

Research Paper

기계학습을 통한 토양오염물질 농도 예측 및 분포 매핑

강혜원* · 박상진** · 이동근***

서울대학교 농업생명과학대학 생태조경·지역시스템공학부*,
서울대학교 환경대학원 협동과정 조경학 및 대학원 융합전공 스마트시티 글로벌 융합**,
서울대학교 농업생명과학대학 조경·지역시스템공학부***

Predicting Concentrations of Soil Pollutants and Mapping Using Machine Learning Algorithms

Hyewon Kang* · Sang Jin Park** · Dong Kun Lee***

Department of Landscape Architecture and Rural System Engineering, Seoul National University*
Interdisciplinary Program in Landscape Architecture & Integrated Major in Smart City Global Convergence**
Department of Landscape Architecture and Rural System Engineering, Seoul National University***

요약: 본 연구는 사업시행이 환경에 미치는 부정적 영향을 최소화할 수 있는 방안을 강구하기 위해 환경영향평가 토양 부문을 강조하였다. 영향평가 절차에 대한 일련의 노력으로서 도시개발사업을 대상으로 하는 국가 인벤토리 기반 데이터베이스를 구축하였으며, 세 가지 기계학습 모델 성능 평가 및 토양오염물질 농도분포 매핑을 진행하였다. 여기에서, 가장 우수한 성능을 보여준 Random Forest 모델을 사용하여 대한민국의 수도권 지역을 대상 9가지 토양오염물질을 매핑하였다. 본 연구의 결과는 도시화가 가장 활발한 서울 지역에서 아연(Zn), 불소(F) 및 카드뮴(Cd) 농도가 상대적으로 우려되는 것을 발견하였다. 또한, 수은(Hg)과 크롬(Cr6+)의 경우 농도가 기준 이하로 검출되었는데, 이는 중금속 농도에 영향을 미치는 산업 및 농업 단지과 같은 오염원 부족이 원인으로 도출되었다. 토양오염물질 공간분포 매핑을 통해 토양특성 및 토지이용 유형과 오염물질 간의 유의한 상관관계를 유추하였다. 이를 통해 사업 현장 위치에 관한 토양오염 최소화 및 계획 결정에 대한 효율적인 토양관리 방안을 구축할 수 있을 것으로 기대한다.

주요어: 도시개발, 토양 중금속, 공간통계, 환경영향평가, 기계학습

Abstract: This study emphasized the soil of environmental impact assessment to devise measures to minimize the negative impact of project implementation on the environment. As a series of efforts for impact assessment procedures, a national inventory-based database was established for urban development projects, and three machine learning model performance evaluation as well as soil pollutant concentration distribution mapping were conducted. Here, nine soil pollutants were mapped to the metropolitan area of South Korea using the Random Forest model, which showed

First Author: Hyewon Kang, Tel: +82-2-880-4885, Email: kanghyewon@snu.ac.kr, ORCID: 0000-0003-1727-1213

Corresponding Author: Dong Kun Lee, Tel: +82-2-880-4875, E-mail: dklee7@snu.ac.kr, ORCID: 0000-0001-7678-2203

Co-Author: Sang Jin Park, Tel: +82-2-880-4885, Email: parkssang87@snu.ac.kr, ORCID: 0000-0003-4240-5682

Received: 8 April, 2022. Revised: 28 June, 2022. Accepted: 6 July, 2022.

the best performance. The results of this study found that concentrations of Zn, F, and Cd were relatively concerned in Seoul, where urbanization is the most active. In addition, in the case of Hg and Cr6+, concentrations were detected below the standard, which was derived from a lack of pollutants such as industrial and industrial complexes that affect contents of heavy metals. A significant correlation between land cover and pollutants was inferred through the spatial distribution mapping of soil pollutants. Through this, it is expected that efficient soil management measures for minimizing soil pollution and planning decisions regarding the location of the project site can be established.

Keywords: urban development, heavy metals in soil, spatial statistics, environmental impact assessment, machine learning

I. 연구 배경 및 목적

기후변화의 적응 및 완화를 위한 지속 가능한 전략으로서 토양은 생물 생산 및 분해 그리고 생물다양성 보전 및 탄소흡수원의 역할 등 주요 기능을 담당하고 있는 생태계의 필수 요소이다(MOE 2010). 물·에너지 관련 요소로서 녹지의 생산성 증대에 관여하고 온실가스의 저장소 역할을 담당하기 때문에 오늘날 글로벌 환경 이슈를 해결하기 위해 토양의 환경적 요인과의 상호작용을 이해하는 것이 요구되고 있다. 토양은 지구 중심 임계 요소이자 가장 큰 육상 생태계로서 다양한 규모로 토양 생태계 서비스에 관여한다. 또한, 토양오염물질 중 특히 중금속은 생물권에 노출되는 가장 주요한 성분으로 축적된 오염물질은 동·식물 및 인류를 위협할 수 있다(Yang et al. 2021). 즉, 토양오염물질은 환경의 위협 요소로 간주하며 그중에서도 중금속은 생태적 리질리언스에 대해 생리적 영향에 작은 요소만으로도 큰 영향을 미칠 수 있다(Azizi et al. 2022).

