Journal of Environmental Science International
35(4); 307~323; April 2026

pISSN: 1225-4517 eISSN: 2287-3503
https://doi.org/10.5322/JESI.2026.35.4.307

ORIGINAL ARTICLE

비전-정형 융합 인코더 기반 멀티모달 딥러닝을 활용한 초미세먼지 농도 예측

송영호 · 황수빈 · 백도경 · 송수영 · 남기전*
국립부경대학교 환경공학전공

Prediction of Fine Particulate Matter Concentration Using Multimodal Deep Learning Based on a Vision-Tabular Fusion Encoder

YoungHo Song, SuBin Hwang, DoGyeong Baek, SuYoung Song, KiJeon Nam*
Department of Environmental Engineering, Pukyong National University, Busan 48513, Korea

Abstract

Ambient fine particulate matter (PM2.5) has emerged as a global public health concern. The accurate forecasting of its concentration is essential for air quality management. Conventional air quality prediction models primarily rely on tabular data of meteorological variables and air pollutants from ground-based monitoring stations. Although sky images directly reflect the visual state of the atmosphere, research using these images alongside real-time meteorological tabular data remains relatively scarce. This study proposes a PM2.5 concentration prediction framework based on multimodal deep learning that effectively combines spatial visual information with meteorological factors. A pre-trained EfficientNet served as the vision encoder, while TabNet, based on sequential attention, served as the tabular encoder. In addition, to overcome the inherent limitations of simple point estimation, quantile regression was introduced to estimate the 90th percentile of prediction intervals, thereby providing robust uncertainty quantification for high-concentration events. Evaluated across diverse observation sites with varying geographic characteristics, the EfficientNet-TabNet fusion model consistently outperformed the single-modality baselines, achieving high predictive accuracy. Furthermore, Explainable AI (XAI) analysis was conducted to validate the model's physical interpretability. The results demonstrated that the vision encoder successfully captured optical degradation patterns caused by Mie scattering in distant skylines. In contrast, the tabular encoder accurately captured atmospheric dynamics, including regional ventilation driven by wind speed and aerosol hygroscopic growth driven by relative humidity. These findings quantitatively demonstrate the synergistic effects of multimodal fusion, offering a highly reliable, interpretable, and scalable approach for air quality forecasting.

Key words : Fine particulate matter, Multimodal deep learning, Vision, Tabular Data, Air quality forecasting

Received 11 March, 2026; **Revised** 8 April, 2026;
Accepted 13 April, 2026
***Corresponding author** : KiJeon Nam, Department of Environmental Engineering, Pukyong National University, Busan 48513, Korea
Phone : +82-51-629-6530
E-mail : kjnam@pknu.ac.kr

1. 서 론

초미세먼지(particulate matter 2.5, PM2.5)는 직경 2.5 μm 이하의 미세한 입자상 물질로, 급속히 발전하는 사회에서의 산업 및 기계적 인간 활동으로 인해 다량으로 발생하고 있다. PM2.5는 질병 및 사망 위험을 증가시키고 기타 다양한 건강 위협을 초래한다. 중금속, 황산염, 질산염 등을 함유한 미세입자는 특히 심혈관계 및 호흡기 질환, 암, 정신질환을 유발하거나 악화시키는 데 큰 영향을 미쳐서 인체에 심각한 악영향을 미칠 수 있다(Russell et al., 2009; Park et al., 2025). 또한, PM2.5 입자는 세포 내로 직접 침투하여 심각한 손상을 유발할 수 있다(Choi et al., 2026). 따라서, PM2.5는 PM10과 같은 더 큰 입자보다 인체에 대한 위해성이 높으므로, 정확한 농도 모니터링에 의한 장단기적인 대책 마련이 중요하다. 또한, PM2.5 농도를 정밀하게 예측함으로써 대기오염을 통제하고, 공중보건을 보호하며, 도시계획 의사결정을 지원하고, 그리고 기후 영향에 대한 통찰을 얻을 수 있다(Zhou et al., 2024).

세계적 연구에서는 PM2.5에 관련된 대기질 지수를 산출하고 고농도 PM2.5 발생을 사전에 예보하기 위한 고도화된 시스템 구축에 매진하고 있다(Zhang et al., 2018; Do et al., 2023). 전통적인 PM2.5 예측은 주로 3차원 화학 수송모델인 CMAQ (Community Multi-scale Air Quality)에 기반을 두고 있다. 이러한 물리적·화학적 모델은 대기 중 오염물질의 이류, 확산, 침적 및 2차 생성 과정을 편미분 방정식으로 계산하여 시공간적 대기질 모의가 가능하다(Lee et al., 2016; Kwak et al., 2025). 그러나, 화학 수송모델은 대기 역학 및 화학 반응의 복잡성으로 인해 막대한 컴퓨터 연산 자원과 긴 수행 기간이 요구되고, 입력자료인 배출원 인벤토리를 구축하는데 시간과 비용이 추가적으로 소요가 되며, 국지적 PM2.5 변화를 즉각적으로 반영하는데 한계가 있어 실제 관측값과 큰 오차를 유발할 수 있다(Huang et al., 2021).

최근 인공지능과 딥러닝 기술의 비약적인 발전으로, 관측 데이터를 기반으로 한 데이터 기반 예측 모델이 활발히 주목받고 있다(Zheng et al., 2015). 대부분의 딥러닝 기반 대기질 예측 연구는 기온, 풍향, 풍속, 상대습도 등 기상청 관측소에서 제공하는 정형 시계열 데이터를 주로 활용한다(Athira et al., 2018). 그러나 지상 센서 데이터는 관측 지점의 국지적 수치만을 제공하여, 대기 전체 공간의 광학적 특성이나 미세먼지로 인한 빛의 산란 정도를 직접적으로 설명하지 못한다. 뿐만 아니라, PM2.5 관측소가 해당 지역에 존재하지 않는다면 결국 예측 대상인 PM2.5의 농도를 직접적으로 예측하는 데 한계가 분명히 존재한다.

기상 관측 정형 데이터와 다르게 카메라를 통해 수집된 영상은 PM2.5에 의해 기인한 에어로졸의 산란 현상으로 인해 가시거리 감소, 채도 저하, 혼탁도 증가 등 대기의 상태에 대한 풍부한 시각적 정보를 포함하고 있다. 이미지 내 특징 추출 알고리즘과 multi-layer perceptron (MLP)을 결합한 모델로 도시 이미지로부터 PM2.5 농도를 예측한 바 있다(Chen et al., 2022). 교통 카메라 이미지로부터 PM2.5 농도를 예측하기 위하여 Residual Network (ResNet) 모델이 활용된 연구가 있다(Liu et al., 2024). 이전 연구들은 단일 장소의 이미지에 국한되어 진행되었으며, 다른 장소에 적용될 가능성에 대한 일반화에 관한 연구가 부족하다. 또한, 이미지 정보 단독으로는 대기 습도 및 바람에 의한 대기 정보를 기반으로 실제 PM2.5 농도를 명확히 추정하기 어렵다는 한계가 존재한다.

최신 이미지 기반 Artificial Intelligence (AI) 분야 연구에서는 멀티 모달리티(Multi-modality)에 관한 연구가 활발히 진행되고 있다. 기상 정형 데이터 또는 이미지를 단일 입력값으로 사용하는 단일 모달리티 모델에 기반한 독립적인 예측 및 해석은 PM2.5의 예측에 부분적인 정보만을 활용하게 된다. 반면, 멀티 모달리티 기반 예측 및 해석은 PM2.5의 변동성에 대해 보다 강건하고 신뢰할 수 있는 정보를 제공하게 된다. 동일한 PM2.5에 대해서도 서로 다른 모달리티에서 수집된 데이터는 각기 다른 표현 방식과 예측력을 가질 수 있고, 이는 서로 상호 보완적인 관계일 수 있다(Zhao et al., 2024). 그러나 멀티모달 데이터의 상호 보완적 특성을 이해하고 이를 효과적으로 융합하여 예측 정확도를 확보하는 것은 매우 어려운 과제로, 특히나 PM2.5에 대한 멀티모달 기반 예측 연구는 진행된 바가 드물다.

본 연구는 이미지와 기상 정형 데이터 두 가지 모달리티의 한계를 상호 보완하기 위해, 두 가지 모달리티 데이터를 융합적으로 활용하는 멀티모달 딥러닝 기반 PM2.5 예측 구조를 제안하였다. 이를 통해 모델은 시각적인 대기 혼탁도와 물리적인 기상 조건(습도, 풍향

등)을 활용하여 더욱 정밀한 PM2.5 농도를 추론할 수 있었다. 더욱이 PM2.5 및 기타 대기오염물질의 농도를 입력으로 사용하지 않아 대기오염 관측소의 부재에도 활용이 가능하다. 또한, 결정론적 예측을 넘어서 분위수 회귀를 결합하여 농도 예측의 신뢰 구간을 제공함으로써, 급격히 상승 증가하는 고농도 초미세먼지 발생 리스크에 대한 불확실성을 평가하였다.

2. 연구이론

2.1. 이미지 기반 대기질 예측 모델

대기 중의 초미세먼지(PM2.5) 입자는 태양광을 산란시키고 흡수하는 미 산란(Mie scattering) 효과를 유발하여, 카메라 영상 내에서 가시거리의 감소, 윤곽선의 흐려짐, 그리고 전반적인 채도의 저하를 일으킨다. 컴퓨터 비전 기반의 딥러닝 모델들은 이러한 미세한 광학적 변화와 픽셀 수준의 패턴을 포착하여 대기질 농도를 역산하는 데 탁월한 성능을 발휘한다.

가장 기초적인 형태인 합성곱 신경망(Convolutional Neural Network, CNN)은 이미지의 지역적 특징을 추출하기 위해 합성곱 층과 풀링 층을 반복적으로 거친다. 2차원 이미지 입력 x 에 대하여 지역적 특징을 추출하기 위한 커널 w 가 적용되는 합성곱 연산의 기본 수식은 다음과 같이 정의된다(Faraji et al., 2022).

$$y_{i,j} = \sum_m \sum_n w_{m,n} \cdot x_{i+m,j+n} + b \quad \cdots\cdots (1)$$

여기서 $y_{i,j}$는 출력되는 특징 맵 i, j 위치에서의 픽셀값이며, $w_{m,n}$은 커널 w의 m행, n열에 위치한 가중치, 그리고 b는 편향이다. 단순한 형태의 CNN은 네트워크의 층이 깊어질수록 학습 과정에서 기울기가 소실되거나 폭발하는 현상이 발생하여, 복잡한 대기 혼탁도를 정밀하게 학습하는 데 한계가 있다.

