

논문

Deep Neural Network 모델을 이용한 초미세먼지 취약지역 분석: 서울시 사례연구

PM_{2.5} Vulnerability Areas Analysis Using Deep Neural Network: Application to Seoul City

박문조, 이예경, 김연준, 정동희¹⁾, 안찬중¹⁾, 정형섭*

서울시립대학교 도시과학대학 공간정보공학과,
¹⁾국립환경과학원 기후대기연구부 대기환경연구과

Moonjo Park, Yekyeong Lee, Yeonjoon Kim, Donghee Jung¹⁾,
Changjung An¹⁾, Hyung-Sup Jung*

Department of Geoinformatics, University of Seoul, Seoul, Republic of Korea

¹⁾Department of Air Quality Research, Climate and Air Quality Research Division,
National Institute of Environment Research, Incheon, Republic of Korea

접수일 2025년 5월 5일
수정일 2025년 5월 16일
채택일 2025년 5월 19일

Received 5 May 2025
Revised 16 May 2025
Accepted 19 May 2025

*Corresponding author
Tel: +82-(0)2-6490-2892
E-mail: hsjung@uos.ac.kr

Abstract Fine particulate matter (PM_{2.5}), defined as particles with a diameter of 2.5 micrometers or less, poses significant health risks as it can penetrate alveoli upon inhalation. Research on the vulnerability to PM_{2.5} has predominantly focused on indicators derived from the Intergovernmental Panel on Climate Change (IPCC) vulnerability assessments on health in the context of climate change (2007). This study aims to analyze PM_{2.5} vulnerability areas using a Deep Neural Network (DNN) model. The research area is set in Seoul, South Korea, characterized by high population density and a basin-like topography that inhibits atmospheric dispersion, thus presenting environmental vulnerability to air pollution. In the DNN model, the dependent variable is defined as the PM_{2.5} vulnerability areas, which correspond to currently designated intensive management zones for PM_{2.5}. For the independent variables, data were collected based on the criteria for the designation of these management zones, including the annual average concentration of PM_{2.5}, the number of high PM_{2.5} concentration days, the annual average concentration of PM₁₀, the number of high PM₁₀ concentration days, demographics of vulnerable populations aged 65 and older and under 15, facilities utilized by these vulnerable groups, sources of particulate matter emissions, and pollution sources from industrial activities, totaling ten datasets. Through a data preprocessing process, the study established a spatial database structured in a 100 m × 100 m grid across Seoul, with the independent variables serving as attribute values. The DNN model was quantitatively evaluated using performance metrics, and based on these results, a vulnerability map for PM_{2.5} in Seoul was visualized and analyzed. The findings of this study are anticipated to serve as foundational data for the Ministry of Environment's policy formulation aimed at data and AI-based environmental management.

Key words: PM_{2.5}, Vulnerability, Deep learning, Deep neural network, DNN

1. 서론

초미세먼지(PM_{2.5})란 지름이 2.5 μm 이하로서 지름이 10 μm 이하인 미세먼지(PM₁₀)에 비해 크기가 매우 작아 흡입 시 기도에서 걸리지 못하고 대부분 폐포까지 침투하여 심장질환과 호흡기 질환을 유발하는

대기오염물질이다.

World Health Organization (WHO)는 전문가 검토를 기반으로 초미세먼지 노출과 건강 영향 간의 인과관계(Causal relationships)를 분석하여, 초미세먼지 단기·장기 노출과 건강 영향을 표 1에 요약 제시하고, 1급 발암물질로 분류하였다(WHO, 2021).

그리고 WHO는 표 1의 연구 결과를 바탕으로 표 2의 2021년 새로운 대기질 기준(Air quality goal, AQG)과 잠정기준(Interim target)을 제안하였다(WHO, 2021). 잠정기준은 WHO 권고 대기질 기준 달성이 어려운 국가나 지역에서 단계적으로 목표를 설정하고 개선 노력을 추진할 수 있도록 설정된 중간 목표 수준을 의미한다.

이러한 WHO가 제안한 새로운 대기질 기준 및 잠정기준을 참고하여, 국내 초미세먼지 관리기준은 국내외 대기오염물질 노출로 인한 건강 영향 자료를 수집하고 허용 위해 수준을 결정한 후, 국내 규제 준수를 위해 필요한 공학적 기술 수준, 경제적 효과, 초미세먼지 농도에 영향을 주는 요인 등을 고려하여 국내 실정에 가장 적합한 기준을 설정했다(표 3).

그림 1을 살펴보면, 국내 초미세먼지 농도는 측정을 시작한 2015년부터 2023년까지 $26 \mu\text{g}/\text{m}^3$ 에서 $18 \mu\text{g}/\text{m}^3$ 로 약 30% 감소 수준을 보였다. 그러나 여전히 국내 $\text{PM}_{2.5}$ 연평균 농도 기준($15 \mu\text{g}/\text{m}^3$)을 충족하지 못하고 있다.

환경부는 초미세먼지 대응 정책에 있어 전국 도심 지역 중점 관리, 인체 위해성 최우선 고려, 통합적이고 과학적인 관리 등 새로운 패러다임의 전환을 제시하며 빅데이터 분석에 기반한 정책을 강조하고 있다.

그동안 초미세먼지 관련 연구는 초미세먼지 배출원 및 특성(Park *et al.*, 2023a), 초미세먼지 집중배출지역(Kim *et al.*, 2019), 머신러닝을 이용한 초미세먼지 농도 예측 모델 개발(Park *et al.*, 2023b; Ryu *et al.*, 2022; Gil and Lee, 2021; Cho *et al.*, 2019) 등 초미세먼지 현상

Table 1. WHO $\text{PM}_{2.5}$ causal relationships setting.

	Exposure term	Health outcomes	Causality determination
$\text{PM}_{2.5}$	Long-term	<ul style="list-style-type: none"> All-cause mortality Cardiovascular mortality Respiratory mortality Lung cancer mortality 	<ul style="list-style-type: none"> Causal for all-cause mortality Causal for cardiovascular mortality Causal for respiratory mortality Likely causal for lung cancer mortality
	Short-term	<ul style="list-style-type: none"> All-cause mortality Cardiovascular mortality Respiratory mortality 	<ul style="list-style-type: none"> Causal for all-cause mortality Causal for cardiovascular mortality Causal for respiratory mortality Causal for cardiovascular effects Causal for respiratory effects

Table 2. WHO recommended AQG level and interim targets for $\text{PM}_{2.5}$.

	Exposure term	Interim target ($\mu\text{g}/\text{m}^3$)				AQG ($\mu\text{g}/\text{m}^3$)
		1	2	3	4	
$\text{PM}_{2.5}$	24 hour	75	50	37.5	25	15
	1 year	35	25	15	10	5

Table 3. South Korea and international $\text{PM}_{2.5}$ management standards.

	Exposure term	South Korea	United States	Japan	Canada	Australia	Hong Kong	China	United Kingdom	European Union	WHO
$\text{PM}_{2.5}$ ($\mu\text{g}/\text{m}^3$)	24 hour	35	35	35	27	25	75	75	-	-	15
	1 year	15	15	15	8.8	8	35	35	25	25	5

자체의 이해와 예측이 주요 연구대상이 되어 왔다. 또한 미세먼지 취약성과 취약지역에 관련된 연구는 Intergovernmental Panel on Climate Change (IPCC) 에서 제시한 기후변화 건강 취약성 평가모형 (IPCC, 2007)을 준용하여 진행되었다 (Mun and Jung, 2024; Jung and Jun, 2020). 그러나 이러한 기존 연구들은 환경부의 초미세먼지 대응 정책 방향에서 강조하는 인체 위해성 최우선 고려, 데이터에 기반한 통합적이고 과학적인 관리와 같은 측면을 충분히 다루지 못하였다.

본 연구 목적은 대표적인 딥러닝 모델 중 하나인 심층 신경망 (Deep Neural Network, DNN)을 활용하여, 빅데이터 기반 초미세먼지 대응 정책 지원을 위한 초미세먼지 취약지역을 탐색하고 분석하는 것이다. 결과변수인 초미세먼지 취약지역을 정의한 후, 초미세먼지 취약지역 설명변수와 결과변수를 공간 데이터베이스로 구축하고 DNN 모델에 적용하였다. DNN 모델은 정량적 성능평가를 통해 검증되었으며, 이 결과를 활용하여 초미세먼지 취약지역 지도로 시각화하여 분석하였다.

