Journal of the Korea Academia-Industrial cooperation Society
Vol. 27, No. 10 pp. 9-14, 2026

https://doi.org/10.5762/KAIS.2026.27.10.9
ISSN 1975-4701 / eISSN 2288-4688

국내 지상관측자료 기반 머신러닝 초미세먼지($PM_{2.5}$) 예측 모델의 지역별 성능 평가

오승민
강원대학교 삼척캠퍼스 첨단AI공학과

Regional Performance Evaluation of Machine Learning-Based $PM_{2.5}$ Prediction Models Using Ground Observation Data in Korea

Sungmin O
Department of Advanced AI Engineering, Samcheock, Kangwon National University

요 약 본 연구는 국내 지상관측자료를 활용하여 지역별 초미세먼지($PM_{2.5}$) 단기 예측 성능을 평가하였다. 2018년부터 2024년까지의 시간별 기상 및 대기질 관측자료를 17개 시·도 행정구역 단위로 집계하고, 기상 변수, 시간 주기 변수, 지역 더미변수, $PM_{2.5}$ 지연 변수를 입력자료로 활용하였다. 지속성 기준모델과 비교한 결과, Random Forest, XGBoost, LightGBM 등 머신러닝 기반 모델은 모두 더 우수한 성능을 보였다. 세 모델 간 성능 차이는 크지 않았으나, LightGBM은 전체 검증자료에서 전반적으로 가장 우수한 예측 성능을 나타내어 주 분석을 위한 대표 모델로 선정하였다. 모델 예측 오차는 행정구역에 따라 뚜렷한 차이를 보였으며, 이는 동일한 모델을 적용하더라도 지역별 농도 변동성 및 관측자료의 대표성에 따라 예측 난이도가 달라질 수 있음을 의미한다. 기상 변수와 $PM_{2.5}$ 간 선형 상관관계는 전반적으로 크지 않았으나, 변수 중요도 분석 결과 직전 시점의 $PM_{2.5}$ 농도, 계절성, 상대습도 및 기온 등이 주요 예측 요인으로 나타났다. 또한 고농도 구간에서 예측오차가 증가하는 경향을 보였으며, 향후 고농도 예측 성능 향상을 위한 추가 변수의 결합, 고농도 사례에 대한 가중학습 등의 보완 전략이 필요할 것으로 판단된다. 본 연구 결과는 국내 초미세먼지 예측모델의 효과적인 구축과 활용을 위해 모델의 평균 성능뿐만 아니라 지역 및 농도 구간별 예측 성능을 함께 검토할 필요가 있음을 제시한다. 또한 본 연구는 지상관측자료만으로 확보 가능한 예측 성능을 제시함으로써, 향후 외부 자료 결합에 따른 성능 향상을 평가하기 위한 기준선을 제공했다는 점에서 의의가 있다.

Abstract This study evaluated regional short-term $PM_{2.5}$ prediction performance using ground observation data in South Korea. Hourly meteorological and air quality observations from 2018 to 2024 were aggregated for 17 administrative regions. The meteorological variables, temporal cycle variables, regional dummy variables, and lagged $PM_{2.5}$ concentrations were used as inputs. Compared to the persistence baseline, Random Forest, XGBoost, and LightGBM all performed better. Although the three models perform similarly, LightGBM showed the best validation performance and was selected as the representative model. The prediction errors vary substantially across regions, suggesting that prediction difficulty differs according to the regional concentration variability and the representativeness of observation data, even when the same model was applied. Although linear correlations between meteorological variables and $PM_{2.5}$ are generally weak, feature importance analysis identified lagged $PM_{2.5}$ concentration, seasonality, relative humidity, and temperature as major predictors. The prediction errors also tended to increase at high concentrations, suggesting the need for complementary strategies, including additional variables and weighted learning for high-concentration cases. These results highlight the importance of evaluating performance according to the region and concentration range as well as overall accuracy. By presenting the predictive performance achievable using only ground observations, this study provides a baseline for evaluating the improvements achieved by integrating external data sources.