이를 극복하기 위해 제안된 ResNet은 잔차 연결이라는 혁신적인 구조를 도입하였다. 이는 특정 층의 입력값을 몇 개 층을 건너뛰어 출력값에 직접 더해주는 방식으로, 수식으로는 다음과 같이 표현된다(He et al., 2016).

$$y = F(x, W_i) + x \quad \cdots\cdots (2)$$

이때, x는 입력 벡터, W_i는 i번째 층의 학습 가능한 가중치, $F()$는 잔차 함수, y는 최종 출력벡터를 나타낸다. 잔차 함수 $F()$는 입력과 출력값의 차이를 나타내며, 기존 정보를 보존한 상태에서 학습이 필요한 잔차를 추출한다. 이 구조는 매우 깊은 네트워크에서도 정보의 손실 없이 안정적인 특징 추출을 가능하게 하며, PM2.5로 인한 옅은 안개 현상이나 미세한 질감 변화를 깊이 있게 추론하는 데 기여할 수 있다(Song et al., 2020).

최근에는 모델의 연산 효율과 예측 성능을 동시에 극대화한 EfficientNet이 대기질 영상 분석의 핵심 백본으로 주목받고 있다. 기존 모델들이 네트워크의 깊이, 너비, 해상도 중 하나만을 임의로 늘려 성능을 향상한 것과 달리, EfficientNet은 복합 계수 ϕ를 사용하여 세 가지 차원을 일정한 비율로 동시 확장한다(Tan et al., 2019).

$$d = \alpha^{\phi}, w = \beta^{\phi},$$
$$r = \gamma^{\phi} \; subject \; to \; \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad \cdots\cdots (3)$$

이때, d, w, r은 각각 네트워크의 깊이(depth), 너비(width), 입력 해상도(resolution)을 나타내며, ϕ는 전체 모델 크기를 조절하는 스케일링 계수이다. α, β, γ는 각 차원의 확장 비율을 나타내는 상수로, 격자 탐색(grid search)을 통해 결정된 상수이다. 이를 통해, d, w, r 세 요소를 균형 있게 확장하는 동시에 연산량이 약 두 배씩 증가하는 제약 조건을 만족하게 된다. EfficientNet은 적은 수의 파라미터로도 하늘 영상의 광역적 조도 변화와 국소적 픽셀 왜곡을 매우 정밀하게 추출할 수 있어, 기존 모델 대비 상대적으로 적은 파라미터 수로도 높은 예측 성능을 내며, 하늘 영상에 포함된 미세한 색상 왜곡이나 혼탁도를 정밀하게 추출할 수 있어 대기질 환경 데이터를 분석하는 데 적합한 고성능 백본 네트워크로 평가받고 있다. 최근 모바일이나 CCTV 이미지를 활용한 최신 PM2.5 예측 연구에서 최고의 성능을 입증하고 있다(Kamble et al., 2024).

2.2. 정형 데이터 기반 대기질 예측 모델

온도, 습도, 풍속, 강수량 등 기상청 관측소에서 수집되는 정형 데이터는 PM2.5의 이동, 축적, 소산 메커니즘을 설명하는 필수적인 물리적 맥락을 제공한다. 이

러한 수치형 데이터를 처리하기 위해 전통적으로 사용된 MLP는 입력 변수들을 가중치와 선형 결합한 후 비선형 활성화 함수를 통과시키는 전방향(feed-forward) 신경망 구조로, 여러 개의 은닉층(hidden layer)을 거쳐 입력 변수 간의 비선형적 관계를 학습한다. l번째 은닉층의 연산은 다음과 같이 정의된다(He et al., 2015).

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}) \cdots\cdots (4)$$

이때, $h^{(l)}$는 l번째 은닉층의 연산 결과, $h^{(l-1)}$는 이전 은닉층인 $l-1$번째 은닉층의 연산 결과, $\sigma()$는 활성화 함수, $W^{(l)}$는 l번째 은닉층의 가중치, $b^{(l)}$는 l번째 은닉층의 편향을 나타낸다. MLP는 훌륭한 비선형 근사기지만, 정형 데이터에 내재한 변수 간의 복잡한 상관관계나 불필요한 변수의 노이즈를 스스로 걸러내는 능력이 부족하다는 단점이 있다.

기존 모델의 한계 극복을 위해 자연어 처리에서 주로 쓰이던 어텐션 알고리즘이 정형 데이터 기반 모델에 도입되었다. 대표적으로 FT-Transformer (Feature Tokenizer + Transformer)는 연속형 및 범주형 정형 데이터들을 각각 독립적인 토큰 벡터로 변환한 후, 다중 헤드 어텐션을 통해 변수 간의 전역적인 상호작용을 학습한다(Gorishniy et al., 2021). FT-Transformer의 핵심 연산은 다음과 같다.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \cdots (5)$$

이때, Q (query), K (key), V (value)는 입력된 변수 토큰들로부터 선형 변환된 행렬이며, $\sqrt{d_k}$는 스케일링 벡터이다. Q와 K의 내적을 통해 변수 간 유사도를 계산하고, $\sqrt{d_k}$로 스케일링한 다음, softmax로 정규화하여 각 변수의 중요도를 얻는다. 이후 중요도에 V를 곱해줌으로써 중요한 변수일수록 모델에서 더 크게 반영되도록 한다. 이러한 메커니즘을 통해 특정 기상 조건에서 어떤 변수가 PM2.5 농도 변화에 가장 큰 영향을 미치는지 자동으로 학습할 수 있다.

의사결정 나무의 직관적인 해석력과 심층 신경망의 표현력을 완벽히 결합한 TabNet이 대기 환경 분석에 활발히 적용되고 있다. TabNet은 순차적 어텐션 기법을 사용하여, 의사결정 단계마다 예측 대상값에 가장 기여도가 높은 변수를 선택적으로 활성화하는 마스크를 생성한다. 단계 i에서의 마스크 $M[i]$는 다음과 같이 산출된다(Son et al., 2023).

$$M[i] = sparsemax(P[i-1] \cdot h_a[i-1]) \cdots\cdots (6)$$

이때, $P[i-1]$는 이전 단계까지 변수가 사용된 정도를 나타내는 페널티 척도이며, $h_a[i-1]$는 어텐션 트랜스포머의 출력값이다. 활성화 함수로 softmax 대신 sparsemax를 적용함으로써 불필요한 노이즈 기상 변수의 가중치를 0으로 만들어 희소한 마스크를 형성하며, 중요한 변수만을 선택적으로 반영할 수 있다. 이를 통해 대기 역학을 보다 강건하게 반영하는데 기여할 수 있다.

2.3. 멀티모달 딥러닝 모델 구조

단일 모달리티에 의존하는 예측 모델은 근본적인 정보의 결손을 피할 수 없다. 이미지만을 활용할 경우 고습도로 인해 단순히 짙은 안개와 실제 PM2.5 고농도 스모그를 명확히 구분하지 못하며, 반대로 기상 데이터만을 활용할 경우 관측소 인근의 불법 소각 등 돌발적인 국지적 대기오염 징후를 즉각적으로 포착하지 못한다. 이를 근본적으로 해결하기 위해, 본 연구는 비전 인코더(CNN, ResNet, EfficientNet)가 추출한 영상 기반의 시각적 대기 혼탁도(미시적 정보)와 정형 인코더(MLP, FT-transformer, TabNet)이 분석한 정형 기상 관측값(거시적 대기 역학 정보)을 융합하는 멀티모달 딥러닝(multimodal deep learning) 모델 아키텍처를 제안한다.

본 모델은 두 개의 독립적인 인코더 모듈과 하나의 융합 모듈로 구성된다. 각 모달리티의 독립적인 인코더에서는 특징을 완전히 추출하게 된다. 이후, 네트워크의 후반부에서 하나로 합치는 후기 융합 방식을 채택하였다. 두 네트워크에서 각각 도출된 이미지 특징 벡터와 기상 특징 벡터는 후기 융합 방식의 하나인 결합기법을 통해 하나로 융합된다. 이를 식으로 나타내면 아래와 같다.

$$h_{fusion} = [h_{img} \oplus h_{tab}] \cdots\cdots (7)$$

h_{fusion}은 융합 벡터로 h_{img}와 h_{tab}의 벡터 결합 연산 $\oplus$을 통해 계산된다. 여기서 h_{img}는 비전 인코더에서 도출된 1차원 이미지 임베딩 벡터, h_{tab}은 정형 인코더에서 도출된 기상 임베딩 벡터이다.

이후, 융합된 고차원 피처 벡터인 h_{fusion}은 MLP로 구성된 최종 회귀 헤드를 통과하며, 가중치 행렬 W_j와 결합하여 최종적인 PM2.5 농도 예측값 $\hat{y}$ 을 추론한다.

$$\hat{y} = W_f h_{fusion} + b_f \quad \cdots\cdots (8)$$

여기서 W_f와 b_f는 모델 학습 과정에서 손실 함수를 최소화하도록 데이터로부터 자동으로 학습되는 파라미터이다. 이미지 임베팅 벡터 h_{img}와 기상 임베딩 벡터 h_{tab}가 결합 된 h_{fusion}이 PM2.5 농도를 가장 잘 예측할 수 있도록, 두 모달리티 간의 상호작용을 효과적으로 반영하는 방향으로 최적화된다. 이러한 융합구조는 습도가 매우 높지만(h_{tab}), 영상에서 빛의 산란 패턴이 일반 수증기의 특성을 띤다(h_{img})는 식의 고도화된 교차 추론을 가능하게 만들어, 최신 연구들에서 증명된 바와 같이 융합적인 정보는 대기오염 데이터셋에서 단일 모델 대비 획기적으로 오차를 감소시키고 예측의 신뢰도를 높일 수 있다(Kalajdjieski et al., 2020; Wu et al., 2024).

