy of Forecasters and CMAQ Model Outputs (Day1 and Day2)

Air pollutant	Prediction source	Prediction accuracy(%) (East / West / South / Central)				
		Total	Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5} (Day1)	Forecaster	83.0 (86.7 / 78.8 / 84.2 / 82.3)	92.7 (94.0 / 92.1 / 91.3 / 93.3)	61.9 (66.7 / 57.9 / 64.2 / 60.9)	0 (- / 0 / - / -)	-
	CMAQ (GFS)	78.4 (79.5 / 73.5 / 83.2 / 77.5)	96.5 (97.2 / 97.8 / 95.4 / 95.9)	40.9 (36.9 / 33.3 / 51.9 / 43.6)	25.0 (0 / 40.0 / 0 / 33.3)	-
	CMAQ (UM)	77.3 (79.9 / 71.1 / 82.2 / 75.8)	94.6 (95.3 / 95.1 / 93.5 / 94.3)	40.6 (40.5 / 31.5 / 51.9 / 41.6)	50.0 (100 / 40.0 / 50.0 / 33.3)	-
PM ₁₀ (Day1)	Forecaster	86.9 (87.7 / 84.2 / 88.2 / 87.7)	94.6 (94.5 / 95.1 / 93.3 / 95.5)	61.1 (59.0 / 57.6 / 67.5 / 61.7)	-	-
	CMAQ (GFS)	83.1 (83.6 / 80.2 / 86.6 / 82.2)	97.5 (99.2 / 98.6 / 95.0 / 97.4)	32.1 (16.1 / 33.3 / 50.0 / 28.8)	-	-
	CMAQ (UM)	83.7 (84.9 / 80.5 / 85.2 / 84.2)	97.1 (99.2 / 98.1 / 93.8 / 97.8)	36.3 (23.2 / 35.7 / 48.2 / 37.9)	-	-
O ₃ (Day1)	Forecaster	91.1 (90.6 / 89.7 / 92.6 / 91.6)	0 (0 / 0 / - / 0)	96.5 (97.7 / 95.4 / 97.8 / 95.1)	58.7 (58.6 / 60.0 / 54.5 / 62.5)	0 (- / - / 0 / -)
	CMAQ (GFS)	89.9 (88.9 / 89.9 / 88.6 / 92.3)	0 (0 / 0 / - / 0)	96.6 (96.9 / 96.9 / 96.2 / 96.4)	40.5 (40.5 / 45.7 / 28.1 / 50.0)	0 (- / - / 0 / -)
	CMAQ (UM)	88.2 (86.6 / 87.9 / 88.6 / 89.6)	0 (0 / 0 / - / 0)	94.4 (95.3 / 93.5 / 95.1 / 93.8)	42.1 (32.4 / 54.3 / 37.5 / 45.5)	0 (- / - / 0 / -)
PM _{2.5} (Day2)	Forecaster	81.3 (84.2 / 77.7 / 83.2 / 80.2)	93.6 (93.2 / 96.0 / 91.2 / 94.7)	54.8 (59.3 / 49.3 / 60.4 / 52.9)	57.1 (100 / 33.3 / 100 / 50.0)	-
	CMAQ (GFS)	77.3 (78.2 / 73.2 / 81.5 / 76.2)	96.8 (97.2 / 97.8 / 95.0 / 97.4)	35.6 (31.3 / 31.8 / 46.2 / 35.0)	33.3 (0 / 40.0 / 0 / 66.7)	-
	CMAQ (UM)	76.8 (79.5 / 72.8 / 79.2 / 74.8)	93.7 (94.8 / 94.6 / 91.3 / 94.4)	39.9 (41.0 / 36.4 / 46.2 / 38.0)	41.7 (50.0 / 40.0 / 50.0 / 33.3)	-
PM ₁₀ (Day2)	Forecaster	86.0 (86.6 / 82.7 / 88.1 / 86.6)	91.8 (92.9 / 91.9 / 90.5 / 92.0)	62.7 (54.5 / 57.4 / 76.5 / 65.0)	-	-
	CMAQ (GFS)	84.1 (83.9 / 80.5 / 88.6 / 83.2)	98.8 (99.6 / 99.1 / 97.5 / 99.1)	31.4 (16.1 / 32.5 / 50.0 / 27.3)	-	-
	CMAQ (UM)	83.6 (84.2 / 79.9 / 86.9 / 83.6)	97.5 (99.6 / 97.7 / 95.0 / 97.8)	34.1 (17.9 / 33.7 / 51.8 / 33.3)	-	-
O ₃ (Day2)	Forecaster	90.8 (87.6 / 90.1 / 93.1 / 92.6)	0 (0 / 0 / - / 0)	96.6 (95.4 / 97.1 / 98.9 / 95.2)	51.1 (44.4 / 50.0 / 52.2 / 64.3)	0 (- / - / 0 / -)
	CMAQ (GFS)	89.1 (87.9 / 87.9 / 88.6 / 91.9)	12.5 (33.3 / 0 / - / 0)	97.2 (97.3 / 97.3 / 97.0 / 97.1)	27.0 (27.0 / 25.7 / 21.9 / 36.4)	0 (- / - / 0 / -)
	CMAQ (UM)	87.8 (85.6 / 86.9 / 89.3 / 89.6)	0 (0 / 0 / - / 0)	94.5 (94.2 / 93.5 / 95.8 / 94.5)	38.1 (32.4 / 45.7 / 37.5 / 36.4)	0 (- / - / 0 / -)

value : Highest value

3.2. DNN 모델 평가 데이터셋 분석

평가 데이터로 사용한 2024년 1월~10월의 대기질 관측 등급은 Table 5에 제시되어 있다. 학습 데이터와 유사하게, (초)미세먼지는 “좋음” 등급, 오존은 “보통” 등급으로 편향되어 있었다. 이 중 황사 일자와 데이터 수집 누락 일자를 포함하여, GFS 기반 평가 데이터는 총 8일, UM 기반 평가 데이터는 7일의 데이터를 제외하였다. DNN 모델의 비교 지표로 사용된 평가 데이터와 동일한 일자의 예측 담당자 예측 결과 및 CMAQ 모델 결과는 Table 6에 나타냈다.

CMAQ 모델의 예측 적중률은 편향성이 높은 등급에서 높은 값을 보였으며, GFS의 미세먼지를 제외하면 전체적으로 익일(Day1) 예측이 더 높은 적중률을 기록하였다. 특히 편향성이 낮은 등급에서도 익일(Day1) 예측이 상대적으로 우수한 적중률을 보였다. 데이터의 편향성에서 차이가 존재하지만, 익일(Day1)의 미세먼지를 제외하면 GFS 모델을 기반으로 한 CMAQ 모델의 예측 적중률이 전반적으로 높은 것으로 평가되었다. 특히, 예측 담당자가 CMAQ 모델의 결과에 다양한 외부 변수를 추가적으로 고려한 경우, 편향성이 낮은 등급에서의 예측 적중률이 크게 향상되었으며, 전체적인 예측 적중률 또한 개선되는 것을 확인하였다.

한편, 예측 담당자의 초미세먼지 “나쁨” 등급 적중률이 0%로 나타난 것은 관측 데이터와 예보 업무 일정의 제한에 따른 결과이다. 올해 초미세먼지 “나쁨” 등급은 총 24회 관측되었으나, 황사 일자가 CMAQ 모델에서 제외됨에 따라 관측 일수는 12일로 감소하였다. 이 중 예보 담당자가 업무를 수행하지 않는 휴일을 제외하면 실제 예측 가능한 날은 단 1일에 불과했다. 결과적으로, 예측이 실제 업무에 반영될 기회 자체가 극히 제한적이었으며, 이는 예측 적중률 저조의 주요 원인으로 작용하였다. 이를 바탕으로, DNN 모델이 예측 담당자 수준 혹은 그 이상의 결과를 얻기 위해서는, 편향성이 낮은 데이터의 적중률을 개선해 나가면서 전체적인 예측 정확도를 높이는 것과 CMAQ 모델 결과 및 다양한 인자를 함께 고려하여 이러한 한계를 극복하는 것이 중요하다고 판단되었다.

3.3. DNN 모델의 학습 성능을 위한 최적 은닉층 설계

DNN 모델은 선행 연구와 동일하게 입력 변수를 일 평균 형태로 구성하고, 최적의 은닉층 설계를 위해 은닉층의 개수를 3개에서 7개까지 조정하여 학습을 수행하였다. 그 결과는 Table 7과 Table 8에 나타냈으며,

Table 7은 GFS 데이터를 기반으로 한 결과이고 Table 8은 UM 데이터를 기반으로 한 결과이다.