Keywords : Air Quality, $PM_{2.5}$, Machine Learning, LightGBM, Ground Observation

본 연구는 2025년도 강원대학교 대학회계 학술연구조성비의 지원을 받았으며, 기상청 재원으로 한국기상산업기술원의 기상기후데이터 융합분석 특성화대학원 사업 지원을 받아 수행되었음.
*Corresponding Author : Sungmin O(Kangwon National University)
email: sungmino@kangwon.ac.kr
Received June 4, 2026 Revised July 6, 2026
Accepted October 2, 2026 Published October 31, 2026

1. 서론

국내 미세먼지(PM: Particulate Matter)는 호흡기 및 심혈관계 질환 위험을 증가시키는 대표적인 대기오염물질로 알려져 있으며[1], PM 농도를 정확하게 예측하고 관리하는 것은 국민 건강 보호와 지역 대기관리 정책 수립에 필수적이다. 국내에서는 기상청과 한국환경공단을 중심으로 기상 및 대기질 지상관측자료가 장기간 축적되어 왔으며, 이를 활용한 통계모형, 머신러닝 및 딥러닝 기반 미세먼지 예측 연구가 활발히 수행되고 있다[2-6].

국내 선행연구에서는 머신러닝 및 딥러닝 기반의 PM 농도 예측 모형을 개발하거나, 서로 다른 알고리즘의 예측 성능을 비교한 바 있다[2,3]. 예를 들어, 트리 기반 머신러닝을 사용하여 서울 지역 미세먼지 단기 예측 성능을 평가하거나[5], 딥러닝과 화학수송모델을 결합한 전국 또는 예보권역 단위의 예측 시스템이 제안된 바 있다[7,8]. 그러나 대부분의 연구는 특정 도시, 개별 측정소 또는 예보권역 전체의 평균 성능에 초점을 맞추고 있어, 전국 규모의 행정구역별 예측 난이도와 지역별 예측 성능 차이를 체계적으로 분석한 연구는 미흡한 실정이다.

국내 미세먼지 농도는 기상 조건, 배출 특성 및 계절적 요인 등에 따라 지역별로 서로 다른 시공간적 변동성을 보인다[9]. 따라서 동일한 머신러닝 모델이라도 지역에 따라 예측 성능이 달라질 수 있으며, 전국 평균 성능만으로 모델의 적용성을 판단하는 것은 한계가 있다. 또한 행정구역 단위의 예측 오차와 주요 영향 변수를 함께 분석하는 것은 대기질 예측 모델의 지역별 활용 가능성과 한계를 평가하는 데 중요하다.

이에 본 연구는 국내 17개 시·도 행정구역의 지상관측 기반 기상자료와 $PM_{2.5}$ 관측자료를 활용하여 1시간 후 $PM_{2.5}$ 농도를 예측하는 머신러닝 모델을 구축하고, 지역별 성능 차이를 정량적으로 평가하는 것을 목표로 한다. 또한 최적 모델의 변수 중요도 분석을 통해 $PM_{2.5}$ 예측에 기여하는 주요 인자를 도출하고, 국내 지상관측자료 기반 머신러닝 예측 모델의 지역별 적용 특성을 제시하고자 한다.

2. 본론

2.1 관측자료

본 연구에서는 기상청 종관기상관측시스템(ASOS) 및 자동기상관측장비(AWS)의 기상자료와 한국환경공단 도시대기측정망(AirKorea)의 $PM_{2.5}$ 관측자료를 활용하였다. 분석 대상 기간은 2018년부터 2024년이며, 분석 대상 지역은 국내 17개 시·도 행정구역이다. 본 연구에서 사용한 행정구역별 ASOS 관측소는 1-4개, AWS 관측소는 4-143개, AirKorea $PM_{2.5}$ 관측소는 5-142개 범위였으며, 각 행정구역 내 가용 관측소의 시간별 자료를 평균하여 지역별 대표값을 산출하였다. 이때 각 시점에서 해당 행정구역 내 관측소 자료의 80% 이상이 가용한 경우에만 평균을 산출하였으며, 최종 분석자료는 지역별로 19,041-29,213개, 전체 412,888개이다. 기상 입력자료는 기온(TA, °C), 상대습도(HM, %), 풍속(WS, m/s), 풍향(WD, degree), 강수량(RN, mm)을 사용하였다.