3. 연구방법

본 연구는 데이터 수집 및 전처리, 단일 모달리티 베이스라인 평가, 멀티모달 융합 모델링, 그리고 예측 및 모델 해석의 네 단계로 구성된다. 첫 번째 단계에서는 연구 대상 지역의 대기 광학적 상태(빛의 산란, 가시거리 감소, 혼탁도 등)와 지면 환경이 명확히 관측되는 실시간 하늘 영상을 수집하였으며, 이와 시간상으로 완벽히 동기화된 정형 기상 관측 데이터(기온, 풍향, 풍속, 상대습도 등) 및 실제 지상 PM2.5 측정값을 확보하였다. 이후, 동일한 시계열 분할 조건에서 이미지만을 사용하는 비전 전용 모델(CNN, ResNet, EfficientNet 등)과 기상 변수만을 사용하는 정형 전용 모델(MLP, FT- Transformer, TabNet 등)을 각각 학습시켰다. 이를 통해 각 단일 모달리티 베이스라인의 예측 성능을 평가하고, 대기질 예측에 있어 비전 정보와 정형 데이터가 가지는 고유한 장점과 한계를 사전 분석하는 소거 연구(ablation study)를 수행하였다.

두 번째 단계에서는 앞선 단일 모달리티 실험을 통해 특징 추출 능력이 가장 뛰어난 것으로 입증된 아키텍처를 선별하여 멀티모달 융합 프레임워크를 구축하였다. 구체적으로 비전 인코더로는 파인 튜닝된 Efficient Net을, 정형 데이터 인코더로는 순차적 어텐션 기반의 TabNet을 활용하였다. 두 네트워크에서 각각 추출된 고차원 특성 벡터는 후기 융합 방식의 하나인 결합 계층을 통해 병합되며, MLP 헤드를 거쳐 최종 PM2.5 농도를 추론하도록 설계되었다. 모델의 정량적 예측 성능 평가는 실제 PM2.5 관측값과 모델 예측값을 비교하여 결정 계수(coefficient of determination, R^2), 평균 제곱근 오차(root mean squared error, RMSE), 평균 절대 오차(mean absolute error, MAE)를 산출하여 실시하였다.

마지막 단계에서는 제안된 융합 모델이 도출한 예측값의 블랙박스 특성을 해소하고, 대기 환경적 타당성을 부여하기 위한 설명 가능한 AI (Explainable AI, XAI) 분석을 수행하였다. 비전 모달리티의 경우, 비전 인코더가 농도 예측 시 영상 내 어느 영역에 가장 강하게 집중했는지를 히트맵으로 시각화하여 대기 혼탁도와의 연관성을 검증하였다. 정형 데이터 모달리티의 경우, 정형 인코더가 특정 농도 예측에 가장 중요한 핵심 변수로 선정한 기상 인자를 추출하고, 이를 습도에 따른 에어로졸의 흡습 성장이나 풍향에 따른 오염물질 확산 등 실제 기상학적 메커니즘과 교차 검증하였다. 더 나아가 단일한 점 추정 결과에 분위수 회귀모델링을 결합함으로써 예측값의 90% 신뢰 구간을 도출하고, 고농도 미세먼지 예측의 불확실성을 정량적으로 가시화하였다.

3.1. 이미지 및 정형 데이터 수집 및 전처리

본 연구에서 멀티모달 딥러닝 모델의 학습 및 성능 평가를 위해 구축한 이미지, 기상 변수, 그리고 예측 대상 변수인 PM2.5 농도 데이터의 세부 메타정보를 Table 1에 요약하여 나타내었다. 데이터 수집의 공간적 범위는 도심지 배출원의 특성과 광학적 가시거리의 차이를 포괄적으로 반영할 수 있도록 지리적·환경적 특성이 각기 다른 세 지점을 선정하였다. 구체적으로, 고밀도 교통량으로 인한 도심 협곡 현상이 나타나는 서울특별시의 서울역, 높은 곳에서 넓은 공간적 가시거리를 확보할 수 있는 초고층 빌딩인 롯데월드타워, 그리고 넓

Table 1. Image and tabular variables used to prediction PM2.5 concentration

	Variables	Unit	Data source
Imange variables	CCTV image	-	KT GiGAeyes
Tabulara variables	Temperature	%	Korea Meteorological Administration weather data service
	Wind direction	°	
	Wind speed	m/s	
	Precipitation	mm	
	Humidity	%	
Target variables	PM2.5	$\mu g/m^3$	AirKorea

은 면적의 호수가 인접해 있어 상대습도에 의한 대기 산란 영향이 뚜렷하게 관측되는 경기도 수원시의 광교호수공원을 연구 대상지로 삼았다. 데이터 수집의 시간적 범위는 2024년 4월부터 2025년 4월까지 총 1년간으로 설정하여, 계절적 요인에 따른 대기질 및 기상 패턴의 주기적인 변동성을 모델이 충분히 학습할 수 있도록 구성하였다.

비전 모달리티를 위한 하늘 및 도심 전경 이미지 데이터는 실시간 스트리밍 서비스인 KT GiGAeyes Live TV (https://www.youtube.com/@GiGAeyesLiveTV)에서 제공하는 고해상도 CCTV 영상을 활용하였다. 대기질 및 기상 측정 주기와 시간적 동기화를 맞추기 위해, 매시간 정각을 기준으로 해당 유튜브 스트리밍 화면을 캡처하여 크롤링하였다. 이와 동시에, 대기의 물리적 맥락을 제공하는 정형 데이터는 기상청 기상자료개방포털(data.kma.go.kr)에서 제공하는 기상관측 자료를 활용하여 기온(℃), 풍향(°), 풍속(m/s), 강수량(mm), 상대습도(%) 등 핵심 기상 인자를 수집하였다. 예측 대상인 1시간 단위의 실측 PM2.5 농도($\mu g/m^3$)는 한국환경공단에서 운영하는 에어코리아(AirKorea)의 해당 지역 도시대기측정망 데이터를 매칭하였다.

수집된 1년 치(총 8,760시간) 데이터 중, 모델의 학습 안정성을 저해할 수 있는 노이즈를 제거하기 위해 전처리를 수행하였다. CCTV 영상의 경우, 일시적인 스트리밍 끊김 및 화면 조정 오류가 발생한 시점을 배제하였다. 기상 및 PM2.5 정형 데이터 측면에서도 장비의 정기 점검이나 통신 오류로 인해 발생한 결측치 구간은 제거하였다. 이러한 결측 및 이상치 필터링을 거친 결과, 최종적으로 멀티모달 딥러닝 모델의 입력으로 사용된 유효 데이터 샘플 수는 서울역 6,227개, 롯데월드타워 7,344개, 광교호수공원 7,084개로 확정되었다.

선별된 유효 데이터는 각각의 멀티모달 인코더 특성에 맞게 추가 전처리 과정을 거쳤다. 이미지 데이터는 인코더 입력으로 사용하기 위해 해상도를 224×224 픽셀 크기로 리사이즈한 후, 픽셀값을 0에서 1 사이의 실수 단위로 스케일링 하였다. 기상 데이터는 각 기상 변수 간의 단위 및 스케일 차이로 인한 특정 변수에 가중치가 편향 학습되는 현상을 방지하고자, 모든 수치 데이터에 대해 표준화(standardization)를 수행하였다. 표준화는 데이터셋을 변환하여, 변환된 데이터의 평균이 0이고 표준편차가 1이되도록 한다(Thara et al., 2019). 전처리된 이미지 및 기상 데이터셋은 이후 단일 모달리티 베이스라인 모델 및 최종 멀티모달 딥러닝 모델의 훈련, 검증, 평가를 위한 입력 데이터로 활용되었다.

3.2. 멀티모달 딥러닝 모델 구조 설정 및 모델링

본 연구에서는 전처리가 완료된 이미지 및 정형 기상 데이터를 활용하여 PM2.5 농도를 추론하는 단일 모달리티 베이스라인 및 멀티모달 딥러닝 모델을 구축하였다. 데이터 내의 시간적 의존성 및 정보 손실을 방지하기 위해 각 지역별 2024년 4월부터 2025년 4월까지의 전체 데이터를 시간순으로 정렬하여, 앞선 60%의 데이터를 모델 학습 자료로, 이어지는 20%를 과적합 방지 및 하이퍼파라미터 튜닝을 위한 내부 검증 자료로, 마지막 20%를 최종 모델 성능 평가를 위한 테스트 자료로 분할하였다. 대기오염 데이터는 대부분의 기간이 저농도 또는 보통 수준에 머무르고 극단적인 고농도 사례는 드물게 발생하는 심각한 클래스 불균형 문제가 존재한다. 모델이 다수의 저농도 패턴에만 편향되어 학습되는 것을 방지하기 위해, 학습 데이터 내 고농도 구간에 대하여 회귀용 오버샘플링을 수행하였다. 이와 더불어, 손실 함수 계산 시 고농도 타겟 샘플의 오차에 더 큰 페

Table 2. Summary of model architectures and training hyperparameters for the multimodal deep learning model predicting PM2.5 concentration

Category	Component	Descriptions
Image encoder	CNN	Convolution(5x5 kernel, 2 stride) - Convolution(3×3 kernel, 2 stride) - Convolution(3×3 kernel, 2 stride) - Global average pooling – Linear
	ResNet	ResNet50 (pre-trained) - Linear
	EfficientNet	EfficientNet_b0 (pre-trained) - Linear
Tabular encoder	MLP	Linear(128) - Linear (128)
	FT-Transformer	Feature tokenizer - Transformer encoder (64 dimension, 4 head, 2 layer 2)
	TabNet	Sequential attention and feature transformer (64 dimension, 3 step)
Hyper-parameters	Optimizer	AdamW (weight decay = 0.0001)
	Learning rate	0.001
	Batch size	32
	Epoch	100
	Scheduler	ReduceLROnPlateau
	Early stopping	Patience 20

널티를 부여하는 가중치 스케일링 기법을 동시 적용하여 예측의 강건성을 확보하였다.