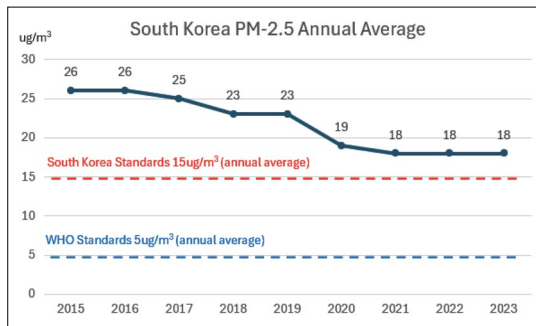


Fig. 1. PM_{2.5} Annual average concentration of South Korea.

2. 연구 범위 및 연구 자료

2.1 초미세먼지 취약지역

본 연구에서는 초미세먼지 취약지역 분석을 위해 DNN 모델을 적용하고자 한다. 이 과정에서 결과변수로 필수적으로 활용될 초미세먼지 취약지역 정의가 중요하다.

본 연구에서 초미세먼지 취약지역은 현재 운영 중인 미세먼지 집중관리지역으로 정의되며, 이를 통째로 구체적이고 실질적인 분석이 가능해진다. 미세먼지 집중관리지역은 환경부의 「미세먼지 집중관리지역 지정 운영 가이드 라인」에 따라 지정된 지역으로, 해당 가이드 라인에 명시된 지정요건을 바탕으로 선정된다. 표 4는 미세먼지 집중관리지역의 지정요건을 요약한 것이다.

2.2 연구지역

본 연구의 대상 지역은 서울특별시로 설정하였다 (그림 2). 서울시는 한반도의 중심부에 위치하며, 북한산, 도봉산, 관악산 등 높은 산으로 둘러싸인 분지 형태이다. 이러한 지형적 특성으로 인해 대기 확산이 제한되어, 서울시는 초미세먼지가 쉽게 축적될 수 있는 환경적 취약성을 가지고 있다.

서울시는 인구밀도가 높은 대도시로, 산업 및 교통 활동이 집중되어 있어 다양한 초미세먼지 배출원이 존재한다. 특히 수도권 급격한 도시화 및 중국의 공업화로 인해 발생하는 대기오염물질의 영향을 크게 받고 있으며, 이는 서울시의 대기질 악화에 기여하고 있다.

Table 4. The criteria for Particle Matter (PM) management zone.

Criteria	
PM _{2.5} (μg/m ³) or PM ₁₀ (μg/m ³)	<ul style="list-style-type: none"> • PM_{2.5}: (for 3 years) Exceeding Annual average concentration 15 μg/m³ • PM₁₀: (for 3 years) Exceeding Annual average concentration 50 μg/m³
Facilities density	A minimum of 10 facilities catering to PM _{2.5} vulnerable groups (daycare centers, kindergartens, elementary schools, senior welfare facilities, postpartum care centers, and hospitals)
Area	Distance within 200 meters of density of vulnerable groups facilities, with a maximum distance of 1 kilometer (In cases where the distance extends to 1 kilometer, a management plan for source of PM must be included)

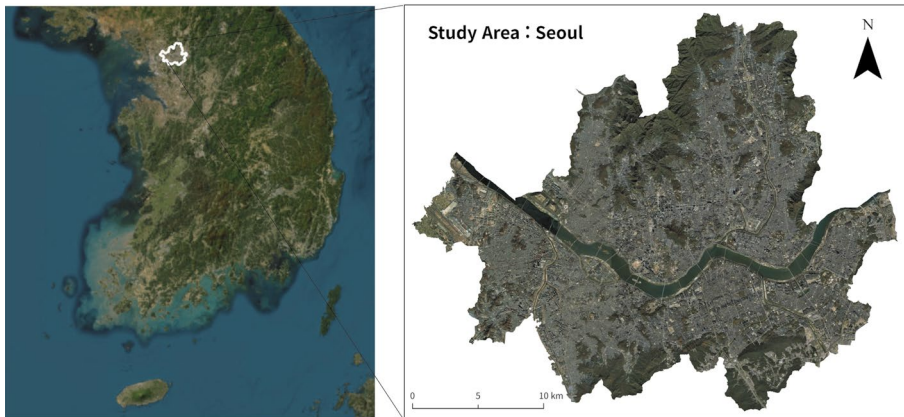
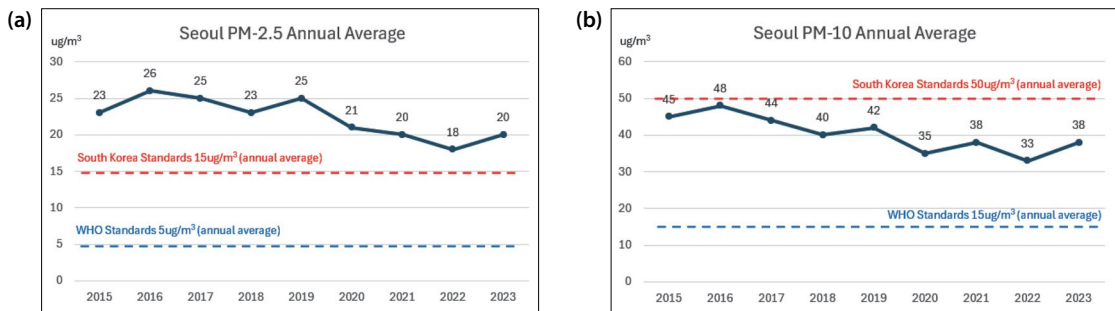


Fig. 2. Study area: Seoul.

Fig. 3. (a) PM_{2.5} annual average concentration and (b) PM₁₀ annual average concentration in Seoul.

이와 같은 대기오염 문제를 해결하기 위해 서울시는 여러 정책을 시행해 오고 있다. 2013년 10월, 서울시는 전국 최초로 초미세먼지 경보제를 도입하였으며, 초미세먼지에 취약한 계층인 15세 미만 유소년과 65세 이상 고령 인구의 건강 보호를 위한 정책을 강화하고 있다. 그러나 서울시의 초미세먼지 농도는 여전히 국내 초미세먼지 연평균 농도 기준인 15 $\mu\text{g}/\text{m}^3$ 를 충족하지 못하고 있어, 지속적인 정책 개선과 연구가 필요하다(그림 3).

2.3 연구 자료

초미세먼지 취약지역 분석을 위해 표 4의 지정 요건에 기반하여, 표 5와 같은 연구 자료가 수집되었다. 본 연구에 사용된 연구 자료들은 설명변수와 결과변수로

구분된다.

설명변수는 PM_{2.5} 연평균 농도, PM_{2.5} 고농도 발생일 수, PM₁₀ 연평균 농도, PM₁₀ 고농도 발생일 수, 취약계층인 15세 미만 및 65세 이상 인구수, 취약계층 이용시설, 비산먼지 발생 사업장, 대기오염 배출 사업장, 도로오염원 PM 배출량 등 총 15개의 자료 항목이 10개의 설명변수로 활용되었다. 결과변수는 현재 운영 중인 서울시 미세먼지 집중관리구역을 초미세먼지 취약지역으로 정의하고 사용하였다.

먼저 PM_{2.5} 연평균농도, PM_{2.5} 고농도 발생일 수, PM₁₀ 연평균농도, PM₁₀ 고농도 발생일 수는 서울시 경계 지역의 가장자리 효과를 완화하기 위해, 서울 반경 30 km 이내에 위치한 총 84개의 도시대기측정망 및 도로변대기측정망의 관측 자료를 활용하였다. 표 4

Table 5. Variables of PM_{2.5} vulnerability area in this study.