예측 대상은 1시간 후 $PM_{2.5}$ 농도이며, 예측 기준 시점의 기상 및 $PM_{2.5}$ 농도(PM25_lag1)와 24시간 전 $PM_{2.5}$ 농도(PM25_lag24)를 입력변수로 사용하였다. 시간대별 및 계절별 주기성을 반영하기 위해 시간(hour)과 연중일(day of year)을 각각 사인 및 코사인 함수로 변환하였고, 지역별 고유 특성을 고려하기 위해 행정구역 정보는 one-hot encoding 방식으로 지역 더미 변수로 변환하였다. 따라서 최종 입력변수는 기상 변수 5개(TA, HM, WS, WD, RN), 시간 주기 변수 4개(hour_sin, hour_cos, doy_sin, doy_cos), 지역 더미 변수 17개, $PM_{2.5}$ 과거 변수 2개(PM25_lag1, PM25_lag24)이다.

2.2 머신러닝 모델 및 성능평가

$PM_{2.5}$ 예측 성능을 비교하기 위하여 지속성 기준모델(Persistence), Random Forest, XGBoost, LightGBM을 적용하였다. 지속성 기준모델은 예측 기준 시점의 $PM_{2.5}$ 농도가 1시간 후에도 유지된다고 가정하는 단순 예측 방식으로, 단기 미세먼지 예측 성능 비교를 위한 기준선으로 활용하였다.

$PM_{2.5}$ 농도는 직전 농도와의 시간적 자기상관성 뿐 아니라 다양한 기상 요인의 영향을 받으며, 이들 영향은 선형적·독립적으로만 나타나지 않는다[13]. 예를 들어 풍속 증가는 오염물질의 확산과 희석을 촉진할 수 있으나, 그 효과는 대기 안정도 및 습도 조건 등에 따라 달라질 수 있다. 따라서 $PM_{2.5}$ 단기 예측에서는 변수 간 상호작용과 비선형적 효과를 고려할 필요가 있다. Random Forest는 다수의 의사결정나무를 결합하는 bagging 기반 앙상블 알고리즘이며[10], XGBoost와 LightGBM은 gradient boosting 기반의 대표적인 트리 앙상블 알고리즘이다[11,12]. 이들 모델은 비선형 관계와 변수 간 상호작용을 학습하는 데 강점이 있어, 복합적인 영향 요인

에 의해 변동하는 $PM_{2.5}$ 농도 예측에 적합하다. 본 연구는 모델별 최적 하이퍼파라미터 탐색보다 동일한 입력자료 조건에서 지역별 예측 성능을 비교하는 데 초점을 두었으므로, 각 모델의 주요 하이퍼파라미터는 과도한 모델 복잡도를 피하는 범위에서 예비 실험을 통해 고정하였다. Random Forest의 주요 설정값은 n_estimators=200, max_depth=22, min_samples_leaf=2이며, XGBoost는 n_estimators=500, max_depth=6, learning_rate=0.05, subsample=0.85, colsample_bytree=0.85로 설정하였다. LightGBM은 n_estimators=700, learning_rate=0.04, num_leaves=63, subsample=0.85, colsample_bytree=0.85, min_child_samples=25로 설정하였다.

전체 자료는 시간 순서를 유지한 상태에서 초기 80%를 모델 학습자료(2018년 1월 - 2023년 7월)로 사용하고, 이후 20%를 검증자료(2023년 8월 - 2024년 12월)로 활용하였다. 무작위 분할은 미래 시점의 정보가 학습자료에 포함되는 정보 누수를 유발할 수 있으므로, 본 연구에서는 실제 예측 상황을 고려하여 시간순 분할을 적용하였다. 모델 성능은 평균절대오차(MAE: Mean Absolute Error), 평균제곱근오차(RMSE: Root Mean Squared Error), 결정계수(R^2)를 사용하여 평가하였다. 전체 검증자료에서 모델별 성능을 비교한 뒤, 대표 모델을 선정하여 행정구역별 성능과 변수 중요도를 산출하였다.

3. 분석결과

3.1 예측 모델 성능

Table 1은 검증자료에 대한 모델별 $PM_{2.5}$ 예측 성능을 보여준다. 지속성 기준모델은 MAE 1.843 $\mu g/m^3$, RMSE 3.431 $\mu g/m^3$, R^2 0.880의 성능을 보였으며, 이는 직전 시점의 $PM_{2.5}$ 농도만으로도 1시간 후 농도를 상당 부분 설명할 수 있음을 의미한다. 세 가지 머신러닝 모델은 모두 기준모델보다 우수한 성능을 보였다. 세 모델의 성능 차이는 크지 않으나, LightGBM은 MAE 1.757 $\mu g/m^3$, RMSE 3.111 $\mu g/m^3$, R^2 0.901로 가장 좋은 성능을 나타냈다. 따라서 본 연구에서는 LightGBM을 대표 모델로 선정하고 추가 분석을 수행하였다.