모달리티별 모델 구조 설정에 있어서, 비전 모달리티의 CNN 모델은 무작위 가중치에서 초기화하여 전체 학습을 진행한 반면, ResNet과 EfficientNet은 대규모 이미지 데이터 셋으로 사전 학습된 가중치를 백본으로 하여, 본 연구에서 사용된 이미지에 맞게 미세조정을 수행하였다. 정형 데이터 모달리티 모델인 MLP, FT-Transformer, TabNet은 별도의 사전 학습 없이 전체 학습을 수행하였다. 최종적인 멀티모달 딥러닝 융합 모델은 비전 모달리티와 정형 모달리티에서 각각 도출된 고차원 특징 벡터를 결합하고 MLP 기반의 예측 헤드를 통과하도록 설계되었다.

본 연구의 멀티모달 프레임워크는 결정론적인 단일 PM2.5 값 추정의 한계를 극복하고자 분위수 회귀(quantile regression)를 도입하였다. 목적 함수로 평균 제곱 오차(MSE) 대신 핀볼 손실(pinball loss)을 적용하여 $\tau \in \{0.05, 0.50, 0.95\}$의 세 가지 분위수를 동시에 예측하도록 구성하였으며, 이를 통해 중앙값 예측값과 함께 90% 신뢰 구간을 산출하여 예측의 불확실성을 정량화하였다. 모델 학습 최적화를 위해 AdamW 옵티마이저(학습률 0.001, Weight decay 0.0001)를 사용하였고, 검증 손실이 정체될 때 학습률을 감소시키는 ReduceLROnPlateau 스케줄러와 과적합 시 학습을 조기 종료하는 Early Stopping (Patience=20)을 적용하였다. 단일 및 멀티모달 딥러닝 모델의 구조 및 하이퍼파라미터 설정에 관한 구체적인 내용은 Table 2에 요약하여 나타내었다.

본 연구에서 도출된 τ=0.50 중앙값 예측 결과를 바탕으로 모델의 정량적 예측 성능을 평가하기 위해, 결정 계수(coefficient of determination, R^2), 평균 제곱근 오차(root mean squared error, RMSE), 평균 절대 오차(mean absolute error, MAE)를 평가 지표로 사용하였다. R^2는 시계열 모델의 분산 설명력을 나타내는 지표로 1에 가까울수록 모델이 예측 대상인 PM2.5의 변동성을 잘 설명한다는 것을 의미한다. RMSE와 MAE는 모델의 예측값과 실제 측정된 데이터 간의 물리적 오차 크기를 나타내며, 0에 가까울수록 모델의 성능이 우수함을 지시한다. 각 평가 지표의 계산식은 다음과 같다.

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y_i} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \cdots\cdots (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n}} \quad \cdots\cdots (10)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y_i}\right| \quad \cdots\cdots (11)$$

여기서 y_i는 i번째 측정값, $\hat{y_i}$는 예측 모델을 통한 i번째 예측값, $\bar{y}$는 측정값의 평균을 나타낸다.

3.3. 멀티모달 딥러닝 모델의 설명가능(XAI) 분석

최근 딥러닝 모델의 복잡성이 증가함에 따라 예측 결과의 도출 과정을 이해하고 모델의 신뢰성을 확보하기 위한 XAI 기법의 중요성이 대두되고 있다. 본 연구에서 제안한 멀티모달 딥러닝 모델의 예측 결과에 대한 대기환경적 타당성을 검증하기 위해, 비전 모달리티와 기상정형 모달리티 각각의 특성에 따른 해석 기법을 적용하여 모델의 예측 과정을 시각화하였다.

첫째로, 비전 모달리티의 해석을 위해 모델의 순전파 과정에서 고유하게 도출되는 활성화 맵을 이용하여 시각화하였다. 순전파 활성화 맵은 입력 이미지에 대한 네트워크의 직접적인 반응을 왜곡 없이 보여주는 장점이 있다(Wang et al., 2020). 본 연구에서 사용하는 비전 인코더 내부의 주요 합성곱 모듈에 순전파 훅을 이용하여 학습 및 추론 중 도출되는 특징 맵을 포착하고, 최소-최대 정규화로 원본 이미지의 공간 해상도 규격으로 맞춘다. 이미지 내 임의의 위치 i, j에서의 최종 활성화 맵 $A_{i,j}$은 아래와 같이 계산된다(Zagoruyko et al., 2016; Woo et al., 2018).

$$A_{i,j} = Normalize\left(\frac{1}{C}\sum_{k=1}^{C} M_{k,i,j}\right) \quad \cdots\cdots (12)$$

이때, $M_{k,i,j}$는 특징 맵의 k번째 채널 특성 맵 값이다. 이와 같이 추출된 활성화 맵은 원본 이미지 위에 히트맵 형태로 표현되어, 모델이 미세먼지 농도 추론 시 대기의 혼탁도, 빛의 산란광, 혹은 가시거리 저하 현상 등 이미지 픽셀 내 어느 지점에 가장 강하게 집중하는지를 설명할 수 있다.

둘째로, 기상 관측 데이터로 구성된 정형 모달리티의 해석을 위해 각 기상 변수가 최종 농도 예측에 영향을 준 개별 기여도를 그래디언트 기반으로 계산하였다. 정형 모달리티의 출력값에 대하여 역전파 연산을 수행하여 입력 변수에 대한 편미분 벡터를 구한다. 다음으로 특정 샘플의 기상 변수별 기여도는 원본 입력 텐서와 산출된 그래디언트 텐서의 요소별 곱으로 정의하여 계산한다(Ancona et al., 2017).

$$\mathrm{attr} = \mathrm{x_{tab}} \odot \nabla_{\mathrm{x_{tab}}} y = \mathrm{x_{tab}} \odot \frac{\delta y}{\delta \mathrm{x_{tab}}} \quad \cdots\cdots (13)$$

여기서, attr은 특정 샘플의 기상 변수별 기여도, $\mathrm{x_{tab}}$은 원본 입력 텐서, y는 정형 모달리티 모델의 출력 값이다. 산출된 기여도 attr는 히트맵을 통해 해당 시점의 예측값 도출에 기상 요인 중 어떤 변수가 PM2.5 예측에 강한 영향을 주었는지 알 수 있다. 이를 통해 멀티모달 딥러닝 모델의 예측 결과가 대기 물리학적 메커니즘과 통계적으로 부합한 것을 증명하며, 본 연구에서 제시하는 예측 시스템의 신뢰성을 확인 할 수 있다.

4. 연구결과

4.1. 이미지 및 정형 데이터 분석

본 연구에서 수집한 이미지 데이터 중 서울의 롯데월드타워에서 촬영된 2025년 1월 20일 13시와 2월 7일 13시 이미지를 각각 Fig. 1(a)와 (b)에 나타내었다. 1월 20일은 지상 PM2.5 실측 농도가 80 μg/m³로 대기환경기준 매우 나쁨 단계에 속해 있었으며, 반대로 2월 7일에는 14 μg/m³로 좋음 단계에 속해 있었다. Fig. 1에서 뚜렷하게 관측되듯이, PM2.5 농도가 매우 높을 경우 광학적 영상 열화가 발생하여 멀리 있는 빌딩의 윤곽 및 산의 능선이 전혀 보이지 않는 등 시정이 심각하게 감소하는 특징을 나타낸다. 미세입자가 시정거리에 이처럼 명확한 영향을 주는 이유는 대기 에어로졸에 의한 빛의 산란 및 흡수 현상 때문이다. PM2.5의 입자 크기는 가시광선의 파장대(0.4~0.7 μm)와 유사하여, 빛이 대기를 통과할 때 입자와 강하게 상호작용을 하며 흩어지게 된다. 결과적으로 목표물에서 반사된 빛이 카메라에 도달하기 전 대기 중에서 산란하여 이미지의 대비와 해상도가 크게 훼손된다(Chen et al., 2021; Yang et al., 2022). Fig. 1(a)의 짙은 연무 현상은 딥러닝 비전 인코더가 학습할 수 있는 직관적인 특징 벡터를 제공할 수 있다.

Fig. 2는 본 연구에서 제안하는 멀티모달 딥러닝 기반 PM2.5 농도 예측 모델의 정형 데이터 입력값으로 사용되는 세 지역에서 측정된 5가지 기상 변수-기온, 풍향, 풍속, 강수량, 습도-의 통계적 특성을 히스토그램

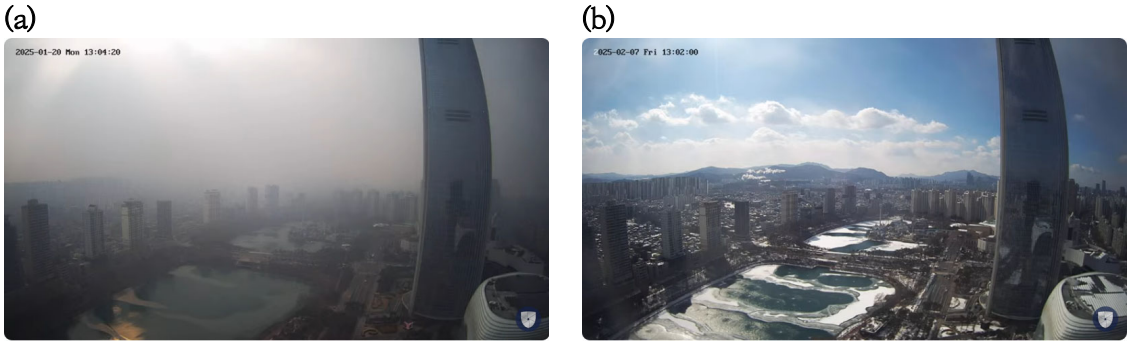

Fig. 1. Atmospheric images of Seoul under different PM2.5 concentrations: (a) at 13:00 on January 20, 2025, with a Very poor air quality level (80 $\mu g/m^3$); (b) at 13:00 on February 7, 2025, with a Good air quality level (14 $\mu g/m^3$).