	Factors	Data type	Resolution	Source	
Explanatory variables	PM _{2.5} annual average concentration	Point	Daily	www.airkorea.or.kr	
	High PM _{2.5} frequencies	Point	Yearly	www.airkorea.or.kr	
	PM ₁₀ annual average concentration	Point	Daily	www.airkorea.or.kr	
	High PM ₁₀ frequencies	Point	Yearly	www.airkorea.or.kr	
	Facilities of PM _{2.5} vulnerable groups	Daycare centers	Point	-	data.seoul.go.kr
		Kindergartens	Point	-	data.seoul.go.kr
		Elementary schools	Point	-	data.seoul.go.kr
		Senior welfare facilities	Point	-	data.seoul.go.kr
		Postpartum care centers	Point	-	data.seoul.go.kr
		Hospitals	Point	-	data.seoul.go.kr
	Population aged under 15	Grid	-	sgis.kostat.go.kr	
	Population aged 65 and older	Grid	-	sgis.kostat.go.kr	
	Sources of fugitive dust	Point	-	cis.seoul.go.kr	
	Sources of air pollutant emissions	Point	-	cleanair.seoul.go.kr	
	Sources of PM by road mobile	Polygon	Weekly	www.bigdata-transportation.kr	
Outcome variable	PM _{2.5} management zone	Grid	-	www.seoul.go.kr	

의 미세먼지 집중관리 지정요건을 준용하여, 수집된 자료 중 2021년부터 2023년까지 3년간의 측정값을 기반으로 각 지점별 3년 평균 농도 및 고농도 발생일 수를 산출하여 분석에 적용하였다(그림 4(a)).

취약계층 이용시설 자료는 서울 열린데이터 광장 포털에서 제공하는 서울시 어린이집, 유치원, 초등학교, 산후조리원, 노인복지시설, 요양병원 등 6개 유형 시설의 위치 정보를 담고 있는 포인트 자료를 수집하여, 단일한 포인트 자료 형태로 통합되었다(그림 4(b)).

취약계층인 15세 미만 및 65세 이상 인구 자료는 통계지리정보서비스(Statistical Geographic Information Service, SGIS)에서 제공하는 격자경계 100 m 단위 서울시 유소년 및 고령 인구 자료를 사용하였다.

비산먼지 발생 사업장 자료는 서울시 건설 알림이 포털에서 제공하는 비산먼지 발생 사업 정보 포인트

자료를 수집하여 구축하였다(그림 4(c)).

대기오염 배출 사업장 자료는 서울시 대기환경정보 포털에서 대기오염물질 배출 사업장 현황 자료를 활용하여 포인트 자료로 생성하였다(그림 4(d)).

도로오염원 PM 배출량 자료는 국가교통 데이터 오픈마켓 포털에서 제공하는 대기오염물질 배출량 행정 구역 읍면동 단위 자료 중 2021년부터 2023년까지의 월평균 자료를 활용하여 3년 연평균 자료로 산출하였다(그림 4(e)).

3. 연구 방법

그림 5는 본 연구에서 수행된 전체 과정을 나타낸 흐름도이다. 연구 자료 수집 및 전처리 후, 데이터셋

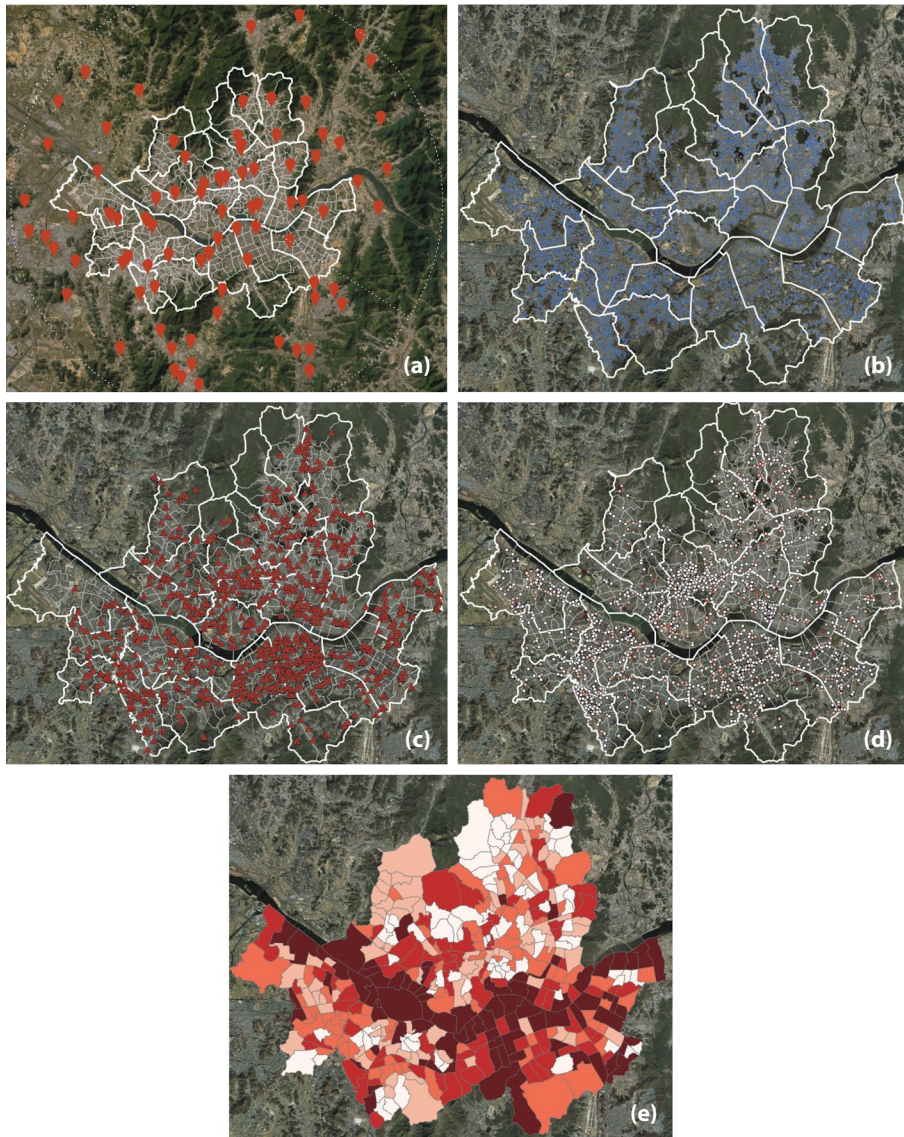


Fig. 4. Data of explanatory variables in this study: (a) 84 air quality measurement sites, (b) Facilities of $PM_{2.5}$ vulnerable groups, (c) Facilities of fugitive dust, (d) Facilities of air pollutant emissions, (e) PM by road mobile by Dong.

는 DNN 모델의 학습 데이터세트와 성능 및 검증평가를 위한 테스트 데이터세트로 나누었다. 학습을 마친 DNN 모델은 정량적 성능 지표를 통해 검증하고, 설명변수의 중요도 평가를 진행하였다. 최종적으로 서울시 초미세먼지 취약지역 매핑을 수행하였다.

3.1 연구 자료 전처리

그림 6은 효과적인 DNN 모델의 학습 및 예측을 위해, 서울시를 $100\text{ m} \times 100\text{ m}$ 단위 Grid로 격자화하여, 설명변수 및 결과변수를 속성값으로 가지는 서울시 공간데이터베이스를 구축하였다. 격자화된 총 61,650 개 Grid 중 산과 강에 해당하는 Grid를 제외한 41,557

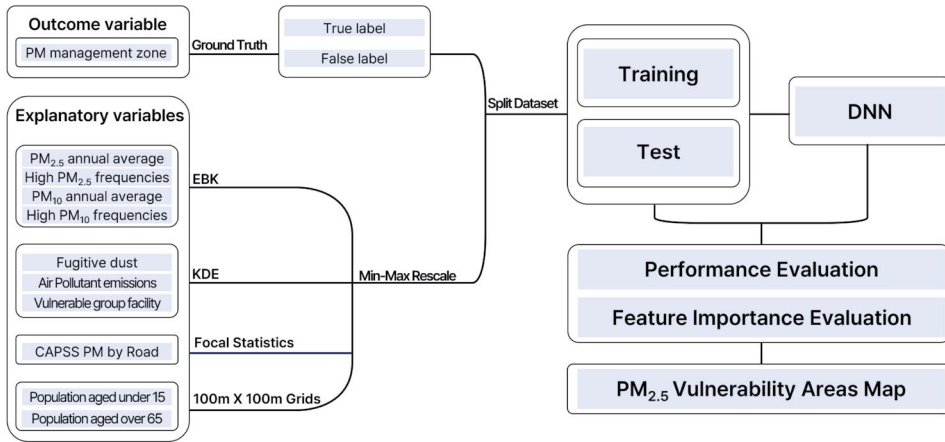


Fig. 5. This flow chart summarizes the study process from data collection and preprocessing, through DNN model training and evaluation, to PM_{2.5} vulnerability mapping in Seoul.