은닉층의 개수를 변화시킨 결과, 학습 데이터의 20%를 사용한 검증 데이터의 정확도와 테스트 데이터의 변화폭은 크지 않았으며, 전반적으로 검증 데이터로의 과적합 현상 없이 안정적이고 일관된 성능을 갖춘 모델을 도출할 수 있었다. 모델별로 비교한 결과, (초)미세먼지의 경우 UM이 전체 예측 적중률이 높았으며, 편향성이 큰 “좋음” 등급을 제외한 다른 등급에서도 높은 적중률을 보여 데이터 편향성에 대한 대응 능력이 뛰어난 것으로 나타났다.

미세먼지의 경우 전체 예측 적중률은 GFS와 UM이 유사하였으나, “좋음” 등급 외 적중률은 UM이 상대적으로 우수하였다. 반면, 오존의 경우 GFS가 전체 예측 적중률과 편향성이 낮은 “보통” 등급 외 적중률에서 더 우수한 성능을 보였다. 일자별로 분석한 결과, 대체적으로 내일(Day1)의 예측 적중률이 모레(Day2)보다 높게 나타났으나, GFS 기반의 (초)미세먼지의 일부 항목에서는 모레(Day2)의 예측 적중률이 내일(Day1)과 유사하거나 더 높은 경향도 확인되었다. 다만 권역별 세부 분석에서는 일부 항목에서 편차가 크게 나타났으며, (초)미세먼지의 “좋음” 항목과 오존의 “보통” 항목처럼 편향성이 큰 등급에 예측이 집중되거나, 반대로 편향성을 극복하며 예측 적중률이 향상되는 양상이 모델의 조건과 항목에 따라 다르게 나타났다.

이러한 결과를 바탕으로, 각각의 항목에 대해 최적의 은닉층 개수를 별도로 산정할 경우 데이터의 상관성이나 조건에 따라 일시적으로 우수한 결과가 도출될 수 있어 결과론적 접근에 그칠 우려가 있었다. 따라서 전체 항목을 고려하면서도 예측 적중률이 높고 편향성이 큰 데이터에 대한 대응력을 확보할 수 있는 은닉층 개수를 대표값으로 선정하는 것이 필요하였다.

전체적인 예측 적중률을 비교한 결과, 은닉층 개수가 4개인 경우가 5회로 가장 높았고, 그다음으로는 5개(3회), 3개·6개·7개(각 2회) 순으로 나타났다. 또한, 편향성이 큰 (초)미세먼지 “좋음” 항목과 오존 “보통” 항목을 제외한 등급별 예측 적중률에서는 은닉층 4개가 6회로 가장 우수하였으며 이 외 5개와 7개(각 5회), 6개(2회), 3개(1회) 순으로 나타났다. 권역별 세부 분석에서도 전체 예측 적중률은 은닉층 개수 4개와 7개가 각각 12회로 가장 높았고, 그 다음으로는 6개(10회), 3개와 5개(각 9회) 순으로 나타났다. 편향성이 큰 등급을 제외한 등급별 예측 적중률은 은닉층 4개가 17

회로 가장 높았으며, 그 뒤를 이어 7개(15회), 5개(14회), 6개(13회), 3개(12회) 순으로 나타났다.

이러한 분석 결과를 종합하면, 은닉층 개수 4개가 전체적인 예측 적중률과 편향성을 극복하는 측면에서

가장 우수한 성능을 보였으며, 이를 바탕으로 대표성을 가지는 최적의 은닉층 구조로 4개를 선정하였다.

Table 7. Prediction Accuracy of DNN Model Using GFS-Based Data by Number of Hidden Layers

Air pollutant	Number of hidden layer	Validation accuracy (%)	DNN-GFS, Prediction accuracy(%) (East / West / South / Central)				
			Total	Good	Moderate	Unhealthy	Very unhealthy
			PM _{2.5} (Day1)	3	85.2	79.0 (81.8/71.7/82.2/80.5)	95.0 (97.2/90.8/95.4/96.4)
4	85.4	79.7 (84.5/77.1/80.1/77.1)		94.8 (95.7/90.2/94.0/99.0)	48.9 (58.3/56.5/43.0/37.6)	25.0 (0/40.0/50.0/0)	-
5	85.2	79.2 (81.8/71.4/82.5/75.1)		96.3 (96.2/92.4/97.7/98.4)	44.6 (47.6/55.6/43.0/31.7)	8.3 (0/0/0/33.3)	-
6	85.4	80.0 (82.2/78.8/80.1/78.8)		95.0 (96.2/89.1/95.8/98.4)	49.5 (48.8/63.9/38.0/43.6)	16.7 (0/20.0/50.0/0)	-
7	85.4	79.6 (80.8/75.4/83.2/79.1)		96.1 (95.7/93.5/96.8/98.4)	46.5 (45.2/48.1/48.1/44.6)	0 (0/0/0/0)	-
PM ₁₀ (Day1)	3	86.9	83.9 (83.8/81.5/85.9/84.5)	95.0 (97.5/91.5/93.4/97.4)	44.7 (25.0/56.0/53.6/39.4)	-	-
	4	87.0	84.8 (84.8/81.5/87.2/85.5)	95.9 (97.9/94.4/94.2/97.0)	45.4 (28.6/48.8/57.1/45.5)	-	-
	5	86.7	84.8 (84.2/84.8/85.2/84.8)	96.1 (98.3/95.3/93.8/97.0)	44.7 (23.2/53.3/48.2/42.4)	-	-
	6	86.5	83.2 (85.5/80.8/84.2/82.2)	94.4 (96.7/92.5/90.9/97.4)	43.5 (37.8/44.1/12.5/45.5)	-	-
	7	86.7	84.8 (85.9/82.8/85.5/84.8)	95.9 (97.9/93.4/95.4/96.5)	45.4 (33.9/56.0/42.9/43.9)	-	-
O ₃ (Day1)	3	94.9	90.6 (89.2/90.6/90.6/91.9)	0 (0/0/-/0)	98.0 (98.8/97.3/99.2/96.7)	34.4 (29.7/47.1/21.9/40.9)	0 (-/-/0/-)
	4	94.9	89.8 (89.6/91.2/90.6/87.9)	0 (0/0/-/0)	96.7 (98.4/98.8/99.2/90.5)	38.4 (35.1/41.2/21.9/63.6)	0 (-/-/0/-)
	5	94.6	91.5 (90.9/91.2/90.6/93.3)	0 (0/0/-/0)	97.9 (97.3/99.2/99.2/96.0)	44.0 (54.1/38.2/21.9/68.2)	0 (-/-/0/-)
	6	95.1	89.5 (90.9/88.9/89.9/88.2)	0 (0/0/-/0)	96.8 (99.6/95.8/99.6/92.3)	34.4 (37.8/44.1/12.5/45.5)	0 (-/-/0/-)
	7	94.6	90.7 (91.2/90.6/89.2/91.9)	0 (0/0/-/0)	99.0 (98.4/98.8/99.6/98.9)	28.0 (48.6/35.3/6.3/13.6)	0 (-/-/0/-)
PM _{2.5} (Day2)	3	81.6	80.6 (80.5/77.1/83.5/81.5)	95.3 (96.2/95.2/95.4/94.4)	50.5 (42.2/48.1/52.6/58.6)	8.3 (0/20.0/0/0)	-
	4	81.8	79.5 (80.1/75.1/82.5/80.1)	94.4 (95.8/95.7/92.6/93.8)	48.9 (42.2/42.5/56.4/55.6)	0 (0/0/0/0)	-
	5	81.8	79.3 (80.1/75.4/80.8/80.8)	92.8 (94.3/90.3/93.1/93.3)	51.9 (45.8/52.8/48.7/58.6)	0 (0/0/0/0)	-
	6	82.3	79.9 (81.5/76.4/81.5/80.1)	95.6 (95.3/93.5/97.2/95.9)	47.8 (48.2/50.0/39.7/51.5)	0 (0/0/0/0)	-
	7	82.1	80.1 (80.5/78.8/82.5/78.5)	95.3 (93.9/93.5/97.2/96.4)	48.6 (48.2/55.7/43.6/45.5)	8.3 (0/20.0/0/0)	-
PM ₁₀ (Day2)	3	86.1	85.6 (87.5/83.5/84.5/86.9)	93.2 (92.9/94.4/90.9/94.8)	58.5 (64.3/54.9/57.1/59.1)	-	-
	4	86.2	84.3 (84.2/83.2/85.5/84.5)	94.9 (98.8/92.6/91.7/96.5)	46.5 (21.4/58.5/58.9/42.4)	-	-
	5	86.5	83.9 (82.5/82.2/85.2/85.9)	90.6 (88.4/92.1/89.6/92.6)	60.0 (57.1/56.1/66.1/62.1)	-	-
	6	86.1	85.1 (87.2/82.8/86.9/83.5)	94.9 (95.4/94.9/92.9/96.5)	50.0 (51.8/51.2/60.7/37.9)	-	-
	7	86.0	83.8 (85.5/77.4/84.8/87.5)	93.6 (96.7/89.3/92.5/95.7)	48.8 (37.5/46.3/51.8/59.1)	-	-
O ₃ (Day2)	3	94.8	89.2 (88.6/88.6/90.2/89.6)	0 (0/0/-/0)	97.3 (97.3/99.6/98.9/93.4)	26.8 (35.1/11.8/19.4/47.6)	0 (-/-/0/-)
	4	95.1	90.5 (89.6/89.9/90.2/92.3)	0 (0/0/-/0)	98.2 (97.3/98.8/99.2/97.4)	30.9 (43.2/29.4/16.1/33.3)	0 (-/-/0/-)
	5	95.2	90.2 (88.2/90.2/90.6/91.9)	0 (0/0/-/0)	97.8 (96.1/99.2/99.2/96.7)	31.7 (40.5/29.4/19.4/38.1)	0 (-/-/0/-)
	6	94.9	89.8 (86.9/89.6/90.9/91.9)	0 (0/0/-/0)	97.4 (94.6/99.2/98.1/97.8)	30.9 (40.5/23.5/32.3/23.8)	0 (-/-/0/-)
	7	95.2	90.3 (89.2/88.9/89.9/93.3)	0 (0/0/-/0)	97.1 (95.7/96.9/97.0/98.5)	39.0 (51.4/35.3/32.3/33.3)	0 (-/-/0/-)