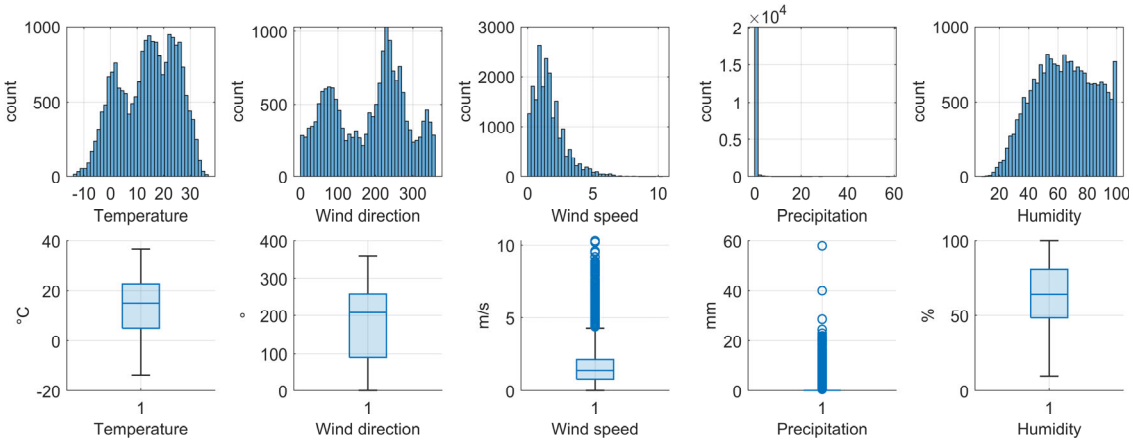

Fig. 2. Comprehensive statistical analysis of meteorological input features from three regions for the PM2.5 prediction model.

과 박스 플롯으로 시각화하였다. 각 변수의 분포 형태는 계절적 특성과 PM2.5의 물리적 거동을 설명할 수 있는 기초 자료가 될 수 있다. 먼저, 기온은 약 -15℃에서 35℃에 이르는 넓은 범위를 가지고 있으며, 히스토그램 상에서 0℃, 15℃, 그리고 20℃ 부근을 중심으로 다봉 분포를 가지고 있다. 이는 1년간의 데이터 수집을 통해 동절기 및 하절기의 계절적 온도 변동성이 반영되었음을 시사한다. 풍향은 80°(동풍 계열), 240°(서남 서풍 계열), 그리고 360°(북풍 계열) 부근에서 빈도를 보이는 삼봉 분포를 가진다. 계절풍 및 국지적 풍계가 복합적으로 작용함을 의미하며, 특히 서풍 계열의 높은 빈도는 외부 오염물질의 장거리 수송 가능성을 모델이 학습할 수 있는 특징이다(Nam et al., 2021). 풍속의 경우 2 m/s 부근에 데이터가 집중된 강한 우측 꼬리 분포를 가지고 있으며, 그 이하의 풍속에 많은 데이터가 있어 대기 정체에 의한 고농도 초미세먼지 환경을 유추할 수 있다. 박스 플롯 상에 많은 이상치를 가지고 있는데, 이는 간헐적인 강풍 현상으로 인한 PM2.5의 대기 확산을 나타낼 수 있다. 강수량은 대부분의 데이터가 0 mm에 분포되어 있어, 비가 내리는 횟수 자체는 드문 것을 알 수 있다. 다만, 강수는 대기 중의 에어로졸 상태로 존재하는 PM2.5를 씻어 내리는 세정효과를 가지고 있으므로 정형 모델에서 포착해야 하는 비선형적 데이터라고 할 수 있다(Fujino et al., 2022). 상대습도는 20%에서 100%까지 넓게 분포하며, 60~80% 구간에서 최대치를 형성하는 좌측 꼬리 형태를 띤다. 습도가 높은 환경일수록 미세먼지 입자가 수분을 흡수하여 크기가 성장하는 흡습 성장이 발생하여, PM2.5로 인한 시정거리 악화를

Table 3. Comparison of PM2.5 prediction performance among single-modality (vision, tabular) models (Bold values indicate the best performance in each modality)

Modality	Model	RMSE	MAE	R^2
Image data	CNN	13.21	9.71	0.24
	ResNet	12.81	8.73	0.28
	EfficientNet	**10.06**	**6.97**	**0.56**
	Average	12.03	8.47	0.36
Tabular data	MLP	15.84	11.39	-0.09
	FT-Transformer	16.17	11.51	-0.13
	TabNet	**15.79**	**11.31**	**-0.08**
	Average	15.93	11.40	-0.10

Table 4. Performance of the multimodal deep learning model in predicting PM2.5 concentrations by observation site

Region	Sub-region	RMSE	MAE	R^2
Total		8.24	5.83	0.70
Seoul	Seoulstation	7.36	5.35	0.73
	LotteWorldTower	7.77	5.50	0.56
Suwon	Gwanggyo	9.72	6.87	0.71

유발하는 요인이다(Won et al., 2021).

수집된 기상 정형 데이터는 PM2.5의 대기 환경학적 및 물리적 역학을 설명할 수 있는 변수라고 할 수 있다. 이미지만을 활용한 단일 모달리티 모델은 고습도나 대기 정체 상황에서 예측 성능이 저하될 우려가 있다. 그러나 상대습도 및 풍속과 같은 정형 데이터를 병행으로 함께 입력하는 멀티모달 딥러닝 융합 모델의 경우 이러한 한계를 극복하고 다양한 기상 조건에서 모델의 PM2.5 농도 예측 강건성을 항상 시킬 수 있을 것으로 판단된다.

4.2. 단일 모달리티 모델 기여도 분석

본 연구에서는 멀티모달 딥러닝 융합 모델의 성능 향상 정도를 객관적으로 비교 분석하기 위해, 이미지 데이터와 기상 정형 데이터를 각각 독립적으로 사용한 단일 모달리티 모델들의 전체 지역 PM2.5 농도 예측 성능을 통합적으로 평가하였다. 이미지 기반의 비전 모델(CNN, ResNet, EfficientNet)과 정형 데이터 기반의 예측 모델(MLP, FT-Transformer, TabNet)에 대한 RMSE, MAE, 그리고 R^2 평가 결과는 Table 3에 나타내었다.

이미지를 단독으로 이용한 비전 모델 그룹에서는 EfficientNet이 RMSE 10.06 $\mu g/m^3$, MAE 6.97 $\mu g/m^3$로 오차를 가장 크게 줄였으며, 0.56의 가장 높은 R^2 값을 기록하여 유의미한 예측 성능을 입증하였다. 특히 EfficientNet은 가장 낮은 성능을 보인 CNN 대비 RMSE를 약 23.8% 감소시켰으며, R^2 값은 0.24에서 0.56으로 2배 이상 향상하며 뚜렷한 PM2.5 농도 예측 성능 개선 효과를 보였다. EfficientNet은 전통적인 CNN이나 망을 깊게 쌓는 데 집중한 ResNet과 달리, 네트워크의 깊이, 너비, 입력 이미지의 해상도를 최적의 비율로 동시 확장하는 복합 스케일링 기법을 적용한 비전 모델이다(Tan et al., 2019). 이러한 구조적 특성 덕분에 한정된 연산량으로도 대기 이미지 전반에 걸친 거시적인 안개 현상부터, 원거리 구조물 윤곽선의 미세한 광학적 열화 현상까지 다양한 크기의 특징을 가장 정밀하게 추출할 수 있었다. 비전 모델 그룹의 평균 R^2는 0.36을 나타내었으며, 이는 기상 변수 없이 이미지만으로도 PM2.5 농도의 일정 부분을 설명할 수 있음을 시사한다. 이는 대기의 시각적 혼탁도가 PM2.5 입자에 의한 빛의 미 산란을 직접적으로 반영하는 강력한 대리 지표임을 나타낸다.

반면, 기상 정형 데이터만을 단독으로 학습한 모델 그룹(MLP, FT-Transformer, TabNet)은 세 모델 모두 R^2 값이 음수(-0.08 ~ -0.13)로 나타나며 예측 성능

이 낮았다. 정형 데이터 모델들의 평균 RMSE (15.93 $\mu g/m^3$)는 비전 모델 그룹의 RMSE (12.03 $\mu g/m^3$)보다 약 32.5% 높게 나타나 오차가 큼을 보였다. 회귀 분석에서 R^2 값이 음수인 것은 모델의 예측 결과가 실제 PM2.5 농도의 평균값을 추종하는 것보다 오차가 크고, 주어진 기상학적 입력 변수만으로는 PM2.5 농도의 분산을 설명하기에 부족함을 의미한다. 이러한 결과는 기상 관측 데이터가 초미세먼지의 농도를 결정짓는 환경적인 조건 정보는 제공하나, 실제 대기 중에 존재하는 초미세먼지의 직접적인 양에 대한 정보는 포함하고 있지 않기 때문이다. PM2.5 예측에 있어 기온, 풍향, 풍속, 습도, 강수량과 같은 기상 관측값 단일 모달리티는 한계가 있음을 알 수 있다. 기상 관측 데이터는 미세먼지의 확산, 정체, 소산과 같은 대기 역학 조건에 영향을 주는 간접적인 지표지만, 실제 PM2.5의 발생 및 동적 증감에 대한 직접적인 정보를 포함하지 않는다. 동일 기상 조건이더라도 오염물질 유입 여부에 따라 실제 PM2.5 농도는 달라질 수 있다. 따라서, 인근 지역으로부터의 대규모 오염물질 유입 혹은 국지적 지역 내의 급격한 변동을 포함할 수 없어 PM2.5 예측 오차가 발생한 것으로 추론된다.

결과적으로, 단일 모달리티 기여도 분석을 통해 이미지 데이터의 PM2.5 농도 예측에 대한 가능성을 입증함과 동시에, 단일 정형 데이터 모델의 한계를 통계적으로 확인하였다. 이를 통해 비전 및 기상 정형 데이터 모달리티를 결합하여 시각적 현상과 대기 물리적인 정보를 상호 보완하는 멀티모달 프레임워크를 구성해야 함을 알 수 있다.