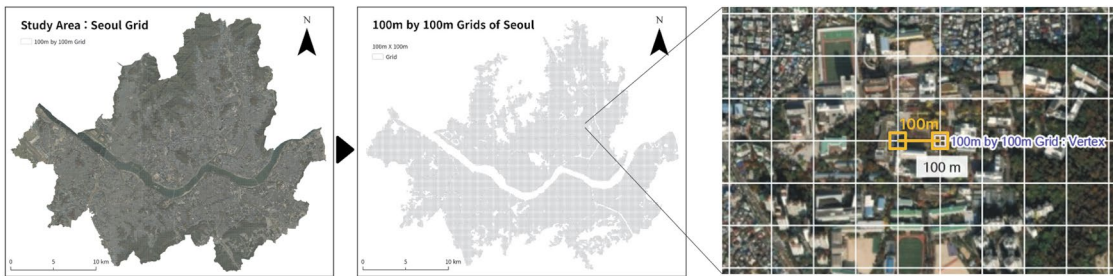


Fig. 6. Seoul grids (100 m by 100 m): 41,557 grids were used in this study.

개의 Grid가 본 연구에 활용되었다.

Empirical Bayesian Kriging (EBK) 보간은 공간적 데이터의 보간 및 예측을 위한 통계적 방법으로, 지리적 데이터에서 공간적 상관관계를 모델링하여 알려지지 않은 위치의 값을 추정하는 기법이다. EBK에 대한 전반적인 내용은 Gribov and Krivoruchko (2020)에 상세히 기술되어 있다. 본 연구에서는 PM_{2.5} 연평균농도, PM_{2.5} 고농도 발생일 수, PM₁₀ 연평균농도, PM₁₀ 고농도 발생일 수 데이터는 EBK를 수행하여 서울시 Grid에 속성값으로 할당하였다(그림 7(a, b, c, and d)).

Kernel Density Estimation (KDE)은 데이터의 확률 밀도 함수 추정하는 비모수적 기법으로, 포인트 데이터 주변에 커널 함수를 배치하여 전체적인 데이터 분포를 추정한다. KDE(식 1)과 같다.

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

K는 가우시안 커널 함수, h는 대역폭(Smoothing factor), n은 데이터 포인트수, x_i는 개별 데이터 포인트이다. 포인트 데이터인 취약계층 이용시설, 비산먼지 발생 사업장, 대기오염 배출 사업장 데이터는 KDE를 통해 서울시 Grid 속성값으로 할당하였다(그림 7(e, f, and h)).

Focal statistics는 각 Grid의 인접 영역 내의 Grids에 대한 통계를 계산하여 해당 Grid의 값으로 할당하는 공간분석 기법이다. 도로오염원이 발생시키는 PM 배출량 데이터는 한국교통연구원에서 제공하는 읍면동 단위의 Clean Air Policy Support System (CAPSS) 도

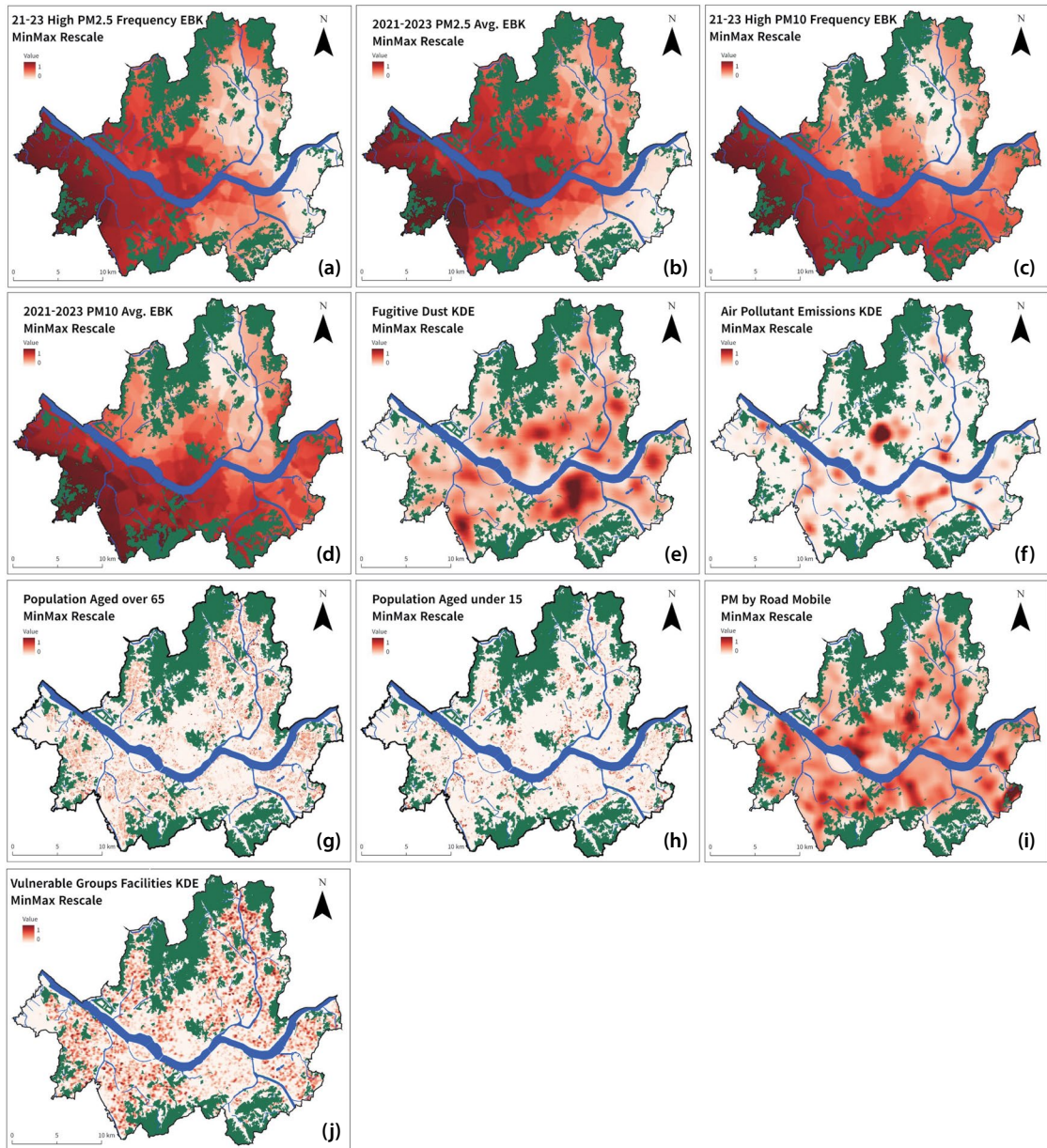


Fig. 7. Variables after Min-Max Re-Scaling used in this study: (a) High $PM_{2.5}$ frequencies EBK, (b) $PM_{2.5}$ annual average EBK, (c) High PM_{10} frequencies EBK, (d) PM_{10} annual average EBK, (e) Fugitive dust KDE, (f) Air pollutant emissions KDE, (g) PM by road mobile, (h) Facilities of $PM_{2.5}$ vulnerable Groups KDE, (i) Population aged under 15, and (j) Population aged under 65 and older. Empirical Bayesian Kriging (EBK).

로오염원 배출량 데이터를 읍면동 면적으로 나눈 후, Grid 면적을 곱하여 각 Grid가 가진 도로오염원 배출량을 계산한다. 각 Grid는 다시 반경 500 m 범위 내의

인접 Grid 평균값으로 도로오염원 배출량을 산출하여 사용한다(그림 7(g)).