value : Highest value

Table 8. Prediction Accuracy of DNN Model Using UM-Based Data by Number of Hidden Layers

Air pollutant	Number of hidden layer	Validation accuracy (%)	DNN-UM, Prediction accuracy(%) (East / West / South / Central)				
			Total	Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5} (Day1)	3	85.4	80.1 (80.2/79.5/82.2/78.5)	93.1 (93.4/94.6/94.9/89.2)	54.3 (48.8/57.4/48.1/60.4)	8.3 (0/0/50.0/0)	-
	4	85.7	81.0 (81.2/78.5/82.6/81.9)	94.6 (95.3/94.1/94.0/94.8)	53.8 (47.6/55.6/50.6/59.4)	16.7 (0/0/100/0)	-
	5	86.0	79.8 (82.6/75.2/82.2/79.2)	92.6 (93.9/91.4/93.1/91.8)	54.6 (56.0/50.9/54.4/57.4)	0 (0/0/0/0)	-
	6	85.9	81.8 (83.2/79.5/82.2/82.2)	93.4 (92.9/93.5/94.5/92.8)	58.9 (60.7/58.3/50.6/64.4)	8.3 (0/20.0/0/0)	-
	7	85.5	82.3 (82.9/81.5/83.6/81.2)	93.1 (93.9/93.5/94.5/90.2)	61.6 (57.1/64.8/55.7/66.3)	0 (0/0/0/0)	-
PM ₁₀ (Day1)	3	87.2	84.9 (86.2/80.9/86.9/85.6)	94.8 (99.2/91.6/96.3/91.8)	49.6 (30.4/53.6/46.4/63.6)	-	-
	4	87.4	83.2 (85.9/79.2/84.2/83.6)	91.8 (97.9/88.3/88.4/93.2)	52.7 (33.9/56.0/66.1/53.0)	-	-
	5	87.4	84.5 (86.9/78.9/85.2/86.9)	94.1 (99.2/89.3/92.1/95.3)	50.4 (33.9/52.4/55.4/57.6)	-	-
	6	86.9	85.3 (85.9/80.5/87.2/87.6)	94.6 (98.8/91.1/92.6/95.7)	52.3 (30.4/53.6/64.3/59.1)	-	-
	7	87.2	83.7 (87.2/80.2/87.2/80.2)	92.6 (99.6/89.7/94.6/85.8)	52.3 (33.9/56.0/55.4/60.6)	-	-
O ₃ (Day1)	3	94.3	89.3 (87.2/88.9/88.3/93.0)	0 (0/0/-/0)	98.6 (100/96.5/98.1/99.6)	18.3 (5.4/40.0/9.4/18.2)	0 (-/-/0/-)
	4	94.3	90.1 (86.9/90.9/89.3/93.3)	12.5 (0/0/-/50.0)	98.9 (99.2/99.2/98.1/98.9)	22.2 (8.1/37.1/18.8/27.3)	0 (-/-/0/-)
	5	94.0	89.3 (87.6/88.6/87.6/93.6)	0 (0/0/0/0)	98.4 (99.2/99.2/95.8/99.3)	19.8 (13.5/17.1/21.9/31.8)	0 (-/-/0/-)
	6	94.9	89.3 (89.9/86.9/88.3/92.3)	0 (0/0/0/0)	98.0 (97.7/98.1/98.9/97.4)	23.0 (43.2/11.4/3.1/36.4)	0 (-/-/0/-)
	7	94.3	89.3 (85.9/89.3/88.9/93.0)	0 (0/0/0/0)	98.5 (96.5/98.8/98.9/99.6)	18.3 (18.9/25.7/9.4/18.2)	0 (-/-/0/-)
PM _{2.5} (Day2)	3	80.3	78.0 (78.2/77.2/80.9/75.8)	93.1 (92.5/91.9/92.2/95.9)	46.7 (43.4/55.4/51.8/37.9)	16.7 (0/40.0/0/0)	-
	4	80.7	77.2 (79.5/72.8/78.9/77.5)	91.3 (91.1/81.2/96.8/94.9)	48.4 (51.8/61.7/30.8/45.0)	8.3 (0/0/0/33.3)	-
	5	80.3	78.9 (80.5/79.5/77.5/77.9)	92.2 (90.6/90.3/92.2/95.9)	51.9 (56.6/64.5/38.5/45.0)	0 (0/0/0/0)	-
	6	80.5	78.2 (77.5/77.5/80.2/77.5)	92.4 (85.9/94.1/93.1/96.9)	49.5 (57.8/52.3/46.2/42.0)	0 (0/0/0/0)	-
	7	80.5	77.3 (74.5/77.2/81.5/76.2)	91.3 (85.0/94.6/93.6/92.3)	48.4 (49.4/48.6/50.0/46.0)	25.0 (0/40.0/0/33.3)	-
PM ₁₀ (Day2)	3	86.2	83.2 (82.6/81.2/85.6/83.6)	92.6 (89.3/91.2/93.4/96.6)	49.8 (53.6/55.4/51.8/37.9)	-	-
	4	86.4	84.1 (85.6/82.2/86.2/82.6)	92.9 (93.0/91.6/91.7/95.3)	52.9 (53.6/57.8/62.5/37.9)	-	-
	5	86.4	82.9 (82.2/79.9/86.2/83.2)	92.4 (91.3/94.0/93.4/90.9)	49.0 (42.9/43.4/55.4/56.1)	-	-
	6	86.7	83.9 (86.6/80.5/85.6/82.9)	93.0 (94.2/90.7/92.1/94.8)	51.3 (53.6/54.2/57.1/40.9)	-	-
	7	86.7	83.3 (85.2/79.5/86.9/81.5)	92.5 (91.3/91.6/93.8/93.1)	50.6 (58.9/48.2/57.1/40.9)	-	-
O ₃ (Day2)	3	93.8	89.2 (87.2/88.6/88.3/92.6)	0 (0/0/-/0)	98.1 (96.1/98.8/98.9/98.5)	20.6 (32.4/20.0/31.1/27.3)	0 (-/-/0/-)
	4	93.3	89.4 (88.6/88.3/89.9/91.9)	0 (0/0/-/0)	99.1 (97.3/98.8/100/100)	15.1 (35.1/17.1/0/0)	0 (-/-/0/-)
	5	93.5	89.2 (87.9/87.2/89.9/91.6)	0 (0/0/-/0)	97.8 (97.7/98.8/97.7/97.1)	23.0 (27.0/8.6/28.1/31.8)	0 (-/-/0/-)
	6	93.4	89.0 (87.2/88.3/87.9/92.6)	0 (0/0/-/0)	98.3 (99.2/96.9/97.4/99.6)	17.5 (10.8/31.4/12.5/13.6)	0 (-/-/0/-)
	7	93.5	89.3 (87.9/88.3/88.9/91.9)	0 (0/0/-/0)	98.4 (97.7/97.3/98.9/99.6)	19.0 (27.0/28.6/9.4/4.5)	0 (-/-/0/-)

value : Highest value

3.4. DNN 모델의 입력 변수 및 가중치 조건 최적화

앞서 도출된 4개의 은닉층 조건을 기반으로 예측 적중률을 향상시키기 위해 입력 변수를 추가로 적용하였다. 기존에는 일평균 자료를 활용하여 딥러닝 모델을 학습시켰는데 이는 시간별 데이터를 활용한 경우보다 모델 경량화와 직관적인 해석이 가능하다는 장점이 있는 반면, 변수의 특징적인 패턴을 세밀하게 반영하기에는 한계가 있었다. 이러한 한계를 보완하기 위해, 성분으로 분리된 풍향 및 풍속 자료를 제외한 모든 변수에 대해 일최고치 변수를 추가로 도입하여 케이스별 특징을 보다 명확하게 반영하고자 하였다.