4.3. 멀티모달 딥러닝 모델 기반 PM2.5 농도 예측

멀티모달 딥러닝 모델을 구성하기 위하여 각 단일 모달리티 PM2.5 예측 모델에서 성능이 좋은 EfficientNet과 TabNet을 각 인코더로 사용하고, 두 인코더의 출력을 결합하여 PM2.5 농도를 예측하도록 하였다. 그리고 제안한 멀티모달 딥러닝 융합 모델(EfficientNet-TabNet)의 범용성과 강건성을 검증하기 위해, 지리적·환경적 특성이 서로 다른 세 관측 지점(서울역, 롯데월드타워, 광교호수공원)의 테스트 데이터 셋에 대한 예측 성능을 독립적으로 평가하였다. 정량적 예측 성능 지표인 RMSE, MAE, 그리고 R^2 산출 결과는 Table 4에 요약하여 나타내었다.

EfficientNet과 TabNet을 각각 인코더로 구성한 멀티모달 딥러닝 모델은 전체 지역에 대해 RMSE 8.24 $\mu g/m^3$, MAE 5.83 $\mu g/m^3$, R^2 0.70의 예측 성능을 보였다. 이는 Table 3에 제시된 단일 모달리티 모델 중 가장 우수한 성능을 보인 EfficientNet과 TabNet 대비, RMSE 기준으로 각각 22.08%와 91.63% 향상된 결과이다. 이러한 결과는 이미지 데이터에서 추출된 대기의 시각적 혼탁도 정보와 정형 기상 데이터 기반의 기상학적 정보가 결합 될 때 PM2.5 예측 성능이 크게 향상될 수 있음을 나타낸다. 특히 단일 모달리티 모델에서 성능이 제한적이었던 기상 변수는 이미지 특징과 결합 되면서, 단순 시각 정보만으로는 파악하기 어려운 대기 확산 및 정체 조건을 보완하여 예측 정밀도를 향상시킨 것으로 분석된다.

관측 지점별 성능을 살펴보면, 서울역 지점에서는 RMSE 7.36 $\mu g/m^3$, MAE 5.35 $\mu g/m^3$와 함께 R^2 0.733이라는 정확한 분산 설명력을 보였다. 호수가 인접하여 습도 변화가 높은 광교 지점에서도 R^2 0.71를 달성하여, 본 모델이 도심 뿐만 아니라 다양한 환경적 변수 앞에서도 높은 신뢰도로 PM2.5 농도를 추론할 수 있음을 입증했다. 롯데월드타워의 경우 오차 지표(RMSE 7.77 $\mu g/m^3$, MAE 5.50 $\mu g/m^3$)는 서울역과 유사한 수준으로 낮게 유지되었으나 R^2 값은 0.56으로 상대적으로 낮게 산출되었다. 이는 테스트 기간 중 국지적으로 높이 발생한 고농도 PM2.5 이상치의 영향이며, 전반적인 예측 오차 자체는 안정적인 수준으로 통제되었음을 알 수 있다.

멀티모달 딥러닝 모델의 예측 결과를 시각적으로 평가하기 위하여 실제 PM2.5 농도와 예측 결과를 분위수 회귀와 함께 Fig. 3에 나타냈다. 본 연구에서는 지역별 가용 데이터의 시계열적 특성을 유지하기 위해 전체 데이터 중 시간순으로 마지막 20%를 테스트 자료로 할당하였다. 다만, 데이터 수집 과정 중 CCTV 기기 점검 작업으로 인한 간헐적 결측 구간이 발생함에 따라, 지역별로 최종 성능 평가에 활용된 시점과 종점에 차이가 발생하였다. 그래프에 파란색 실선으로 표시된 바와 같이 모델의 PM2.5 예측 중앙값(τ=0.50)은 검은색 점으로 표시된 실제 PM2.5 농도의 시계열적 변동성을 높은 정확도로 추종하는 것을 알 수 있다. 특히 대기질 관리에 있어 중요한 한국 대기환경기준의 나쁨 수준인 36 $\mu g/m^3$ 이상의 고농도에서도 모델의 정확도가 유지되었다. Fig. 3(a) ~ (c)에 나타난 1월 ~ 2월 매우 나쁨 수준인 76 $\mu g/m^3$ 이상으로 급증한 경우에도 딥러닝 모델에

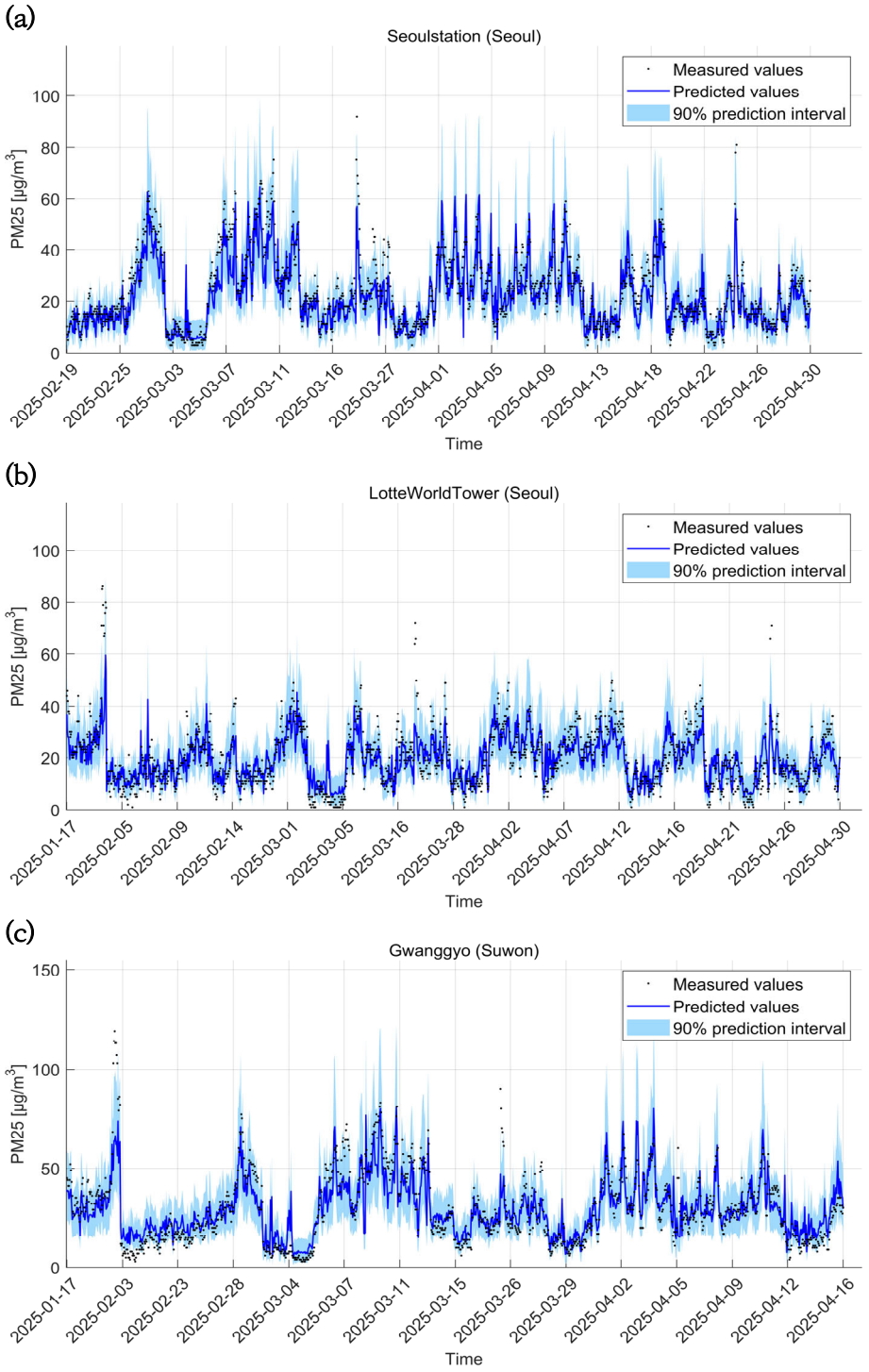

Fig. 3. Time-series prediction results of PM2.5 concentration and 90% prediction intervals of the multimodal model by observation site: (a) Seoul Station, (b) Lotte World Tower, (c) Gwanggyo Lake.

서 발생 가능한 과소평가 현상 없이 실제 농도를 예측할 수 있음을 보였다.

추가적으로, Fig. 3의 그래프에 하늘색 음영으로 표시된 영역은 분위수 회귀를 통해 계산된 τ=0.05와 τ=0.95의 90% 예측 신뢰 구간을 나타낸다. 대기 상태가 양호하고 PM2.5가 저농도로 유지되는 구간에 대해서는 신뢰 구간 밴드의 폭이 좁게 형성되어 모델의 예측 확실성이 높음을 나타낸다. 반면, PM2.5 농도가 급증하는 지점에 대해서는 신뢰 구간 밴드가 넓게 형성되어 잠재적인 예측 리스크에 대한 불확실성을 선제적으로 가시화한다. Fig. 3(a)와 같이 서울역 지역에서 3월 말 경 PM2.5 농도가 급증하여 발생한 76 $\mu g/m^3$ 이상의 매우 나쁨 수준의 경우, 중앙값은 정확한 최고점에 도달하지 못했으나 90% 신뢰구간이 해당 범위를 포괄함으로써 고농도 발생 가능성을 나타내었다. 결론적으로, 이미지와 기상 데이터를 융합한 멀티모달 딥러닝 모델은 단일 데이터 추정 방식의 한계를 넘어 실제 대기의 변동성을 안정적으로 예측하며, 예측의 상하한선 정보를 동시에 제공이 가능하여, 대기오염 대응을 위한 확률론적 의사결정 근거를 제공하여 모니터링을 효과적으로 수행하고 정책 수립에 기여할 수 있을 것으로 사료된다.