DNN 모델의 학습 효율성을 높이고, 각 특성의 중

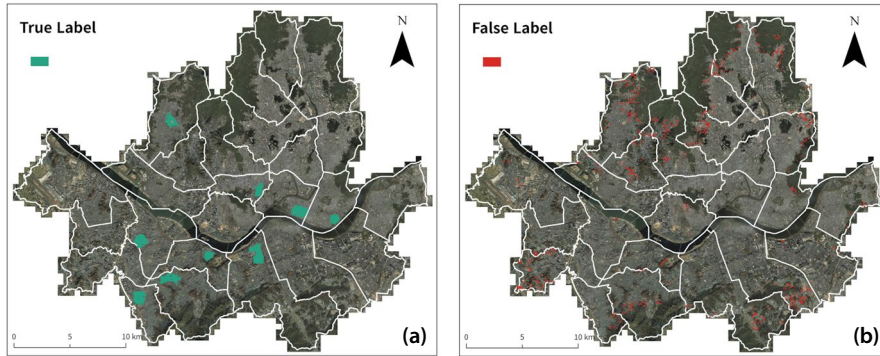


Fig. 8. Ground truth of this study consists of (a) True label and (b) False label.

요성을 동등하게 반영하기 위해서는 데이터의 값의 범위와 단위를 일치시켜야 한다. 데이터의 값을 0과 1 사이값으로 리스케일하는 Min-Max 정규화 기법을 적용하였다(식 2).

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Ground Truth는 DNN 모델이 초미세먼지 취약지역과 취약지역이 아닌 지역을 구별하는 데 중요한 역할을 하며, 신뢰성 있는 Ground Truth 제작은 성능에 직접적인 영향을 미친다(Gong *et al.*, 2022). DNN 모델 학습의 결과변수로 사용되는 Ground Truth를 생성하였다(그림 8).

True label은 초미세먼지 취약지역이다(그림 8(a)). 이는 현재 서울시에서 지정하여 운영 중인 9개의 미세먼지 집중관리구역이다. False label은 초미세먼지 취약지역이 아닌 지역이며, 빈도비(Frequency ratio) 분석 방법을 활용하여 제작되었다(그림 8(b)).

빈도비 분석 방법에 따라 설명변수에 대해 구간을 정하고, 각 구간의 Grid 수를 전체 Grid 수로 나누는 비율을 구한다. 이때 각 구간별 초미세먼지 취약지역 Grid 수의 비율을 구하여 두 비율을 나누어 설명변수의 구간에 대한 빈도비를 계산한다. 이렇게 계산된 빈도비 값이 낮은 구간을 초미세먼지 취약지역이 아닌 지역으로 가정하여 False label을 생성하였다. 이 False label은 True label과 함께 Ground truth로 구성되었다.

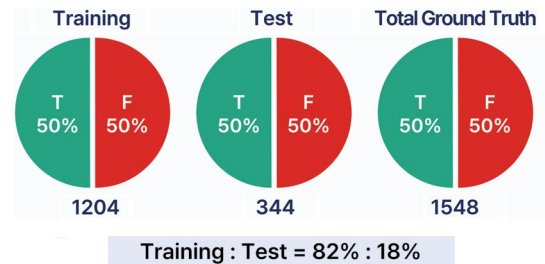


Fig. 9. Dataset is split into training (82%) and test (18%) datasets for the DNN model.

취약지역과 취약지역이 아닌 지역은 모두 동일하게 9개의 클러스터로 그룹화하였다. True label 9개 클러스터의 Grid 수의 합은 766개로, False label의 9개 클러스터 Grid 수의 합도 동일하게 766개로 총 1532개로 설정하였다. Ground Truth는 통계적으로는 유사하지만, 서로 독립적인 학습 데이터세트(82%)와 테스트 데이터세트(18%)로 나누었고, 학습 데이터세트는 DNN 모델 학습에, 테스트 데이터세트는 DNN 모델의 검증과 성능평가에 사용되었다(그림 9).

3.2 Deep Neural Network 모델

심층 신경망은 입력층과 출력층 사이에 다수의 은닉층을 포함하는 인공신경망의 발전된 유형이다. 이 모델은 인간의 신경 시스템 구조 및 작동 원리에서 착안된 머신러닝 기법이다(Du *et al.*, 2018). 인공신경망 모델의 훈련에 핵심적인 역할을 하는 역전파 알고리

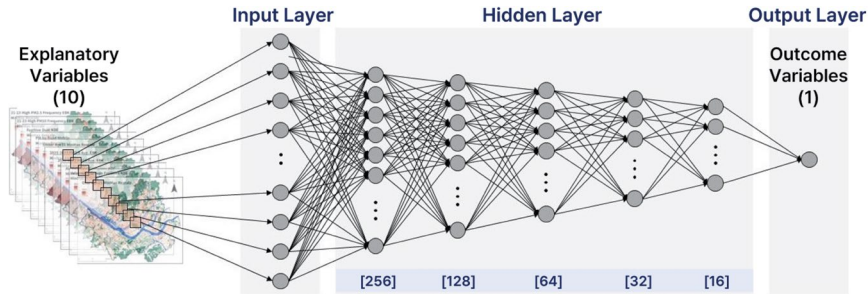


Fig. 10. DNN model architecture proposed in this study. The DNN model, a 6-layer network with 5 hidden layers, classifies $PM_{2.5}$ vulnerability and non-vulnerability areas based on 10 explanatory variables.

좁은 주어진 입력-출력 데이터 쌍을 기반으로 작동한다. 이 알고리즘은 모델의 예측 결과와 실제 관측값 사이의 오차를 최소화하는 방향으로 신경망 내 연결 가중치를 반복적으로 조정함으로써 모델을 최적화한다 (Liu *et al.*, 2017).

DNN 모델의 최적 성능을 확보하기 위해 모델의 주요 설정 요소들에 대한 다양한 조합으로 반복적인 학습 및 검증을 수행하였다. 이러한 과정을 통해 테스트 데이터세트에서 가장 우수한 성능을 보인 모델 설정을 최종적으로 결정하였다. 최적화된 모델 설정은 다음과 같다. 모델은 총 6개 층으로 구성된다. 5개로 구성된 은닉층을 통해, 입력층의 10개의 설명변수의 특징을 학습하여, 이를 바탕으로 출력층에 초미세먼지 취약지역과 취약지역이 아닌 지역을 분류한다. 각 은닉층의 노드 수는 256, 128, 64, 32, 16으로 설정되었다. 과적합을 방지하기 위해 각 은닉층에는 L2 정규화(Weight decay)와 Dropout 기법을 적용하였다. 활성화 함수로는 ReLU (Rectified Linear Unit)를 사용하고, 최적화 알고리즘으로 Adam (Adaptive moment estimation) 옵티마이저를 사용하였다. 배치크기는 64, 학습률은 0.001, 에포크 수는 20으로 설정하였다(그림 10).

3.3 DNN 모델 검증 및 성능평가

서울시 초미세먼지 취약지역 분류 모델의 검증 및 성능평가를 위해, 본 연구에서는 Confusion Matrix를 기반으로 정확도(Accuracy), 정밀도(Precision), 재현

Table 6. Confusion matrix for DNN model validation and performance evaluation.

Actual \ Predicted	Positive	Negative
	Positive	TP (True positive)
Negative	FP (False positive)	TN (True negative)

율(Recall), F1-score을 측정하고, AUC와 같은 다양한 성능 지표를 사용하였다(표 6).