Table 1. Comparison of model performance on the validation dataset

Model	MAE	RMSE	R^2
LightGBM	1.757	3.111	0.901
XGBoost	1.761	3.123	0.900
Random Forest	1.758	3.174	0.897
Persistence	1.843	3.431	0.880

Fig. 1은 LightGBM 모델의 $PM_{2.5}$ 관측값과 예측값 비교 결과이다. 대부분의 값이 1:1 선 주변에 분포하여, 모델이 전반적인 농도 변동을 안정적으로 재현하고 있음을 확인할 수 있다. 고농도 구간의 예측오차를 정량적으로 평가하기 위해 $PM_{2.5}$ 예보등급에 기반하여 농도구간별 성능을 추가로 분석하였다(Table 2). 저농도 구간(≤ 15 $\mu g/m^3$)의 RMSE는 2.468 $\mu g/m^3$였으나, 나쁨 이상 구간(〉35 $\mu g/m^3$)의 RMSE는 7.183 $\mu g/m^3$로 증가하였다. 이는 고농도 사례의 발생 빈도가 낮고, 외부 유입, 국지적 배출 변화, 급격한 기상 변화 등의 영향이 행정구역 평균 입력자료만으로 충분히 설명되지 않기 때문으로 판단된다.

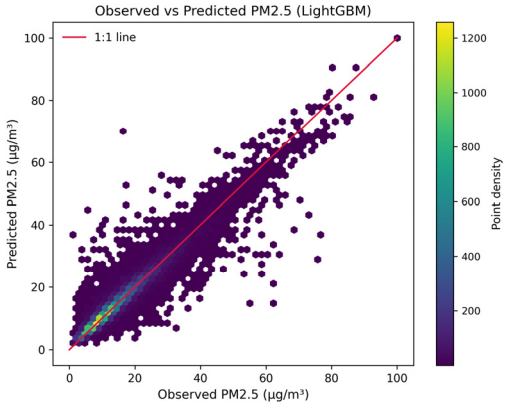


Fig. 1. Comparison between observed and predicted $PM_{2.5}$ concentrations using the LightGBM model. The color scale indicates point density.

Table 2. Model performance by $PM_{2.5}$ concentration range.

$PM_{2.5}$ range	Data size	MAE	RMSE	R^2
Low	55,775	1.363	2.468	0.425
Moderate	31,444	2.071	3.234	0.596
High	4,182	4.642	7.183	0.526

3.2 행정구역별 예측 성능

Fig. 2는 LightGBM의 행정구역별 $PM_{2.5}$ 예측오차를 보여준다. 지역별 RMSE는 경남(2.075 $\mu g/m^3$), 경북(2.108 $\mu g/m^3$), 전남(2.337 $\mu g/m^3$) 등에서 낮게 나타난 반면, 세종(4.209 $\mu g/m^3$), 서울(3.893 $\mu g/m^3$), 인천

(3.877 $\mu g/m^3$) 등에서는 상대적으로 높게 나타났다. 이는 동일한 머신러닝 모델과 입력변수를 적용하더라도 지역에 따라 $PM_{2.5}$ 예측 난이도가 달라질 수 있음을 보여준다.

국내 $PM_{2.5}$ 농도는 지역별 배출 특성, 도시화 수준, 지형 및 기상 조건에 따라 서로 다른 시공간적 변동성을 보인다[9,13]. 특히 서울과 인천은 장거리 이동 오염물질, 교통, 산업, 항만 등 다양한 요인의 영향을 받을 수 있으며[14,15], 행정구역 평균 입력자료만으로 이를 충분히 설명하기 어려워 상대적으로 높은 예측 오차로 이어졌을 가능성이 있다. 세종의 경우 대기질 관측소 수가 상대적으로 적어, 행정구역 평균값이 지역 내부의 농도 변동성을 충분히 반영하지 못했을 가능성이 있다. 다만 본 연구는 배출원 자료와 지형·토지이용 변수를 직접 포함하지 않았으므로, 지역별 오차 패턴에 대한 명확한 해석은 향후 추가 자료를 활용한 검증이 필요하다.