이와 같이 추가된 일최고치 변수를 반영한 데이터는 Table 9와 Table 10의 Option 1에 해당한다. 여기서 Table 9는 GFS 데이터를 기반으로 한 결과이며, Table 10은 UM 데이터를 기반으로 한 결과이다. 모든 경우에 일관되게 나타나는 것은 아니었으나, 전체 예측 적중률이나 편향성을 띠는 등급 외 등급의 적중률이 향상되는 결과를 확인할 수 있었다. 특히 일부 권역의 모델에서는 전체 등급 적중률을 유지하면서도 편향성을 가지지 않는 등급의 예측 적중률이 60% 이상으로 나타나, 편향성을 크게 극복한 사례도 발견되었다.

다음으로 데이터 편향성을 극복하기 위해 Option 1 조건에서 등급별 데이터 수에 비례하여 가중치를 부여하는 방법을 도입하였으며, 그 결과는 Table 9와 Table 10의 Option 2에 나타났다. 전체적으로 편향성은 개선된 결과를 얻었으나, 전체 데이터의 예측 적중률은 감소하는 경향을 보였다. 이는 등급별 데이터에 비례하여 가중치를 적용할 경우, 편향성이 큰 데이터의 영향이 감소하는 동시에 편향성이 작은 데이터의 영향이 상대적으로 증가하기 때문에 편향성이 큰 데이터의 본연의 가치를 저하시킨 것으로 판단된다.

이를 개선하기 위해 등급별 최소 가중치를 1.0으로 고정하여, 편향성이 큰 데이터의 영향을 유지하면서도 동시에 편향성을 극복하려는 시도를 하였다. 그 결과는 Table 9와 10의 Option 3에 제시하였다. 일부 예외는 있으나 전체 예측 정확도와 편향성이 개선된 것으로 확인되었다. 특히 오존의 경우, 관측 빈도가 매우 낮은 “좋음” 등급과 “매우 나쁨” 등급이 예측 결과에 포함되었는데, “매우 나쁨” 등급은 학습 데이터에 포함되지 않아 분석에서 별도로 고려하지 않았다. 반면, “좋음” 등급은 가중치를 과도하게 높게 설정하더라도 예측 적중률이 개선되지 않는 동시에 다른 등급에 간섭하여 부정적인 영향을 미칠 우려가 있었다.

이에 따라 “좋음” 등급의 가중치를 1.0으로 고정하여 과대 평가되지 않도록 조정하였으며, 그 결과는 Table 9와 10의 Option 4에 제시하였다. 전체적으로 편향성이 개선되는 결과를 얻었으며, UM을 기반으로 한 익일(Day1) 모델을 제외하고는 전체 등급 적중률 또한 향상된 것으로 확인되었다. (초)미세먼지의 경우 Option 1~3의 적용 결과를 비교했을 때, 전체 예측 적중률과 편향된 데이터에 대한 대응성을 종합적으로 고려하면 Option 3의 결과가 가장 안정적이었으며, 오존의 경우 Option 1~4 중 Option 4의 결과가 가장 안정적인 것으로 나타났다.

결과론적인 접근을 피하기 위해 각 항목에 대해 한 가지 Option을 적용하여 전체적으로 평가한 결과, GFS 모델을 기반으로 한 DNN 모델의 예측 적중률은 다음과 같다. 초미세먼지의 경우 익일(Day1) 예측 적중률은 80.8%(좋음 95.3%, 보통 51.9%, 나쁨 8.3%)였으며, 모레(Day2)는 79.1%(좋음 91.2%, 보통 54.1%, 나쁨 25.0%)로 나타났다. 미세먼지는 익일(Day1) 예측 적중률이 82.5%(좋음 90.6%, 보통 53.8%)였으며, 모레(Day2)는 83.2%(좋음 91.1%, 보통 55.4%)를 기록하였다.

오존의 경우 익일(Day1) 예측 적중률은 90.4%(좋음 0%, 보통 96.0%, 나쁨 49.6%, 매우 나쁨 0%)였으며, 모레(Day2)는 90.4%(좋음 0%, 보통 97.4%, 나쁨 36.6%, 매우 나쁨 0%)로 나타났다. 또한 UM 모델을 기반으로 한 DNN 모델의 예측 적중률은 다음과 같다. 초미세먼지는 익일(Day1) 예측 적중률이 80.5%(좋음 89.9%, 보통 61.8%, 나쁨 25.0%)였으며, 모레(Day2)는 78.4%(좋음 93.7%, 보통 46.2%, 나쁨 25.0%)를 기록하였다. 미세먼지는 익일(Day1) 예측 적중률이 85.4%(좋음 90.9%, 보통 66.0%)였으며, 모레(Day2)는 83.1%(좋음 90.9%, 보통 55.2%)로 나타났다. 오존의 경우 익일(Day1) 예측 적중률은 89.4%(좋음 12.5%, 보통 95.8%, 나쁨 41.3%, 매우 나쁨 0%)였으며, 모레(Day2)는 89.8%(좋음 97.5%, 보통 97.5%, 나쁨 31.0%, 매우 나쁨 0%)로 나타났다.

여기서 GFS 모델을 기반으로 한 DNN 모델은 (초)미세먼지를 제외하고 익일(Day1)의 예측 적중률과 편향성 대응 측면에서 모레(Day2)보다 뛰어난 성능을 보였다. 초미세먼지의 경우, 예측 적중률은 GFS 모델 기반의 DNN 모델이 높았으나, 익일(Day1) 예측에서는 UM 모델 기반의 DNN 모델이 편향성 대응 측면에서 더 우수한 것으로 나타났다. 미세먼지의 경우, 익일

(Day1)에서는 UM 모델 기반의 DNN 모델이 더 뛰어난 성능을 보였으나, 모레(Day2)에는 GFS 모델이 약간 우세한 것으로 나타났다. 오존의 경우, 전체적으로 GFS 모델 기반의 DNN 모델이 예측 적중률과 편향성 대응 측면에서 우수한 성능을 보였다.

Table 9. Prediction accuracy of DNN model with different tuning options based on GFS data