4.4. XAI 기반 멀티모달 딥러닝 모델의 예측 결과 분석

본 연구에서 제안한 멀티모달 딥러닝 모델이 단순한 데이터의 통계적 상관관계를 넘어 실제 대기물리학적 메커니즘을 반영하여 학습되었는지 검증하기 위해, XAI 기법을 활용하여 Fig. 4와 같이 높고 넓은 공간적 가시거리를 가진 롯데월드타워를 대상으로 하여 모델의 예측 과정을 분석하였다. Fig. 4(a)는 2025년 1월 17일 경 PM2.5 농도가 40 $\mu g/m^3$ 이상의 나쁨 수준일 때 예측 값에 대한 비전 인코더의 공간적 활성화 맵을 시각화한 결과다. 순전파 과정을 통해 도출된 비전 인코더의 활성화 맵을 분석한 결과, 모델은 이미지 전반의 하늘 영역이나 근거리 지형지물보다는 산 능선과 빌딩들의 스카이라인에 붉은색 히트맵 영역과 같이 가중치를 집중시키는 경향을 보였다. 이러한 모델의 공간적 어텐션 분포는 대기 광학적 메커니즘과 부합한다. PM2.5의 직경은 가시광선 파장대와 유사하여 빛의 투과를 방해하는 미 산란(Mie scattering)을 유발한다. 이와 같은 산란은 가시거리 감소 현상과 픽셀 대비의 소실을 야기하고, 미세먼지 농도와 강한 상관관계를 가지며 이미지의 선명도를 결정짓는 요인이 된다(Ma et al., 2025). 즉, 미세먼지로 인한 대기 혼탁도는 근거리의 건물 형상보다는 원거리의 산 능선과 스카이라인과 같이 하늘과 지표면이 맞닿는 경계선 부분에서 강하게 발생한다고 할 수 있다(Laohakiat et al., 2024). 결론적으로 Fig. 4(a)의 XAI 기반 활성화 맵은 본 연구에서 제시한 EfficientNet 기반 비전 인코더가 픽셀의 단순한 색상 변화를 학습하는 것이 아니라, 인간의 시각 인지 체계와 동일하게 미 산란으로 인한 원거리의 광학적 열화 현상을 PM2.5 농도 산출의 핵심 지표로 추출하고 있음을 나타낸다.

정형 데이터 인코더가 기상 변수들에 부여한 개별 중요도를 XAI의 기여도 분석 기법으로 추출한 히트맵은 Fig. 4(b)에 나타냈다. 히트맵의 세로축은 PM2.5 농도 예측 시점, 가로축은 기상 데이터 변수, 색은 해당 변수의 PM2.5 예측에 대한 기여 정도를 나타낸다. 인코더는 예측 과정에서 주로 풍속과 상대습도 변수에 매우 높은 설명력을 할당하고 있는 것으로 확인되었다. 이는 미세먼지의 이동 및 2차 생성 과정을 결정짓는 핵심 기상 요인들과 정확히 일치한다. 첫째, 대기 역학적 관점에서 풍속은 오염된 공기의 지역적 유입 및 유출과 대기 정체 현상을 유발하는 인자이다. 풍속이 낮을 경우 배출된 오염물질이 소산되지 못하고 국지적으로 축적되어 PM2.5 고농도 현상이 발생 가능하며, 강한 풍속은 오염물질을 외부로 확산시키는 역할을 한다(Chen et al., 2020). 히트맵에서 풍속이 강한 기여도를 나타내는 것은 모델이 대기의 환기 능력을 실시간으로 반영하여 오염물질의 축적 강도를 나타내고 있는 것이다. 둘째, 상대 습도는 미세먼지 입자의 대기 화학적 상호작용을 설명하는 변수이다. PM2.5는 대기 중의 상대습도가 높아질 경우 수분을 흡수하여 입자의 부피와 질량이 급격히 증가하는 흡습 성장 현상을 가진다. 이는 기존 건조 질량 기준의 PM2.5 농도보다 높은 산란 단면적을 가져 가시거리를 실제 이상으로 악화시키게 된다(Won et al., 2021). 비전 모달리티만을 이용한 예측 모델은 상대습도 상승으로 인한 가시거리 악화 현상을 반영하지 못하고 실제 PM2.5 농도보다 더 높게 과대평가할 가능성이 있다. 정형 인코더 기반의 습도 변수 가중치를 활용하여 PM2.5의 흡습 성장 메커니즘을 반영함으로써 단일 모달리티 모델보다 멀티모달 딥러닝에서 PM2.5 예측 향상도를 증가시킬 수 있었다.

결론적으로, XAI 시각화 분석을 통해, 본 연구의 멀티모달 딥러닝 모델이 비전 모달리티로부터는 광학적

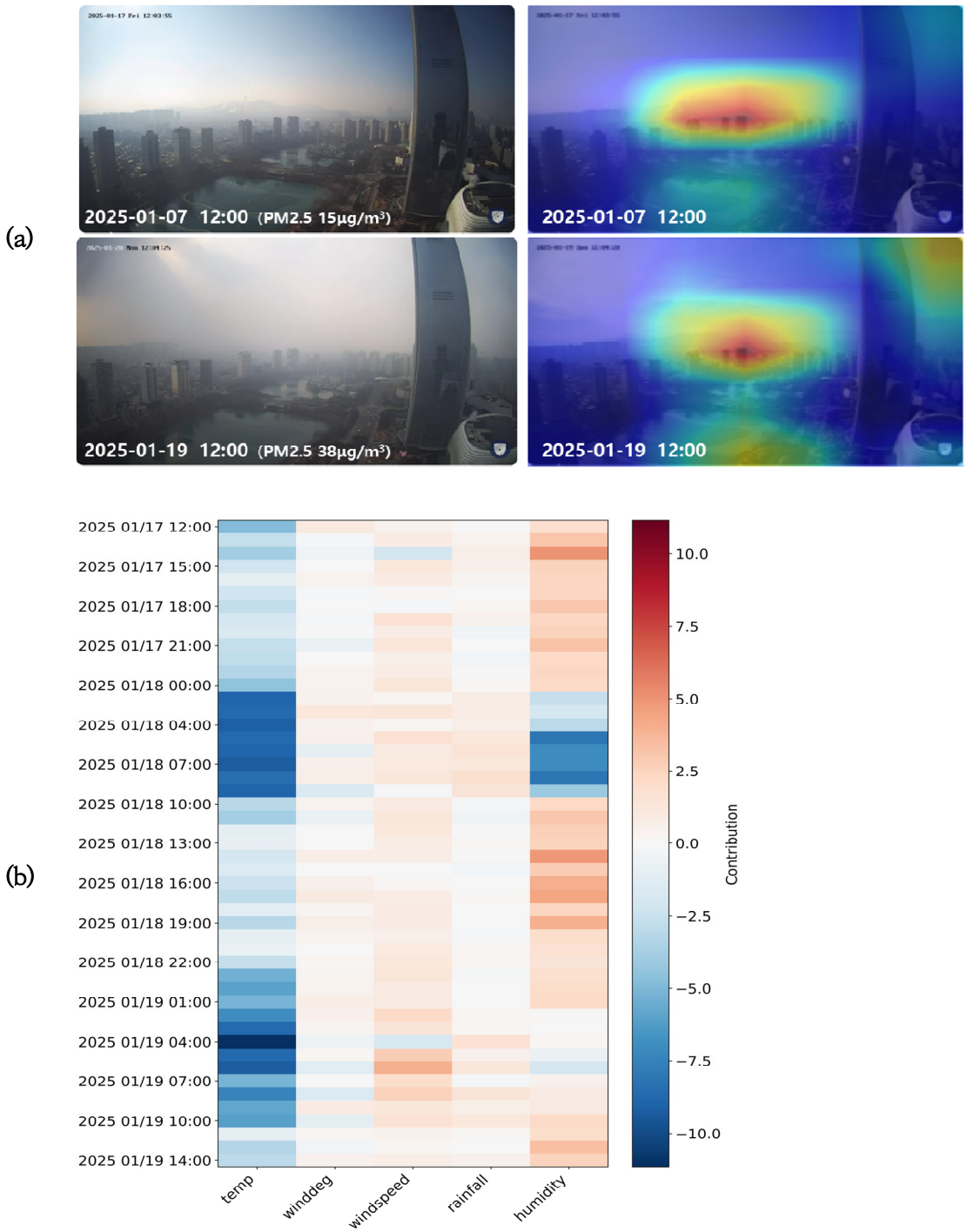

Fig. 4. Explainable AI visualization of the multimodal deep learning model for predicting PM2.5 concentration at the Lotte World Tower site: (a) Spatial activation maps from the vision encoder, and (b) feature contribution heatmap of meteorological variables from the tabular encoder.

열화 패턴을, 기상 정형 모달리티로부터는 대기 수송 및 화학적 메커니즘을 상호 보완적으로 융합하여 PM2.5를 예측함을 알 수 있었다. 이는 제안된 멀티모달 딥러닝반 PM2.5 모델이 단순한 블랙박스 모델이 아니라, 대기 환경 도메인의 물리 화학적 메커니즘을 반영할 수 있는 설명 및 해석 가능성 높은 신뢰성을 갖춘 모델임을 나타낸다.

5. 결 론

본 연구에서는 지상 기상 관측 정형 데이터와 대기 이미지를 융합하여 초미세먼지 (PM2.5) 농도를 정밀하게 예측하고, 그 예측의 불확실성을 정량화하는 멀티모달 딥러닝 모델을 제안하였다. 연구 결과, 대기의 시각적 정보만을 활용한 비전 모델이나 기상 관측값만을 활용한 정형 예측 모델은 예측 성능에 한계를 보였다. 반면, 두 모달리티를 상호 보완적으로 융합한 멀티모달 딥러닝 모델은 지리적·환경적 특성이 다른 세 관측 지점 모두에서 단일 모달리티 모델들보다 우수한 예측 성능으로 R^2 최대 0.73을 달성하며 모델의 범용성과 강건성을 입증하였다. 단일 점 추정의 한계를 극복하기 위해 도입한 분위수 회귀 모델링은 90% 예측 신뢰 구간을 제공함으로써, 급격한 대기질 악화 및 고농도 미세먼지 발생 리스크에 대한 확률론적 해석을 가능하게 하였다. 나아가 설명 가능한 인공지능 분석을 통해 예측의 블랙박스 특성을 극복하였다. 비전 인코더는 원거리 산 능선과 스카이라인에서 발생하는 미 산란 기반의 광학적 열화 현상을 반영하였으며, 정형 인코더는 풍속에 따른 대기 환기 효과 및 상대습도에 따른 에어로졸의 흡습 성장 메커니즘을 반영하고 있음을 확인하였다.