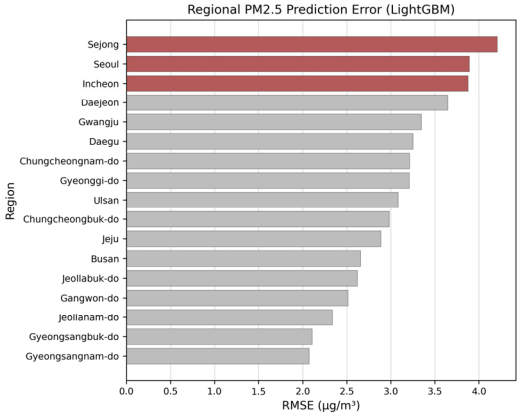


Fig. 2. Regional $PM_{2.5}$ prediction errors (RMSE) using the LightGBM model.

3.3 변수 중요도 분석

Fig. 3은 LightGBM 모델의 변수 중요도이다. 가장 중요한 변수는 PM25_lag1로 나타났으며, 이는 1시간 후 $PM_{2.5}$ 농도가 직전 시점의 농도와 강한 자기상관을 가진다는 점을 의미한다. 이러한 결과는 지속성 기준모델이 비교적 높은 성능을 보인 결과와도 일관된다.

두 번째로 중요한 변수는 연중일의 사인 변환값(doy_sin)으로 나타났다. 이는 $PM_{2.5}$ 농도 예측에서 계절성이 중요한 역할을 한다는 점을 의미한다. 국내 $PM_{2.5}$ 농도는 계절에 따라 뚜렷한 변동성을 보이며, 봄철 황사 및 장거리 이동 오염물질, 겨울철 대기 정체와 난방 수요, 여름철 강수 및 대기 확산 조건 등이 농도 변화에 영향을 줄 수 있다[9,13]. 따라서 연중일 기반의 주기 변수는 이러한 계절적 변동성을 간접적으로 반영한 것으로 해석할 수 있다.

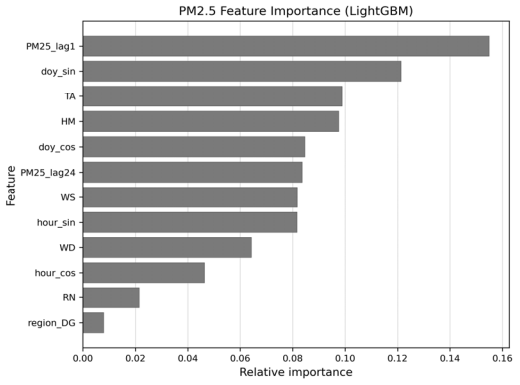


Fig. 3. Feature importance of the LightGBM model

기상 변수 중에서는 상대습도(HM), 기온(TA), 풍속(WS)이 상대적으로 높은 중요도를 보였다. 상대습도는 입자의 흡습성 성장 및 대기 중 입자상 물질의 농도 변화와 관련될 수 있으며, 기온은 대기 안정도와 오염물질 생성 및 확산 조건을 반영한다. 풍속은 오염물질의 확산과 희석에 직접적인 영향을 미치는 변수이다[16,17]. 한편 지역 더미 변수의 중요도는 상대적으로 낮게 나타났는데, 이는 단순한 지역 식별자보다 각 시점의 기상 조건, 계절성, $PM_{2.5}$ 농도 이력이 예측에 더 직접적으로 기여했음을 시사한다. 추가적으로 $PM_{2.5}$와 주요 기상변수 간 상관성을 검토한 결과, 기온(-0.244), 풍속(-0.180), 강수량(-0.157)은 음의 상관관계를 보였으며, 상대습도(-0.052)와 풍향(0.014)의 선형 상관은 상대적으로 작았다. 이는 단순 선형 상관만으로는 $PM_{2.5}$ 변동을 충분히 설명하기 어려우며, 기상 변수와 시간 주기성, 과거 농도 정보 간의 복합적인 관계를 함께 고려할 필요가 있음을 시사한다.

4. 결론

본 연구는 2018년부터 2024년 동안의 국내 17개 시·도 행정구역 단위의 시간별 지상관측자료를 활용하여 1시간 후 $PM_{2.5}$ 농도를 예측하고, 머신러닝 모델의 전체 및 지역별 예측 성능을 평가하였다. 분석 결과, 트리 기

반 머신러닝 모델은 지속성 기준모델보다 우수한 성능을 보였으며, 특히 LightGBM은 전체 검증자료에서 가장 우수한 예측 성능을 보였다.