정확도(Accuracy)는 전체 Grid에서 모델이 정확하게 분류한 Grid의 비율을 나타내는 지표이다(식 3). 0과 1 사이의 값을 가지며, 1에 가까울수록 모델의 성능이 우수함을 의미한다. 하지만 클래스 간의 불균형이 심할 경우 모델의 성능을 과대평가할 위험이 있다. 따라서 정확도는 다른 성능 평가지표들과 함께 고려되어야 하며, 올바른 결과 해석을 위한 주의가 필요하다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

정밀도(Precision)는 모델이 True로 예측한 Grid 중에서 실제로 True인 Grid의 비율을 의미한다(식 4). 이는 모델이 얼마나 정확하게 양성 Grid를 예측하는지를 나타낸다. 0과 1 사이의 값을 가지며, 1에 가까울수록 모델의 성능이 우수함을 의미한다.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

재현도(Recall)는 실제 True가 Grid 중 모델이 True로 올바르게 예측한 Grid 비율을 의미한다(식 5). 모델

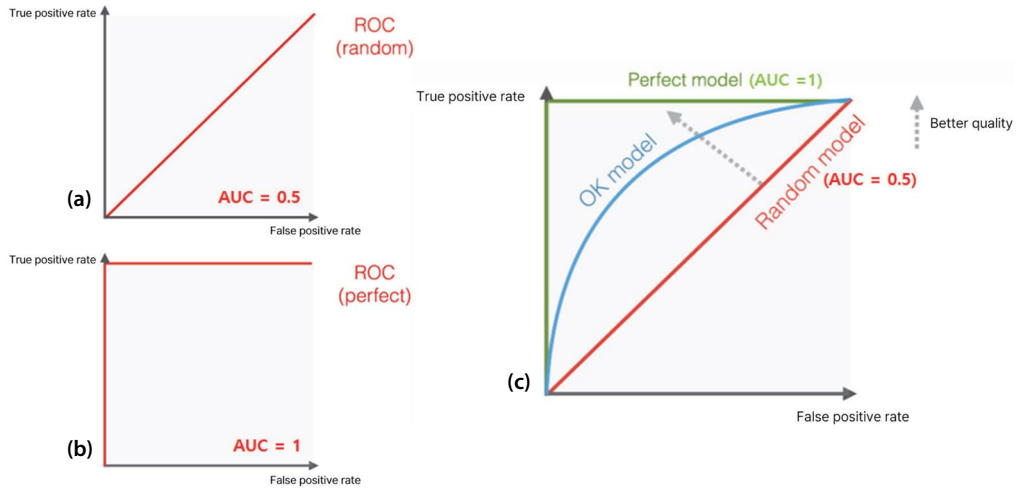


Fig. 11. ROC curve for classifier model evaluation: (a) AUC=0.5 (random), (b) AUC=1 (perfect), (c) ROC curves and AUC are used to evaluate classifier performance.

이 양성 Grid를 얼마나 잘 포착하는지를 측정한다. 0과 1 사이의 값을 가지며, 1에 가까울수록 모델의 성능이 우수함을 의미한다.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

F1-Score는 정밀도와 재현도의 조화 평균으로, 두 지표를 동시에 고려하여 모델의 성능을 종합적으로 평가하기에 적합한 지표이다(식 6). 0과 1 사이의 값을 가지고, 1에 가까울수록 모델의 성능이 우수함을 의미한다.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

AUC (Area Under the Curve)는 ROC (Receiver Operating Characteristic) 곡선 아래의 면적을 나타내며, DNN 모델의 이진 분류 성능을 평가하는 지표로 사용된다. AUC 값이 클수록 모델이 양성(positive)과 음성(negative)을 더 잘 구분한다는 의미를 가지며, 1에 가까울수록 모델의 성능이 우수함을 의미한다(그림 11).

3. 4 DNN 모델 설명변수 중요도 평가

SHAP (Shapley Additive exPlanations)은 모델의 특정 예측값에 대한 개별 설명변수의 기여도를 SHAP

value로 계산한다. 이 SHAP value는 해당 설명변수가 예측값을 변화시키는 정도를 나타낸다(Ko *et al.*, 2025).

서울시 초미세먼지 취약지역 분류 예측에 대한 설명변수 중요도 평가를 위해 본 연구에서는 Mean absolute SHAP value를 사용하였다. Mean absolute SHAP value는 개별 SHAP value의 절댓값을 평균하여 산출된 값으로서, 해당 설명변수가 예측에 미치는 평균적인 영향력의 크기를 의미하며, 설명변수의 전반적인 중요도를 객관적으로 평가하는 핵심 지표로 사용된다(Wang *et al.*, 2022).

4. 연구 결과

4. 1 DNN 모델의 검증 및 성능평가

DNN 모델 검증 및 성능을 평가하기 위해 테스트 데이터세트를 활용하였다. 본 연구의 DNN 모델의 주요 성능 지표인 정확도, 정밀도, 재현율, F1-score를 요약하고 있다(표 7). 이러한 지표들을 통해 모델이 초미세먼지 취약지역 분류에 있어서 얼마나 정확하게 Grid 단위로 예측했는지를 평가할 수 있다.

DNN 모델의 정확도는 0.9667로, 이는 테스트된

Table 7. Confusion matrix-based performance indicators of the trained DNN model.

	True	False	Total
Accuracy	-	-	0.9667
Precision	1	0.9375	0.9688
Recall	0.9333	1	0.9667
F1-score	0.9655	0.9677	0.9832

Grid 중 약 96.67%가 올바르게 분류되었음을 의미한다. 높은 정확도 수치는 학습된 모델이 학습 데이터의 패턴을 성공적으로 일반화하고, 테스트 데이터세트에 대해서도 높은 수준의 성능을 발휘했음을 나타낸다.

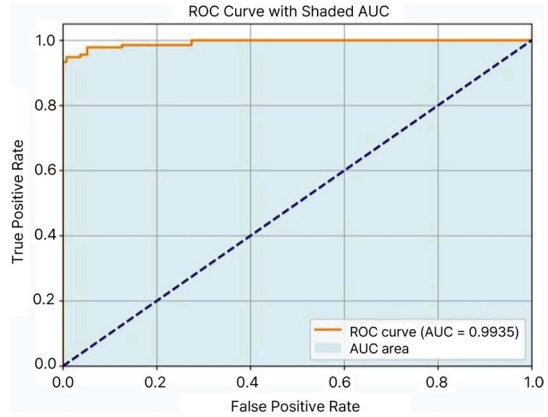
양성클래스(True)의 정밀도는 1로, 이는 모델이 초미세먼지 취약지역으로 분류된 모든 지역이 실제로 취약지역이라는 것을 의미한다. 즉, 잘못된 긍정 예측(False Positive)이 전혀 없음을 보여준다. 반면 음성클래스(False)의 정밀도는 0.9375로, 모델이 음성으로 예측한 지역 중 약 6.25%는 실제로는 취약지역일 가능성이 있음을 나타낸다. 이러한 오탐 가능성에도 불구하고, 전반적인 정밀도가 높은 수준을 유지하고 있다.

양성클래스의 재현율은 0.9333으로 실제 초미세먼지 취약지역 중 93.33%를 모델이 올바르게 탐지했음을 의미한다. 이는 본 연구에서 중요한 지표로 작용하며, 환경 연구에서는 오탐보다 누락을 최소화하는 것이 더 중요하기 때문에, 높은 재현율은 특히 중요한 역할을 한다. 음성클래스의 재현율은 1이므로, 이는 실제 취약하지 않은 지역을 정확하게 분류했음을 의미한다. 즉, 취약하지 않은 지역에서 잘못된 분류가 발생하지 않았음을 의미한다.

F1-score는 양성클래스의 경우 0.9655, 음성클래스의 경우 0.9677로 측정되었다. 높은 F1-score는 모델이 초미세먼지 취약지역 탐지뿐만 아니라 취약지역이 아닌 지역 분류에서도 높은 신뢰성을 보이고, 매우 높은 성능을 발휘하고 있음을 나타낸다.

다음으로, 테스트 데이터세트를 기반으로 DNN 모델의 성능을 시각적으로 평가하기 위해 ROC 곡선을 도출하였다.

그림 12는 학습된 DNN 모델의 ROC 곡선을 나타

**Fig. 12.** ROC curve and AUC (= 0.9935) of trained DNN classifier evaluation.

내며, 이를 통해 모델의 분류 성능을 확인할 수 있다. ROC 곡선 아래의 면적(AUC)은 0.9935로, 모델이 양성클래스와 음성클래스를 구분하는 데 있어 높은 성능을 나타냄을 의미한다.

본 연구에서 학습된 DNN 모델은 테스트 데이터세트에서 높은 정확도, 정밀도, 재현율 및 F1-score를 보였으며, AUC 값이 0.9935로 모델의 성능이 매우 우수함을 확인할 수 있다. 이러한 성능은 DNN 모델이 초미세먼지 취약지역과 취약지역이 아닌 지역을 구별하는 데 있어 매우 높은 분류 성능을 발휘하고 있음을 나타낸다.