Air pollutant	Option	DNN-GFS, Prediction accuracy(%) (East / West / South / Central)				
		Total	Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5} (Day1)	Option 1: Daily mean + maximum inputs	80.6 (81.1/79.8/82.5/78.8)	96.5 (97.6/94.0/96.8/97.4)	48.4 (41.7/59.3/44.3/45.5)	8.3 (0/0/50.0/0)	-
	Option 2: Option 1 + Grade-balanced weights	80.3 (82.5/79.8/82.2/76.8)	93.8 (97.2/90.8/95.4/91.2)	53.5 (47.6/64.8/46.8/51.5)	8.3 (0/0/50.0/0)	-
	Option 3: Option 2 + Minimum weight of 1	80.8 (81.5/78.1/85.5/78.1)	95.3 (96.7/97.8/93.5/93.3)	51.9 (45.2/48.1/65.8/50.5)	8.3 (0/0/50.0/0)	-
PM ₁₀ (Day1)	Option 1: Daily mean + maximum inputs	83.9 (84.8/81.1/87.5/82.2)	93.8 (97.1/92.0/92.5/93.5)	48.9 (32.1/53.6/66.1/42.4)	-	-
	Option 2: Option 1 + Grade-balanced weights	83.2 (84.2/81.5/84.8/82.5)	91.9 (95.9/92.0/88.4/91.3)	52.7 (33.9/54.8/69.6/51.5)	-	-
	Option 3: Option 2 + Minimum weight of 1	82.5 (83.5/81.8/83.5/81.1)	90.6 (95.9/88.3/87.6/90.5)	53.8 (30.4/65.5/66.1/48.5)	-	-
O ₃ (Day1)	Option 1: Daily mean + maximum inputs	90.7 (90.6/90.9/90.6/90.9)	0 (0/0/-/0)	99.1 (98.1/99.6/99.6/98.9)	27.2 (45.9/32.4/18.8/0)	0 (-/-/0/-)
	Option 2: Option 1 + Grade-balanced weights	89.9 (90.2/87.2/88.6/91.2)	12.5 (0/0/-/50.0)	95.1 (96.9/94.6/94.7/94.1)	46.4 (51.4/38.2/40.6/59.1)	0 (-/-/0/-)
	Option 3: Option 2 + Minimum weight of 1	90.0 (89.9/88.9/90.2/90.9)	0 (0/0/-/0)	96.1 (98.4/95.8/96.2/94.1)	44.8 (37.8/44.1/43.8/59.1)	0 (-/-/0/-)
	Option 4: Good class weight adjusted to 1	90.4 (90.2/90.2/89.6/91.6)	0 (0/0/-/0)	96.0 (95.7/97.7/95.8/94.9)	49.6 (59.5/41.2/40.6/59.1)	0 (-/-/0/-)
PM _{2.5} (Day2)	Option 1: Daily mean + maximum inputs	78.6 (78.8/79.1/80.8/75.8)	93.8 (95.3/91.9/96.3/91.3)	47.3 (38.6/60.4/39.7/46.5)	8.3 (0/0/0/33.3)	-
	Option 2: Option 1 + Grade-balanced weights	76.4 (78.8/73.7/80.5/72.7)	84.9 (81.1/88.2/86.2/84.6)	59.8 (74.7/51.9/66.7/50.5)	8.3 (0/0/0/33.3)	-
	Option 3: Option 2 + Minimum weight of 1	79.1 (79.8/75.1/82.8/78.8)	91.2 (92.0/87.1/94.9/90.3)	54.1 (50.6/56.6/50.0/57.6)	25.0 (0/20.0/50.0/33.3)	-
PM ₁₀ (Day2)	Option 1: Daily mean + maximum inputs	84.8 (85.2/83.5/85.5/85.2)	94.6 (96.7/94.0/92.9/94.8)	50.0 (35.7/56.1/53.6/51.5)	-	-
	Option 2: Option 1 + Grade-balanced weights	83.9 (87.2/79.5/86.5/82.5)	93.8 (92.1/93.0/94.2/95.7)	48.8 (66.1/43.9/53.6/36.4)	-	-
	Option 3: Option 2 + Minimum weight of 1	83.2 (85.2/76.4/87.5/83.8)	91.1 (95.9/80.0/91.3/96.1)	55.4 (39.3/67.1/71.4/40.9)	-	-
O ₃ (Day2)	Option 1: Daily mean + maximum inputs	90.2 (86.9/90.6/90.2/93.3)	0 (0/0/-/0)	98.5 (96.1/99.2/99.2/99.3)	26.0 (29.7/32.4/16.1/23.8)	0 (-/-/0/-)
	Option 2: Option 1 + Grade-balanced weights	89.6 (89.6/89.2/88.6/91.2)	0 (0/0/-/0)	97.6 (97.7/100/96.6/96.4)	27.6 (40.5/14.7/22.6/33.3)	0 (-/-/0/-)
	Option 3: Option 2 + Minimum weight of 1	89.5 (87.9/88.2/90.9/90.9)	0 (0/0/-/0)	97.1 (95.3/98.1/99.6/95.3)	30.9 (43.2/20.6/19.4/42.9)	0 (-/-/0/-)
	Option 4: Good class weight adjusted to 1	90.4 (89.9/89.9/90.2/91.6)	0 (0/0/-/0)	97.4 (98.1/96.5/99.2/96.0)	36.6 (40.5/47.1/16.1/42.9)	0 (-/-/0/-)

value : Highest value

Table 10. Prediction accuracy of DNN model with different tuning options based on UM data

Air pollutant	Option	Validation accuracy (%)	DNN-UM, Prediction accuracy(%) (East / West / South / Central)				
			Total	Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5} (Day1)	Option 1: Daily mean + maximum inputs	81.5 (81.5/79.9/80.9/83.6)	92.3 (96.2/92.4/87.6/93.3)	60.5 (46.4/62.0/64.6/67.3)	0 (0/0/0/0)	-	
	Option 2: Option 1 + Grade-balanced weights	79.2 (83.2/77.5/78.9/77.2)	93.7 (95.3/91.4/91.2/96.9)	48.9 (53.6/54.6/45.6/41.6)	41.7 (50.0/60.0/50.0/0)	-	
	Option 3: Option 2 + Minimum weight of 1	80.5 (83.2/80.2/80.2/78.2)	89.9 (91.5/92.4/86.6/89.2)	61.8 (64.3/60.2/64.6/59.4)	25.0 (0/60.0/0/0)	-	
	Option 1: Daily mean + maximum inputs	85.2 (86.2/82.2/87.2/85.2)	98.1 (97.9/85.5/95.0/93.1)	57.3 (35.7/73.8/53.6/57.6)	-	-	
PM ₁₀ (Day1)	Option 2: Option 1 + Grade-balanced weights	84.3 (84.6/81.5/85.2/85.9)	92.0 (93.0/90.7/91.3/93.1)	56.9 (48.2/58.3/58.9/60.6)	-	-	
	Option 3: Option 2 + Minimum weight of 1	85.4 (86.2/83.9/85.9/85.6)	90.9 (95.5/86.4/89.7/91.4)	66.0 (46.4/77.4/69.6/65.2)	-	-	
	Option 1: Daily mean + maximum inputs	90.3 (87.9/89.6/90.3/93.3)	12.5 (0/33.3/-/0)	98.4 (98.1/97.3/98.5/99.6)	27.8 (24.3/37.1/25.0/22.7)	0 (-/-/0/-)	
	Option 2: Option 1 + Grade-balanced weights	88.3 (85.9/88.3/87.9/90.9)	25.0 (66.7/0/-/0)	94.4 (94.2/93.5/94.0/96.0)	41.3 (29.7/57.1/40.6/36.4)	0 (-/-/0/-)	
O ₃ (Day1)	Option 3: Option 2 + Minimum weight of 1	89.7 (87.6/88.6/89.6/93.0)	0 (0/0/-/0)	98.2 (98.8/97.7/96.2/100)	24.6 (16.2/28.6/37.5/13.6)	0 (-/-/0/-)	
	Option 4: Good class weight adjusted to 1	89.4 (88.3/87.6/88.3/93.6)	12.5 (0/33.3/-/0)	95.8 (96.1/92.7/95.8/98.5)	41.3 (40.5/54.3/28.1/40.9)	0 (-/-/0/-)	
	Option 1: Daily mean + maximum inputs	80.5 (80.9/79.9/80.5/80.9)	93.7 (95.8/91.9/92.7/94.4)	54.1 (44.6/62.6/48.7/57.0)	0 (0/0/0/0)	-	
	Option 2: Option 1 + Grade-balanced weights	78.2 (78.2/77.5/79.2/77.9)	93.2 (92.5/89.2/95.9/94.9)	46.7 (43.4/58.9/34.6/46.0)	25.0 (0/40.0/0/33.3)	-	
PM _{2.5} (Day2)	Option 3: Option 2 + Minimum weight of 1	78.4 (78.9/77.9/80.2/76.5)	93.7 (91.5/92.5/95.9/94.9)	46.2 (48.2/53.3/38.5/43.0)	25.0 (0/60.0/0/0)	-	
	Option 1: Daily mean + maximum inputs	84.2 (84.9/80.9/84.6/86.6)	92.1 (92.1/92.1/89.3/94.8)	56.3 (53.6/51.8/64.3/57.6)	-	-	
	Option 2: Option 1 + Grade-balanced weights	82.6 (84.9/80.9/84.6/80.2)	90.0 (92.6/92.1/89.3/86.2)	56.3 (51.8/51.8/64.3/59.1)	-	-	
	Option 3: Option 2 + Minimum weight of 1	83.1 (83.2/81.5/84.9/82.6)	90.9 (90.5/90.7/90.1/92.2)	55.2 (51.8/57.8/62.5/48.5)	-	-	
O ₃ (Day2)	Option 1: Daily mean + maximum inputs	89.4 (87.6/89.6/88.9/91.6)	0 (0/0/-/0)	98.9 (98.1/98.1/100/99.3)	16.7 (21.6/34.3/0/4.5)	0 (-/-/0/-)	
	Option 2: Option 1 + Grade-balanced weights	88.1 (87.9/85.6/87.9/90.9)	12.5 (33.3/0/-/0)	96.3 (97.7/94.6/96.2/96.7)	24.5 (24.3/25.7/21.9/27.3)	0 (-/-/0/-)	
	Option 3: Option 2 + Minimum weight of 1	88.5 (87.2/87.2/89.6/89.9)	0 (0/0/-/0)	97.5 (96.9/100/98.1/95.3)	19.0 (27.0/0/21.9/31.8)	0 (-/-/0/-)	
	Option 4: Good class weight adjusted to 1	89.8 (89.6/86.6/89.6/93.3)	0 (0/0/-/0)	97.5 (97.7/95.8/98.9/97.8)	31.0 (40.5/25.7/15.6/45.5)	0 (-/-/0/-)	

value : Highest value

3.5. DNN 모델의 성능 한계 및 실무 활용 가능성 분석

다음으로 각 DNN 모델의 성능 한계를 평가하기 위해 2021년부터 2023년까지의 데이터를 학습 데이터로 사용하고 2024년 데이터를 검증 데이터로 사용하여 모델을 평가하였고 그 결과는 DNN_{-MAX}로 표기하여 Table 11과 Table 12에 나타냈다. Table 11은 GFS 데이터를 기반으로 한 결과이며, Table 12는 UM 데이터를 기반으로 한 결과이다.