주요 결과는 다음과 같다. 첫째, 지속성 기준모델도 단기 예측에서 높은 기준 성능을 보였으나, 트리 기반 머신러닝 모델은 기상 조건, 시간 주기성, 과거 $PM_{2.5}$ 농도 정보를 함께 반영함으로써 추가적인 예측 성능 향상을 보였다. 둘째, 행정구역별 RMSE는 2.075-4.209 $\mu g/m^3$ 범위로 나타나 지역별 예측 난이도에 뚜렷한 차이가 있음을 확인하였다. 또한 고농도 구간에서 예측 오차가 상대적으로 증가하였다. 셋째, 변수 중요도 분석에서는 직전 $PM_{2.5}$ 농도, 계절성, 상대습도, 기온 등이 주요 예측 요인으로 도출되었다.

본 연구결과는 전국 단위 $PM_{2.5}$ 예측 모델을 평가할 때 전체 평균 성능뿐만 아니라, 지역별 오차 수준과 고농도 구간에서의 예측 특성을 함께 검토할 필요가 있음을 시사한다. 특히 행정구역 평균 자료 기반 모델은 계산 비용과 해석 가능성 측면에서 실무적 장점이 있으나, 지역 내부의 공간 이질성과 국지적 배출 특성을 충분히 반영하기에는 한계가 있을 수 있다. 아울러 고농도 구간 예측 성능을 향상시키기 위해서는 고농도 발생 특성을 반영할 수 있는 추가 변수를 결합하고, 고농도 사례를 고려한 가중 학습 또는 구간별 모델링을 적용할 필요가 있다.

본 연구는 장기간 축적된 국내 관측자료를 기반으로 지역별 $PM_{2.5}$ 단기 예측 성능을 동일한 조건에서 비교했다는 점에서 의의가 있다. 또한 위성자료, 화학수송모델 산출물, 배출원 정보 등 추가적인 외부 자료를 결합하기 전, 관측자료 기반 머신러닝 모델만으로 확보할 수 있는 예측 성능과 지역별 한계를 제시했다는 점에서 향후 고도화된 $PM_{2.5}$ 예측 모델 개발을 위한 기준선으로 활용될 수 있다. 향후 연구에서는 측정소 단위 검증, 지역 내부 변동성 분석, 고농도 사례 중심의 오차 평가 등을 통해 행정구역 평균 기반 예측의 한계를 보완할 필요가 있다.

References

[1] V. A. Southerland, M. Brauer, A. Mohegh, M. S. Hammer, A. van Donkelaar, et al., "Global urban temporal trends in fine particulate matter ($PM_{2.5}$) and attributable health burdens: estimates from global datasets", *The Lancet Planetary Health,* Vol.6, No.2, pp.e139-e146, 2022,
DOI: https://doi.org/10.1016/S2542-5196(21)00350-8

[2] S. H. Seong, S. J. Kim, M. H. Ryu, "Comparative Study of Machine Learning Models for Fine Dust Prediction: Focusing on Domestic Occurrence Data", *Journal of Korea Innovation Society*, Vol.15, No.4, pp.339-357, 2020.

[3] Y. I. Kim, K. H. Lee, "Accuracy Analysis of Machine Learning Methods for Predicting PM Concentration", *Journal of Korean Society for Atmospheric Environment*, Vol.39, No.2, pp.149-164, 2023.
DOI:https://doi.org/10.5572/KOSAE.2023.39.2.149

[4] Y. J. Lee, I. S. Na, Y. G. Son, "Evaluation of Machine Learning Application on the Prediction of Particulate Matter Concentrations in Small/Medium-Sized City", *Journal of Korean Society of Environmental Engineers*, Vol.46, No.10, pp.537-550, 2024.
DOI:https://doi.org/10.4491/KSEE.2024.46.10.537

[5] B. Y. Kim, Y. K. Lim, J. W. Cha, "Short-Term Prediction of Particulate Matter (PM10 and PM2.5) in Seoul, South Korea Using Tree-Based Machine Learning Algorithms", *Atmospheric Pollution Research*, Vol.13, No.10, 101547, 2022.
DOI:https://doi.org/10.1016/j.apr.2022.101547