4.2 DNN 모델 설명변수 중요도 평가

DNN 모델의 결과를 해석하고 각 설명변수의 상대적 기여도를 파악하기 위해 Mean absolute SHAP value를 산출하였다. 그림 13은 산출된 Mean absolute SHAP value 그래프이다. Mean absolute SHAP value는 해당 설명변수가 모델의 초미세먼지 취약지역 분류에 미치는 평균적인 영향력의 크기를 의미한다. 이를 통해 모델의 초미세먼지 취약지역을 판단하는 주요 요인을 파악하고 각 변수의 중요도를 평가하였다.

비산먼지 발생 사업장 변수가 Mean absolute SHAP value가 가장 높아 초미세먼지 취약지역 예측에 가장 큰 영향력을 미치는 주요 변수로 확인되었다. 이는

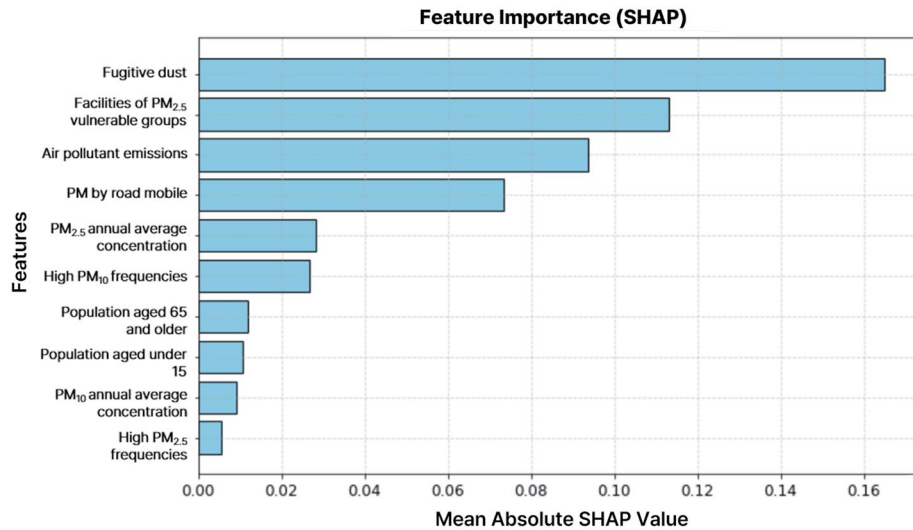


Fig. 13. Mean absolute SHAP values of 10 explanatory variables for the trained DNN.

Table 8. Predicted class and probability of PM_{2.5} vulnerability.

Predicted class	Probability range	Vulnerability label	Number of grids (Total grids = 41557)	Ratio of grids (%)	Ratio of true grids (%)
False	0.000~0.500	Low	7689	18.50	-
	0.500~0.900	Intermediate	7621	18.34	22.49
True	0.900~0.990	High	10577	25.45	31.22
	0.990~0.997	Very High	9570	23.03	28.25
	0.997~1.000	Extremely High	6100	14.68	18.01

DNN 모델이 비산먼지 발생원이 많은 지역을 초미세먼지 취약지역으로 판단하는 데 매우 민감하게 반응함을 나타낸다.

다음으로 취약계층 이용시설, 대기오염 배출 사업장, 도로오염원 PM 배출량 변수가 상대적으로 높은 Mean absolute SHAP value를 나타내며, 이들 역시 초미세먼지 취약지역 예측에 중요한 영향을 미치는 변수군으로 확인되었다.

이러한 Mean absolute SHAP value를 통해 확인된 설명변수 중요도를 바탕으로 모델의 초미세먼지 취약지역 분류에 전반적으로 미치는 평균적인 영향력과 주요 요인을 확인하였다.

4.3 서울시 초미세먼지 취약성

표 8은 DNN 모델 결과를 활용하여 서울시 초미세먼지 취약성을 요약한 것이다.

연구에 사용된 DNN 모델은 초미세먼지 취약지역 분류 모델로서, Probability가 0.5 이하인 지역은 초미세먼지 취약지역이 아닌 지역(False Class)을 의미하며, 초미세먼지 취약성 'Low'로 정의하였다. Probability가 0.5 이상인 지역, 즉 초미세먼지 취약지역(True Class)으로 분류된 Grid는 총 41,557개의 Grid 중에서 약 81.5%에 해당하는 33,868개이다. 이 33,868개 Grid는 Probability Range에 따라 'Intermediate', 'High', 'Very High', 'Extremely High' 4단계로 세분하여 정의하였다.

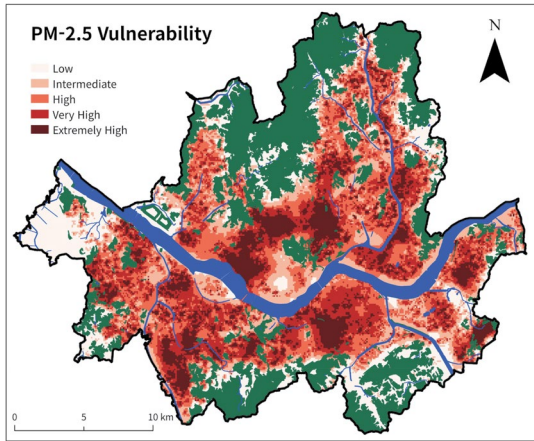


Fig. 14. This map shows the vulnerability of $PM_{2.5}$ in Seoul in 5 levels.

4. 4 서울시 초미세먼지 취약지역 매핑

그림 14는 DNN 모델의 초미세먼지 취약지역 Probability와 앞서 정의한 취약성 단계를 활용하여, 서울시 내 초미세먼지 취약지역을 시각화한 지도이다. 이 지도는 서울시 초미세먼지 취약성을 색상으로 구분하여 5단계로 표현하고 있다. 이런 시각화를 통해 각 Grid 단위로 초미세먼지 취약성 정도를 직관적으로 확인할 수 있다.

4. 5 서울시 초미세먼지 취약성

Extremely High 레벨 지역 분석

서울시 초미세먼지 취약성 분석 결과 중 가장 높은 위험 수준을 나타내는 Extremely high 레벨의 지역에 주목하였다. 특히, 현재 미세먼지 집중관리구역으로 지정되어 운영 중인 9개 구를 제외한 나머지 서울시 16개 구를 대상으로 분석을 수행하였다.

그림 15는 Extremely high 레벨에 해당하는 Grid만을 별도로 시각화한 지도이다(그림 15(a)). DNN 모델의 Probability를 사용하여, 16개 구의 Extremely high 레벨 Grid의 DNN 모델 Probability 평균값을 계산하고, 이를 막대그래프로 나타낸 결과는 송파구, 용산구, 마포구, 종로구 순으로 나타났다(그림 15(b)).

이러한 분석 결과는 서울시 초미세먼지 대응 정책에 있어, 송파구, 용산구, 마포구, 종로구를 우선 대상

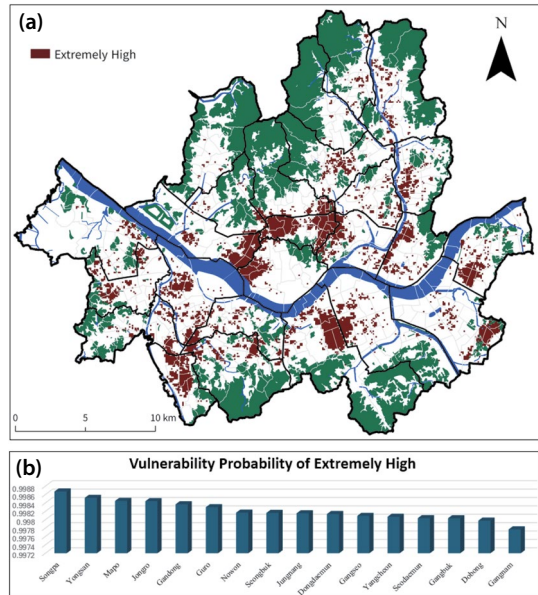


Fig. 15. (a) Map and (b) bar graph of extremely high vulnerability $PM_{2.5}$ in Seoul.

지역의 근거 자료로 활용될 수 있다.