GFS 모델을 기반으로 한 DNN 모델의 경우, 초미세 먼지의 익일(Day1) 예측 한계 적중률은 82.9%(좋은 91.7%, 보통 66.1%, 나쁨 16.7%)였으며, 모레(Day2)

예측 한계 적중률은 83.0%(좋은 92.2%, 보통 65.0%, 나쁨 8.3%)로 나타났다. 미세먼지의 경우 익일(Day1) 예측 한계 적중률은 87.0%(좋은 92.7%, 보통 66.8%)였으며, 모레(Day2) 예측 한계 적중률은 86.6%(좋은 93.5%, 보통 65.0%)를 기록하였다. 오존의 경우 익일(Day1) 예측 한계 적중률은 92.3%(좋은 0%, 보통 97.2%, 나쁨 56.8%, 매우 나쁨 0%)였으며, 모레(Day2) 예측 한계 적중률은 91.7%(좋은 0%, 보통 97.3%, 나쁨 50.4%, 매우 나쁨 0%)를 기록하였다.

UM 모델을 기반으로 한 DNN 모델의 경우, 초미세 먼지의 익일(Day1) 예측 한계 적중률은 82.6%(좋은

92.0%, 보통 64.0%, 나쁨 25.0%)였으며, 모레(Day2) 예측 한계 적중률은 81.7%(좋은 89.4%, 보통 66.8%, 나쁨 16.7%)로 나타났다. 미세먼지는 익일(Day1) 예측 한계 적중률이 88.3%(좋은 94.0%, 보통 68.3%)였으며, 모레(Day2) 예측 한계 적중률은 86.2%(좋은 92.3%, 보통 64.4%)를 기록하였다. 오존의 경우 익일(Day1) 예측 한계 적중률은 91.6%(좋은 0%, 보통 97.4%, 나쁨 49.2%, 매우 나쁨 0%)였으며, 모레(Day2) 예측 한계 적중률은 91.1%(좋은 0%, 보통 97.9a%, 나쁨 40.5%, 매우 나쁨 0%)로 나타났다. 성능 한계는 전체적으로 익일(Day1)이 모레(Day2)보다 우수한 것으로 나타났다.

대기오염물질별로 살펴보면, 모델 최적화 과정에서

와 유사한 경향을 보였는데, 초미세먼지의 경우 예측 적중률은 GFS 모델 기반의 DNN 모델이 높았으나 익일(Day1) 예측에서는 UM 모델 기반의 DNN 모델이 편향성 대응 측면에서 더 우수한 것으로 나타났다. 미세먼지의 경우 익일(Day1)은 UM 모델 기반의 DNN 모델이, 모레(Day2)는 GFS 모델이 더 우수한 성능을 보였다. 오존의 경우 전체적으로 GFS 모델 기반의 DNN 모델이 예측 적중률과 편향성 대응 측면에서 가장 우수한 것으로 나타났다.

Table 11. Prediction accuracy(Limits) of DNN-GFS using training data without split and validation on test dataset

Air pollutant	Option	DNN-GFS, Prediction accuracy(%) (East / West / South / Central)				
		Total	Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5} (Day1)	DNN-MAX Optimum option with no split on training set; validation on test dataset	82.9 (83.5/80.5/85.5/82.2)	91.7 (90.5/89.7/92.6/93.8)	66.1 (67.9/67.6/68.4/61.4)	16.7 (0/20.0/0/33.3)	-
PM ₁₀ (Day1)		87.0 (89.6/83.8/88.2/86.2)	92.7 (91.7/89.2/95.0/94.4)	66.8 (80.4/70.2/58.9/57.6)	-	-
O ₃ (Day1)		92.3 (92.6/90.9/91.6/93.9)	0 (0/0/-/0)	97.2 (95.3/97.7/98.9/97.1)	56.8 (81.1/47.1/34.4/63.6)	0 (-/0/-)
PM _{2.5} (Day2)	DNN-MAX Optimum option with no split on training set; validation on test dataset	83.0 (83.5/82.5/84.8/81.1)	92.2 (92.5/88.2/93.5/94.4)	65.0 (67.9/65.9/69.6/60.6)	8.3 (0/0/0/33.3)	-
PM ₁₀ (Day2)		86.6 (88.6/82.8/88.6/86.5)	92.5 (93.4/89.3/92.9/93.9)	65.8 (67.9/65.9/69.6/60.6)	-	-
O ₃ (Day2)		91.7 (89.6/91.2/91.6/94.3)	0 (0/0/-/0)	97.3 (96.1/97.3/96.6/98.9)	50.4 (51.4/52.9/51.6/42.9)	0 (-/0/-)

Table 12: Prediction accuracy(Limits) of DNN-UM using training data without split and validation on test dataset

Air pollutant	Option	DNN-UM, Prediction accuracy(%) (East / West / South / Central)				
		Total	Good	Moderate	Unhealthy	Very unhealthy
PM _{2.5} (Day1)	DNN-MAX Optimum option with no split on training set; validation on test dataset	82.6 (83.2/80.5/84.6/81.9)	92.0 (93.4/89.2/90.3/94.8)	64.0 (58.3/69.4/70.9/57.4)	25.0 (50.0/0/0/66.7)	-
PM ₁₀ (Day1)		88.3 (88.9/90.9/91.9/93.6)	94.0 (95.0/93.0/93.4/94.4)	68.3 (62.5/67.9/76.8/66.7)	-	-
O ₃ (Day1)		91.6 (89.9/90.9/91.9/93.6)	0 (0/0/-/0)	97.4 (98.4/97.3/95.8/98.2)	49.2 (37.8/51.4/62.5/45.5)	0 (-/0/-)
PM _{2.5} (Day2)	DNN-MAX Optimum option with no split on training set; validation on test dataset	81.7 (80.5/80.5/83.2/82.6)	89.4 (87.8/88.7/92.7/88.2)	66.8 (63.9/68.2/59.0/74.0)	16.7 (0/40.0/0/0)	-
PM ₁₀ (Day2)		86.2 (86.9/83.2/87.9/86.6)	92.3 (94.2/87.4/94.6/92.2)	64.4 (55.4/72.3/58.9/66.7)	-	-
O ₃ (Day2)		91.1 (88.9/91.6/90.9/93.0)	0 (0/0/-/0)	97.9 (97.7/97.7/98.1/98.2)	40.5 (35.1/54.3/34.4/36.4)	0 (-/0/-)

value : Highest value

최적화된 DNN 모델의 성능을 평가하고, 실제 적용 가능성을 확인하기 위해 DNN-MAX(이론적 성능 한계), CMAQ 모델, 그리고 예측 담당자의 예측 결과와 비교를 수행하였다. 이를 통해 최적화된 모델이 어느 수준까지 성능을 개선할 수 있는지를 확인하고, 기존 예측 시스템과의 상대적 성능 차이를 시각화하였다.

Fig. 7과 Fig. 8은 Day1(익일)과 Day2(모레)의 예측 결과를 대기오염물질별로 비교한 것이다. 각 그림에서 항목별 왼쪽 막대 그래프는 전체 예측 적중률을 비교한 것이며, 오른쪽 원 그래프는 등급별(“좋음”, “보통”, “나쁨”) 예측 결과를 시각화한 것이다. 원 그래프의 구성은 중심부터 CMAQ 모델 결과, 최적화된 DNN 모델, DNN-MAX(성능 한계 모델), 예측 담당자의 예측 결과 순으로 배치되어 있다. 전체적으로 최적화된 DNN 모델의 예측 적중률은 CMAQ 모델보다 우세한 경향을 보였으나, 예측 담당자의 예측 결과 수준에는 미치지 못하는 것으로 나타났다. 예측 범위와 항목, 시기가 달라 직접적인 비교는 어렵지만, 국립환경과학원의 인공지능을 활용한 미세먼지 단기 예측 연구에 따

르면, LSTM 기반 예측 모델이 예보관이 예측한 초미세먼지 77.4%의 적중률보다 3% 이내의 결과를 도출하였다.