본 연구는 향후 후속 연구를 통해 PM2.5 농도 예측의 신뢰도를 더욱 제고할 수 있을 것이다. 먼저, 비전 인코더가 시각적으로 유사한 안개와 초미세먼지를 식별하는데 발생할 수 있는 한계점은, 정형 기상 데이터와의 교차 검증 및 기상 조건별 가중치 최적화를 통해 효과적으로 보완될 수 있을 것이다. 아울러, 주간의 태양광과 야간 인공광원에 따른 광학적 변동성을 고려하여, 시간적 모달리티를 추가하여 통합한 주야간 범용 모델로의 확장이 가능할 것으로 판단된다.

결론적으로, 본 연구가 제안한 멀티모달 딥러닝 모델은 복잡한 대기 역학 및 광학적 특성을 학습하는 신뢰성 높은 예측 모델이다. 비록 특이 기상 상황 및 광원 차이에 대한 후속 연구가 요구되나, 본 연구는 향후 고가의 대기오염 측정망이 부재한 지역이나 측정망 설치가 까다로운 환경에서도 기존의 CCTV 네트워크 및 모바일 카메라 영상을 활용하여 고해상도의 대기질 모니터링 체계를 구축하는 데 핵심적인 기술로 활용될 수 있을 것으로 기대된다.

감사의 글

이 논문은 국립부경대학교 자율창의학술연구비(2024년)에 의하여 연구되었음.

REFERENCES

Ancona, M., Ceolini, E., Oztireli, C., Gross, M., 2017, Towards better understanding of gradient-based attribution methods for deep neural networks, Published as a conference paper at ICLR 2018, arXiv preprint arXiv:1711.06104v4.

Athira, V., Geetha, P., Vinayakumar, R., Soman, K. P., 2018, Deepairnet: Applying recurrent networks for air quality prediction, Procedia Comput. Sci., 132, 1394-1403.

Chen, C. W., Tseng, Y. S., Mukundan, A., Wang, H. C., 2021, Air pollution: Sensitive detection of PM2.5 and PM10 concentration using hyperspectral imaging, Appl. Sci., 11(10), 4543.

Chen, Q., Chen, W., Pan, G., 2022, An Improved picture-based prediction method of PM2.5 concentration, IET Image Process., 16(11), 2827- 2833.

Chen, Z., Chen, D., Zhao, C., Kwan, M. P., Cai, J., Zhuang, Y., Xu, B., 2020, Influence of meteorological conditions on PM2.5 concentrations across China: A Review of methodology and mechanism, Environ. Int., 139, 105558.

Choi, W. W., Kim, E. J., Lee, S., 2026, Spatiotemporal variation of particulate matter PM2.5 concentration and hierarchical clustering characteristics in Busan metropolitan area, J. Environ. Sci. Int., 35(2), 91-107.

Do, W. G., Kim, D. Y., Song, H. J., Cho, G. J., 2023, A Study on the PM2.5 forecasting method in Busan using deep neural network, J. Environ. Sci. Int., 32(8), 595-611.

Faraji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S., Downey, K., 2022, An Integrated 3D CNN-GRU deep learning method for short-term prediction of PM2.5 concentration in urban environment, Sci. Total

Environ., 834, 155324.
Fujino, R., Miyamoto, Y., 2022, PM2.5 decrease with precipitation as revealed by single-point ground-based observation, Atmos. Sci. Lett., 23(7), e1088.
Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A., 2021, Revisiting deep learning models for tabular data, Adv. Neural Inf. Process. Syst., 34, 18932-18943.
He, H. D., Lu, W. Z., Xue, Y., 2015, Prediction of particulate matters at urban intersection by using multilayer perceptron model based on principal components, Stochastic Environ. Res. Risk Assess., 29(8), 2107-2114.
He, K., Zhang, X., Ren, S., Sun, J., 2016, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 770-778.
Huang, L., Liu, S., Yang, Z., Xing, J., Zhang, J., Bian, J., Liu, T. Y., 2021, Exploring deep learning for air pollutant emission estimation, Geosci. Model Dev., 14(7), 4641-4654.
Kalajdjieski, J., Zdravevski, E., Corizzo, R., Lameski, P., Kalajdziski, S., Pires, I. M., Trajkovik, V., 2020, Air pollution prediction with multi-modal data and deep neural networks, Remote Sens., 12(24), 4142.
Kamble, A., Aramkul, S., Champrasert, P., 2024, A Mobile image-driven PM2.5 estimation framework using deep learning techniques, IEEE Access, 13, 16196-16207.
Kwak, K. H., Kim, J., Choi, M., Jeon, Y., Kim, T., Lee, G., Kang, B. C., Jung, S. A., 2025, Enhancing the simulation performance of PM2.5 compositions in the WRF-CMAQ modeling system using machine learning techniques, J. Korean Soc. Atmos. Environ., 41(3), 430-447.
Laohakiat, S., Klerkkidakan, S., Wiwatwattana, N., 2024, Visually estimating and forecasting PM2.5 levels using hybrid architecture deep neural network, Curr. Appl. Sci. Technol., e0258074.
Lee, K., Lee, S. H., Kim, E. J., 2016, Assessment of global air quality reanalysis and its impact as chemical boundary conditions for a local PM modeling system, J. Environ. Sci. Int., 25(7), 1029-1042.
Liu, Y., Zhang, Y., Yu, P., Ye, T., Zhang, Y., Xu, R., Guo, Y., 2024, Applying traffic camera and deep learning-based image analysis to predict PM2.5 concentrations, Sci. Total Environ., 912, 169233.
Ma, M., Zhao, Z., Ma, Y., Cao, Y., Kang, G., 2025, PM2.5 concentration simulation by hybrid machine learning based on image features, Front. Earth Sci., 13, 1509489.
Nam, K. J., Li, Q., Heo, S. K., Tariq, S., Loy-Benitez, J., Woo, T. Y., Yoo, C. K., 2021, Inter-regional multimedia fate analysis of PAHs and potential risk assessment by integrating deep learning and climate change scenarios, J. Hazard. Mater., 411, 125149.
Park, J., Lee, J. Y., Choi, Y., Babar, Z. B., Seo, S., Ahn, J. Y., Lim, H., 2025, Characterizing temporal and spatial patterns of PM2.5 and PM10 in Baeangnyeong Island and the Seoul metropolitan area, J. Environ. Sci. Int., 34(11), 669-695.
Russell, A. G., Brunekreef, B., 2009, A Focus on particulate matter and health, Atmos. Environ., 43, 4620-4625.
Son, R., Stratoulias, D., Kim, H. C., Yoon, J. H., 2023, Estimation of surface PM2.5 concentrations from atmospheric gas species retrieved from TROPOMI using deep learning: Impacts of fire on air pollution over Thailand, Atmos. Pollut. Res., 14(10), 101875.
Song, S., Lam, J. C., Han, Y., Li, V. O., 2020, ResNet-LSTM for real-time PM2.5 and PM10 estimation using sequential smartphone images, IEEE Access, 8, 220069-220082.
Tan, M., Le, Q., 2019, Efficientnet: Rethinking model scaling for convolutional neural networks, Proceedings of the International Conference on Machine Learning, PMLR, 6105-6114.
Thara, D. K., PremaSudha, B. G., 2019, Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques, Pattern Recognit. Lett., 128, 544-550.
Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Hu, X., 2020, Score-CAM: Score-weighted visual explanations for convolutional neural networks, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, 24-25.
Won, W. S., Oh, R., Lee, W., Ku, S., Su, P. C., Yoon, Y. J., 2021, Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility, Sci. Rep., 11(1), 16401.
Woo, S., Park, J., Lee, J. Y., Kweon, I. S., 2018, CBAM: Convolutional block attention module, Proceedings of the European Conference on Computer Vision (ECCV), Munich, 3-19.
Wu, Y., Wang, X., Wang, M., Liu, X., Zhu, S., 2024, Time-series forecasting of PM2.5 and PM10 concentrations based on the integration of surveillance images, Sensors, 25(1), 95.
Yang, Z., Wang, Y., Xu, X. H., Yang, J., Ou, C. Q., 2022, Quantifying and characterizing the impacts of PM2.5 and humidity on atmospheric visibility in 182

Chinese cities: A Nationwide time-series study, J. Clean. Prod., 368, 133182.
Zagoruyko, S., Komodakis, N., 2016, Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer, Published as a conference paper at ICLR 2017, arXiv preprint arXiv:1612.03928v3.
Zhang, G., Rui, X., Fan, Y., 2018, Critical review of methods to estimate PM2.5 concentrations within specified research region, ISPRS Int. J. Geo-Inf., 7(9), 368.
Zhao, F., Zhang, C., Geng, B., 2024, Deep multimodal data fusion, ACM Comput. Surv., 56(9), 1-36.
Zheng, Y., Yi, X., Li, M., Li, R., Shan, Z., Chang, E., Li, T., 2015, Forecasting fine-grained air quality based on big data, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, 2267-2276.
Zhou, S., Wang, W., Zhu, L., Qiao, Q., Kang, Y., 2024, Deep-learning architecture for PM2.5 concentration prediction: A Review, Environ. Sci. Ecotechnol., 21, 100400.

• Integrated Bachelor's and Master's course. YoungHo Song
Department of Environmental Engineering, Pukyong National University
syh1854@naver.com
• Integrated Bachelor's and Master's course. SuBin Hwang
Department of Environmental Engineering, Pukyong National University
hab6310@naver.com
• Integrated Bachelor's and Master's course. DoGyeong Baek
Department of Environmental Engineering, Pukyong National University
qorehrud1526@naver.com@naver.com
• Integrated Bachelor's and Master's course. SuYoung Song
Department of Environmental Engineering, Pukyong National University
ssyeong63@naver.com
• Professor. KiJeon Nam
Department of Environmental Engineering, Pukyong National University
kjnam@pknu.ac.kr