5. 결 론

본 연구는 DNN 모델을 이용하여 서울시 초미세먼지 취약지역을 정의하고 분석하였다. 기존의 IPCC (2007) 기후변화 건강 취약성 평가모형에 기반한 초미세먼지($PM_{2.5}$) 취약성 연구에서 한 걸음 더 나아가, DNN 모델을 활용하여 서울시의 공간적 특성과 다양한 환경적·사회적 요인을 반영한 초미세먼지 취약지역을 도출하고자 하였다. 이를 위해 총 10개 설명변수와 1개 결과변수를 기반으로, $100\text{ m} \times 100\text{ m}$ Grid 단위로 하는 서울시 공간데이터베이스를 구축하였으며, 이를 DNN 모델에 적용하였다.

연구 결과, 본 연구의 DNN 모델은 초미세먼지 취약지역의 분류에서 매우 우수한 성능을 보였다. 테스트 데이터셋을 이용한 DNN 모델 검증 및 성능평가 결과, 정확도는 0.9667, 정밀도는 0.9688, 재현율은 0.9667, F1-score는 0.9832였으며, AUC 값은 0.9935였

다. 이는 모델의 초미세먼지 취약지역과 취약지역이 아닌 지역을 분류하는 성능이 정량적으로 매우 우수함을 의미한다. 이러한 결과는 초미세먼지와 관련된 취약지역을 기존의 연구보다 객관적이고 정확하게 평가할 수 있음을 나타낸다.

본 연구의 주요 기여는 다음과 같다.

첫째, DNN 모델을 이용한 데이터 기반의 분석을 통해, 초미세먼지 취약지역을 정량적이고 객관적으로 도출하는 새로운 접근 방법을 제시하였다.

둘째, DNN 모델 결과를 초미세먼지 취약지역 지도를 Grid 단위로 시각화함으로써, 복잡한 데이터를 직관적으로 쉽게 이해할 수 있다.

셋째, 다양한 환경적·사회적 요인과 공간적 특성을 반영한 데이터 기반 결과는 초미세먼지 대응 정책에 필요한 기초자료로 활용될 수 있다.

향후 연구에서는 본 연구에서 사용된 DNN 모델의 성능을 더욱 향상시키기 위해 다양한 머신러닝 기법들과의 비교 분석을 수행할 필요가 있다. 또한 초미세먼지 농도의 시계열적 변화와 기후 요소를 함께 고려한 연구가 이루어져야 할 것이다. 더 나아가, 서울시 외의 다른 대도시나 지역으로 확장 적용하여 대한민국 전역의 초미세먼지 취약지역에 대한 연구를 수행해야 할 것이다.

결론적으로 본 연구는 DNN 모델을 이용하여 초미세먼지 취약지역을 데이터에 기반하여, 정량적이고 객관적으로 분석함으로써, 통합적이고 과학적인 초미세먼지 대응 정책 접근을 가능하게 하였다. 이러한 연구 결과가 향후 환경 정책 수립 및 실행에 실질적으로 기여하여, 국민의 건강과 안전을 보호하는 데 도움이 될 것으로 기대된다.

감사의 글

이 연구는 국립환경과학원에서 주최한 제3회 대학(원)생 미세먼지연구아이디어공모전으로 수행되었습니다(NIER-2024-03-00-005).

References

- Cho, K., Lee, B., Kwon, M., Kim, S. (2019) Air Quality Prediction Using a Deep Neural Network Model, *Journal of Korean Society for Atmospheric Environment*, 35(2), 214-225, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2019.35.2.214>
- Du, S.S., Wang, Y., Zhai, X., Balakrishnan, S., Salakhutdinov, R., Singh, A. (2018) How many samples are needed to estimate a convolutional neural network?, *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 371-381. <https://dl.acm.org/doi/10.5555/3326943.3326978>
- Gil, J., Lee, M. (2021) Calculation of PM_{2.5} in Seoul 12-hours in Advance Using Simple Artificial Neural Network with Measurements of Background Sites, and Analysis of Contribution of Input Variables, *Journal of Korean Society for Atmospheric Environment*, 37(6), 862-870, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2021.37.6.862>
- Gong, S., Baek, W., Jung, H. (2022) Landslide Susceptibility Mapping Using Deep Neural Network and Convolutional Neural Network, *Korean Journal of Remote Sensing*, 38(6), 1723-1735, (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2022.38.6.2.12>
- Gribov, A., Krivoruchko, K. (2020) Empirical Bayesian kriging implementation and usage, *Science of The Total Environment*, 722. <https://doi.org/10.1016/j.scitotenv.2020.137290>
- Intergovernmental Panel on Climate Change (IPCC) (2007) *Climate Change 2007*. https://www.ipcc.ch/site/assets/uploads/2018/03/ar4_wg2_full_report.pdf (accessed May 16, 2025).
- Jung, S., Jun, K. (2020) A Study on the Health Vulnerability Assessment of Particle Matter in Incheon, *Journal of the Korean Society of Hazard Mitigation*, 20(1), 31-40, (in Korean with English abstract). <https://doi.org/10.9798/KOSHAM.2020.20.1.31>
- Kim, D., Choi, M., Yoon, B. (2019) Analysis of PM Hot-spot Emission Zone in Seoul Metropolitan Area, *Journal of Korean Society for Atmospheric Environment*, 35(4), 476-501, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2019.35.4.476>
- Ko, B., Lee, C., Kang, T., Choi, J. (2025) SHAP explanation of machine learning forecasting of PM₁₀ concentration, *The Korean Journal of Applied Statistics*, 38(1), 79-88, (in Korean with English abstract). <https://doi.org/10.5351/KJAS.2025.38.1.079>

- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E. (2017) A survey of deep neural network architectures and their applications, *Neurocomputing*, 234, 11-26. <https://doi.org/10.1016/j.neucom.2016.12.038>
- Mun, H., Jung, J. (2024) A study on the classification of vulnerable areas to PM_{2.5} according to urban characteristics based on vulnerability assessment, *Journal of Environmental Impact Assessment*, 33(5), 187-203, (in Korean with English abstract). <https://doi.org/10.14249/eia.2024.33.5.187>
- Park, J., Park, Y., Han, S., Byun, M., Lee, T., Son, Y., Lee, G., Kang, H., Kim, H., Park, D., Choi, W. (2023a) Source Characteristics and Their Changes in PM_{2.5} in Ulsan: Based on PMF Results for 2020 Observations and a Review of Source Profiles Reported in Korea, *Journal of Korean Society for Atmospheric Environment*, 39(1), 42-61, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2023.39.1.42>
- Park, J., Song, Y., Suh, M., Kim, C. (2023b) Prediction of PM_{2.5} using Super Learner Ensemble, *Journal of Korean Society for Atmospheric Environment*, 39(6), 1038-1049, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2023.39.6.1038>
- Ryu, J., Son, S., Kim, J. (2022) Prediction and Analysis of PM_{2.5} Concentration in Seoul Using Ensemble-based Model, *Korean Journal of Remote Sensing*, 38(6), 1191-1205, (in Korean with English abstract). <https://doi.org/10.7780/kjrs.2022.38.6.1.16>
- Wang, D., Thunell, S., Lindberg, U., Jiang, L., Trygg, J., Tysklind, M. (2022) Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods, *Journal of Environmental Management*, 301. <https://doi.org/10.1016/j.jenvman.2021.113941>
- World Health Organization (WHO) (2021) WHO global air quality guidelines. <https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdf> (accessed May 16, 2025).

Authors Information

- 박문조 (서울시립대학교 도시과학대학 공간정보공학과 학부생)
(mj0017179@uos.ac.kr)
- 이예경 (서울시립대학교 도시과학대학 공간정보공학과 학부생)
(lyk030905@naver.com)
- 김연준 (서울시립대학교 도시과학대학 공간정보공학과 학부생)
(uos.ac.kr@uos.ac.kr)
- 정동희 (국립환경과학원 기후대기연구부 대기환경과 전문위원)
(ehdgm6869@korea.kr)
- 안찬중 (국립환경과학원 기후대기연구부 대기환경과 전문위원)
(chanjung@korea.kr)
- 정형섭 (서울시립대학교 도시과학대학 공간정보공학과 교수)
(hsjung@uos.ac.kr)