또한, 우리 원의 사전 연구에서는 초미세먼지의 예측 적중률이 최대 80.1% 수준으로, 예측 담당자의 예측 수준에는 완전히 미치지 못하였으나, CMAQ 모델의 결과보다는 개선된 경향을 보였다는 점에서 공통된 결과로 나타났다.^{15),17)} 한편, DNN-MAX(학습 데이터셋 전체를 사용하고 테스트 데이터셋으로 검증한 모델)의 경우 예측 담당자의 예측 적중률과 유사하거나 이를 상회하는 결과를 도출하였다. 그러나 초미세먼지의 “나쁨” 등급 예측 적중률은 개선되지 못했는데, 이는 CMAQ 모델이 고농도 데이터에 대해 과도하게 예측하는 경향을 보이기 때문으로 판단된다. 실제로 CMAQ 모델의 예측 결과에서 “나쁨” 등급의 일부 데이터가 “매우 나쁨” 등급으로 과대 예측되는 사례가 확인되었으며, 비율은 익일(Day1) 29.2%, 모레(Day2) 33.3%로 나타났다. 이러한 CMAQ 모델의 특성이 DNN 모델 학습 과정에서 방해 요인으로 작용했을 가능성이 있으

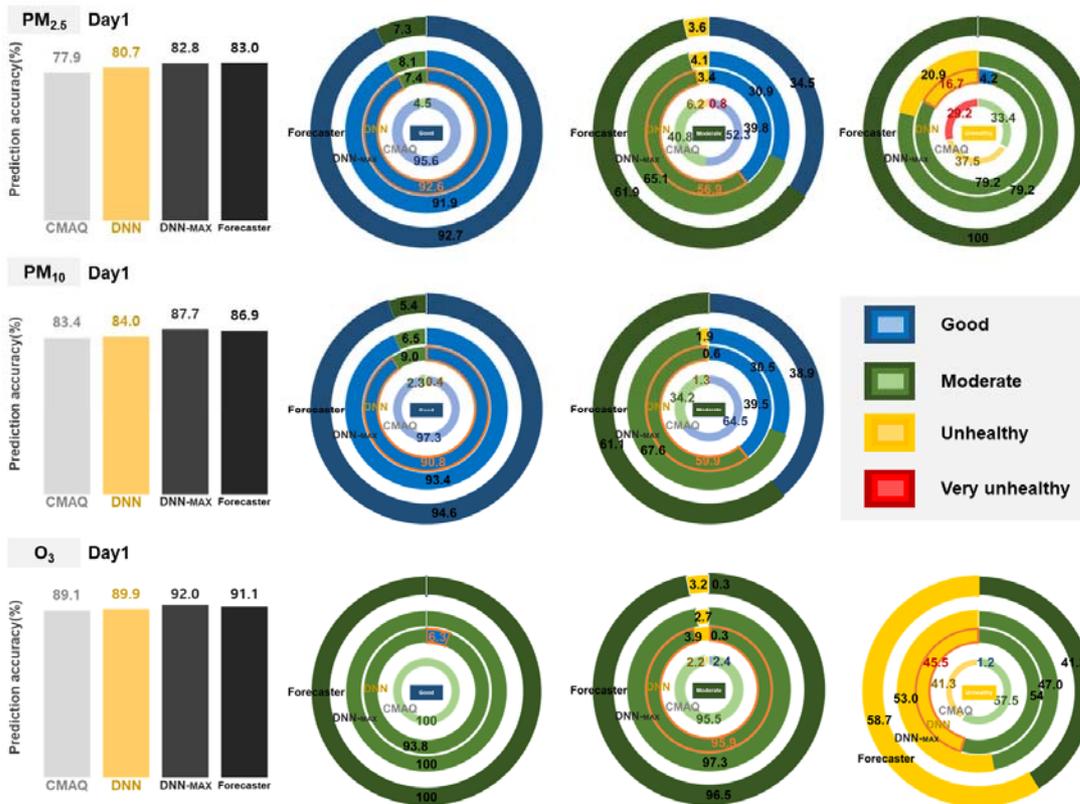


Fig. 7. Comparison of optimized DNN model with DNN-MAX (Theoretical performance limit), CMAQ model, and forecaster predictions for Day1

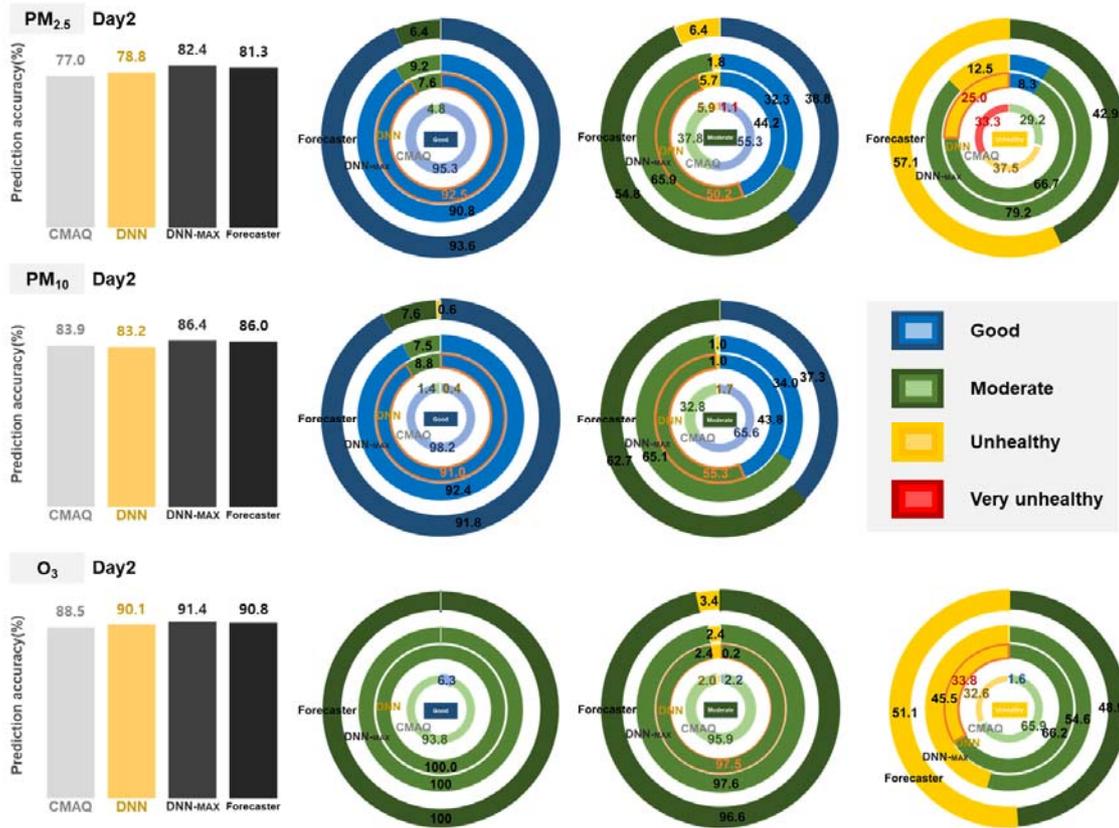


Fig. 8. Comparison of optimized DNN model with DNN-MAX (Theoretical performance limit), CMAQ model, and forecaster predictions for Day2

며, 특히 “나쁨” 등급의 예측 성능 개선에 한계를 가져 온 것으로 보인다.

본 연구에서 최적화된 DNN 모델은 CMAQ 모델 대비 예측 적중률이 우수하고 모델의 과소 모의로 인한 출력 자료의 편향성을 일부 개선한 것으로 나타나 예측 담당자의 보조 자료로 활용될 수 있을 것으로 보인다. 여기서 예측 적중률을 이론적 한계 성능까지 끌어 올릴 경우, 현업에서도 실질적인 지원 도구로 사용이 가능할 것으로 판단된다. 모델 구축에 사용된 훈련 데이터는 2021년부터 2023년까지의 자료로, 이 중 9.6%는 자료가 불충분하여 학습에 활용되지 못하였고 나머지 중 20%는 검증 데이터로 사용되었다. 이러한 제한된 데이터로는 예측 적중률을 충분히 평가하기에 다소 부족한 측면이 있었으며, 보다 장기간에 걸쳐 데이터를 확보한 후 이를 기반으로 모델을 재구축하고 평가해볼 필요가 있다고 판단된다. 또한 CMAQ 모델의 입력 자료와 옵션들은 현업 운영 과정에서 지속적으로 개선되고 있기 때문에, 이러한 변화 속에서도 모델이 안정적으로 작동하는지를 검증할 필요가 있다.

이번 연구에서는 2024년 1월 배출량 입력 자료가 변경되었음에도 비교적 안정적인 결과를 확인할 수 있었다. 향후 모델 성능 한계를 극복하기 위해 예측 담당관이 사용하는 추가 변수의 검토가 필요하며, 일평균 및 일최고 자료를 활용하는 경우에는 DNN 모델이 여전히 적합한 선택으로 보인다. 다만, 예측 모델의 성능을 중량화하고 시계열 특성을 보다 정밀하게 반영하기 위해서는 일간 시계열 자료로 변환한 후, LSTM이나 CNN과 같은 시계열 딥러닝 모델을 적용하는 방안도 함께 고려할 필요가 있다고 판단된다.