[6] H. S. Kim, I. Park, C. H. Song, K. Lee, J. W. Yun, et al., "Development of a Daily PM10 and PM2.5 Prediction System Using a Deep Long Short-Term Memory Neural Network Model", *Atmospheric Chemistry and Physics*, Vol.19, No.20, pp.12935-12951, 2019.
DOI:https://doi.org/10.5194/acp-19-12935-2019

[7] Y. S. Koo, H. Y. Kwon, H. S. Bae, H. Y. Yun, D. R. Choi, et al., "A Development of PM2.5 Forecasting System in South Korea Using Chemical Transport Modeling and Machine Learning", *Asia-Pacific Journal of Atmospheric Sciences*, Vol.59, No.5, pp.577-595, 2023.
DOI:https://doi.org/10.1007/s13143-023-00314-8

[8] J. W. Huh, J. S. Youn, P. M. Park, K. J. Jeon, S. Park, "Development of a Prediction Model for Daily $PM_{2.5}$ in Republic of Korea by Using an Artificial Neural Network", *Applied Sciences*, Vol.13, No.6, 3575, 2023.
DOI:https://doi.org/10.3390/app13063575

[9] S. O, J. W. Yoon, S. K. Park, "Estimating Hourly Ground-Level Aerosols Using Geostationary Environment Monitoring Spectrometer Aerosol Optical Depth: A Machine Learning Approach", *Atmospheric Measurement Techniques*, Vol.18, No.6, pp.1471-1484, 2025.
DOI: https://doi.org/10.5194/amt-18-1471-2025

[10] L. Breiman, "Random Forests", *Machine Learning*, Vol.45, No.1, pp.5-32, 2001.
DOI:https://doi.org/10.1023/A:1010933404324

[11] T. Chen, C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, San Francisco, CA, USA, pp.785-794, 2016.
DOI:https://doi.org/10.1145/2939672.2939785

[12] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, et al., "LightGBM: A Highly Efficient Gradient Boosting

Decision Tree", *Advances in Neural Information Processing Systems* 30, Curran Associates, Long Beach, CA, USA, pp.3146-3154, 2017.
DOI: https://doi.org/10.14249/eia.2019.28.3.183

[13] K. P. Nam, D. G. Lee, L. S. Jang, "Analysis of PM2.5 Concentration and Contribution Characteristics in South Korea according to Seasonal Weather Patterns in East Asia: Focusing on the Intensive Measurement Periods in 2015", *Journal of Environmental Impact Assessment*, Vol.28, No.3, pp.183-200, 2019.

[14] D. Choi, H. Lee, J. Kim, et al., "Distinct Meteorological Mode Associated with High-PM2.5 Episodes in Seoul, South Korea", *Journal of Applied Meteorology and Climatology*, Vol.62, No.7, pp.853-862, 2023.
DOI: https://doi.org/10.1175/JAMC-D-23-0016.1

[15] M. S. Kang, J. H. Kim, Y. Sunwoo, K. H. Hong, "Analysis of Fine Dust Impacts on Incheon and Busan Port Areas Resulting from Port Emission Reduction Measures", *Atmosphere*, Vol.16, No.5, 521, 2025.
DOI: https://doi.org/10.3390/atmos16050521

[16] Z. Li, J. Guo, A. Ding, H. Liao, J. Liu, et al., "Aerosol and boundary-layer interactions and impact on air quality", *National Science Review*, Vol.4, No.6, pp.810-833, 2017.
DOI:https://doi.org/10.1093/nsr/nwx117

[17] Y. Miao, J. Li, S. Miao, H. Che, Y. Wang, et al., "Interaction between Planetary Boundary Layer and PM2.5 Pollution in Megacities in China: A Review", *Current Pollution Reports*, Vol.5, No.4, pp.261-271, 2019.
DOI:https://doi.org/10.1007/s40726-019-00124-5

오 승 민(Sungmin O) [정회원]



- 2010년 8월 : 이화여자대학교 환경공학과 (공학석사)
- 2018년 6월 : 그라츠대학교 기상학과 (이학박사)
- 2018년 6월 ~ 2021년 8월 : 막스 플랑크 생지화학연구소 박사후연구원
- 2024년 9월 ~ 현재 : 강원대학교 첨단AI공학과 조교수

〈관심분야〉
기상기후, 대기환경, 딥러닝 모델링