4. 결론

본 연구에서는 우리원에서 수행하는 (초)미세먼지와 오존의 익일(Day1), 모레(Day2) 예측 업무를 효율적으로 수행하기 위해 DNN 모델을 구축하였다. 이를 위해 2021년부터 2023년까지의 관측 데이터를 학습 데이터로 활용하고, 최적의 모델 구조를 도출한 뒤 2024년 10월까지의 대기질을 예측하였다. 예측 결과는

CMAQ 모델과 예보 담당자의 예측 결과와 비교·분석하여 모델의 성능과 한계를 평가하였으며, 이를 통해 대기질 예측의 실무 적용 가능성을 확인하고자 하였고 다음과 같은 결론을 도출하였다.

1. 우리원에서 활용 중인 대기질 진단평가시스템의 CMAQ 모델 결과를 대기오염물질별 예보 등급으로 환산한 결과, 모델의 과소 모의 영향 등으로 (초)미세먼지는 “좋음”, 오존은 “보통” 등 관측 빈도가 높은 등급에 예측이 집중되는 편향성이 확인되었다. 이에 예측 담당자가 활용하는 모델에 기상 변수와 대기질 자료의 일평균값, 전일 모델과 실측치 간의 보정치를 입력 변수로 적용하고 은닉층 개수를 3~7개 범위로 조정하여 DNN 모델을 구축한 결과, CMAQ 모델 대비 예측 적중률이 개선되었으며 편향성 문제 또한 일부 해소된 것으로 나타났다. 특히 은닉층 개수 4개를 적용한 경우 예측 적중률과 편향성 개선 효과가 가장 뛰어나고 안정적인 성능을 보였다.
2. 예측 적중률과 편향성을 단계적으로 개선하기 위해, 평균화 과정에서 데이터의 세부 특성이 단순화되는 문제를 보완하고자 입력 변수에 일최고 값을 추가하였으며, 학습 자료의 등급별 데이터 편향성을 개선하기 위해 예측 데이터를 축소하지 않으면서 자료의 규모에 비례하여 가중치를 부여하는 방식으로 DNN 모델을 순차적으로 구축하였다. 그 결과, CMAQ 모델 대비 예측 적중률은 GFS 기반 모델의 경우 익일(Day1) 초미세먼지 80.8%, 미세먼지 82.5%, 오존 90.4%, 모레(Day2)는 초미세먼지 79.1%, 미세먼지 83.2%, 오존 90.4%로 나타났다. UM 기반 모델의 경우 익일(Day1) 초미세먼지 80.5%, 미세먼지 85.4%, 오존 89.4%, 모레(Day2)는 초미세먼지 78.4%, 미세먼지 83.1%, 오존 89.8%를 기록하였다. 전체적으로 DNN 모델은 CMAQ 모델 대비 약 3%p 내외 수준에서 예측 적중률이 개선되었으며, 편향성이 낮은 데이터의 적중률 또한 상승하여 CMAQ 모델 결과가 가지는 예측 편향성과 성능 한계를 일정 부분 해결한 것으로 나타났다. 다만, 이러한 개선 수준이 예측 담당자의 예측 결과에는 미치지 못하는 것으로 확인되어 최적화된 DNN 모델은 현업에서 예측 업무의 보조 자료로 사용하는 것이 합당하다고 판단된다.
3. 동일 구조의 DNN 모델로 한계 성능을 분석한 결과, 예측 담당자의 예측 결과와 유사하거나 이를 소폭

상회하는 성과를 도출하였다. 학습 데이터를 보다 충분히 확보할 경우 예측 적중률을 더욱 향상시켜 현업에 활용 가능성을 높일 수 있을 것으로 판단된다. 또한, DNN 모델의 구조적 개선을 위해 입력 인자의 추가를 검토하거나, 입력 변수의 형태를 일간 시계열 자료로 변환하여 다양한 딥러닝 모델(LSTM, CNN 등)을 적용하는 방안도 고려할 필요가 있다고 판단된다.

참고문헌

1. Y. Lim, Y. Choi, E. Kang, Y. Jeong, J. Park and H. W. Han, Association between short- and medium-term exposure to air pollutants and depressive episode using comprehensive air quality index among the population in South Korea, *Journal of Affective Disorders*, 356, pp.307-315 (2024).
2. G. O. Ofremu, B. Y. Raimi, S. O. Yusuf, B. A. Dziwornu, S. G. Nnabuife, A. M. Eze and C. A. Nnajiolor, Exploring the Relationship between Climate Change, Air Pollutants and Human Health: Impacts, Adaptation, and Mitigation Strategies, *Green Energy and Resources*, 2024, 100074.
3. I. G. Canu, P. Wild, T. Charreau, R. Freund, A. Toto, J. Pralong, K. Sakthithasan, V. Jouannique, A. Debatisse and G. Suarez, Long-term exposure to PM₁₀ and respiratory health among Parisian subway workers, *International Journal of Hygiene and Environmental Health*, 256, 114316 (2024).
4. N. R. Rahimi, A. Azhdarpoor and R. Fouladi-Fard, Exposure to tropospheric ozone and NO₂ in the ambient air of Tehran metropolis: Spatiotemporal distribution and inhalation health risk assessment, *Physics and Chemistry of the Earth, Parts A/B/C*, 136, 103777 (2024).
5. Korea Environment Corporation, Airkorea, <https://www.airkorea.or.kr/>.
6. Busan Metropolitan City Institute of Health & Environment, <https://heis.busan.go.kr/environmental/air0>

- 06.aspx
7. D. R. Choi and Y. S. Koo, An Evaluation of the influence of boundary conditions from GEOS-Chem on CMAQ Simulations over East Asia, *Journal of the Korean Society for Atmospheric Environment*, 29(2), pp.186-198 (2013)
 8. J. Kim and Y. K. Jang, Uncertainty assessment for CAPSS emission inventory by DARS, *Journal of the Korean Society for Atmospheric Environment*, 30(1), pp.26-36 (2014).
 9. J. A. Lee, S. Alessandrini, J.-H. Kim, S. Meech, R. Kumar, I. V. Djalalova and J. M. Wilczak, Comparison of CAMS and CMAQ analyses of surface-level $PM_{2.5}$ and O_3 over the conterminous United States (CONUS), *Atmospheric Environment*, 338, 120833 (2024)
 10. K. K. R. Samal, K. S. Babu and S. K. Das, Multi-directional temporal convolutional artificial neural network for $PM_{2.5}$ forecasting with missing values: A deep learning approach, *Urban Climate*, 36, 100800 (2021).
 11. Z. Wang, R. Li, Z. Chen, Q. Yao, B. Gao, M. Xu, L. Yang, M. Li and C. Zhou, The estimation of hourly $PM_{2.5}$ concentrations across China based on a Spatial and Temporal Weighted Continuous Deep Neural Network (STWC-DNN), *ISPRS Journal of Photogrammetry and Remote Sensing*, 190, pp.38-55 (2022).
 12. P. Y. Kow, I. W. Hsia, L. C. Chang and F. J. Chang, Real-time image-based air quality estimation by deep learning neural networks, *Journal of Environmental Management*, 307, 114560 (2022).
 13. C. Menares, P. Perez, S. Parraguez and Z. L. Fleming, Forecasting $PM_{2.5}$ levels in Santiago de Chile using deep learning neural networks, *Urban Climate*, 38, 100906 (2021).
 14. M. Teng, S. Li, J. Xing, C. Fan, J. Yang, S. Wang, G. Song, Y. Ding, J. Dong and S. Wang, 72-hour real-time forecasting of ambient $PM_{2.5}$ by hybrid graph deep neural network with aggregated neighborhood spatiotemporal information, *Environment International*, 176, 107971 (2023).
 15. National Institute of Environmental Research, A Development of Short-term Prediction Tool for PM_{10} and $PM_{2.5}$ Concentrations using Artificial Intelligence (V), (2023).
 16. National Institute of Environmental Research, evelopment of Mid-term Prediction tool for PM_{10} and $PM_{2.5}$ using Artificial Intelligence (V), (2023)
 17. W. G. Do, D. Y. Kim, H. J. Song and G. J. Cho, A Study on the $PM_{2.5}$ forecasting Method in Busan Using Deep Neural Network, *Journal of Environmental Science International*, 32(8), pp.595-611 (2023).
 18. National Center for Meteorological Supercomputer, <https://www.kma.go.kr/super/>
 19. NCAR Research Data Archive, NCEP GFS 0.25 Degree Global Forecast Auxiliary Grids Historical Archive, <https://rda.ucar.edu/datasets/ds084.3/#description>.
 20. WRF MODEL USERS' PAGE, <https://www2.mmm.ucar.edu/wrf/users/>.
 21. I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press (2016).