인간의 주된 인위적 활동을 포함하는 주거 및 교통, 그리고 산업 및 농·공업 활동에 따른 도시의 개발은 토양 내 중금속 함량에 상당한 영향을 미친다. 도시화는 토양 프로파일을 실질적으로 변경할 수 있으며 개발사업에 의해 토양특성 및 식물 기능을 변형시켜 토양의 특성에 영향을 줄 수 있다(Bae et al. 2015).

환경오염에 의한 생태계 훼손 최소화를 위한 일환으로써 환경영향평가 토지환경 토양 부문을 시행하고 있다. 환경영향평가란 도시개발과정에서 대상 사업

의 계획을 수립할 때 그 사업의 실행이 환경에 미치는 영향을 조사·예측·평가하여 해로운 영향을 피하거나 줄일 수 있는 방안을 강구하기 위해 수행되는 법률에 의한 평가 절차이다(K-eco 2020). 관련하여 국내외 환경영향평가 관련 토양오염물질 예측 연구현황을 알아보기 위해 선행논문을 검토하였다(Table 1).

토양오염물질 측정 관련 국내 선행연구를 검토한 결과, Yoon et al.(2019)은 복합 중금속 오염지역의 오염도 평가를 위한 지수화 모델을 제시하였으며, Jun(2011)의 연구에서는 GIS를 활용하여 공간적 분포도를 작성함으로써 대기 및 토양 내 오염 현황을 파악하였다. 그리고 Kim et al.(2017)의 지구 통계학을 기반으로 하여 하천을 따라 이동하는 중금속의 오염 농도를 예측하는 연구도 존재했다. 관련 국내 선행연구는 크게 오염된 토양을 현장 채취 및 물리·화학적 분석법에 기초하여 평가하는 연구가 주를 이루었다. 검토 결과, 국내에서 실제 시행된 연구는 미시적으로 파편화된 지역을 포함하였으며 생물학적 복원 적용을 위한 자원 개발을 목표로 하였다. 환경영향평가 전략의 연결을 통해 토양오염의 훼손을 최소화하기 위한 노력을 강조하여 확립하는 것이 미흡하였으며, 대상 사업의 토양 오염도를 예측할 수 있는 모델 개발을 목적으로 진행된 연구는 없었다. 또한, 기계학습 알고리즘을 활용하여 토양오염물질 농도의 분포를 예측하는 연구는 아직 존재하지 않는 것으로 나타났다.

반면 국외 선행연구에서는 토양의 중요성을 강조하며 오염물질 및 중금속 축적량을 예측하고 분포를 예측하는 연구가 다수 진행된 것을 알 수 있었다. 기

Table 1. Literature review related to the soil pollutants analysis

Division	Author (Year)	Model	Study site
Domestic	Yoon et al. (2019)	Pollution Index Model	Chungcheongbuk-do Industrial Area
	Jeon (2011)	Geographic Information System (GIS)	Daejeon
	Kim (2016)	GIS, Kriging	Samcheok, Gangwon-do
Foreign	Knoll et al. (2019)	GIS, Multiple Linear Regression (MLR), Classification And Regression Tree (CART), Random Forest (RF), Boosted Regression Tree (BRT)	Hesse, Germany
	Shen et al. (2020)	GIS, RF	USA
	Golia et al. (2021)	Robust quadratic regression	Thessaly Area, Greece
	Zhou et al. (2021)	GIS, Support Vector Machine (SVM), RF, BRT	Switzerland
	Forkuor et al. (2017)	MLR, RF, SVM, stochastic gradient boosting	Ioba Province, South-Western Burkina Faso
	Liang et al. (2019)	XGBoost	China
	Feng et al. (2020)	Cellular Automata	China
	Cambou et al. (2018)	GIS	New York, Paris
	Hengl et al. (2017)	XGBoost	Worldwide
	Mahmoudzadeh et al. (2020)	k-nearest neighbor (kNN), RF, Cubist (Cu), XGBoost, Support Vector Regression (SVR)	Western Iran
	Rosset al. (2019)	Bootstraps	Australia
	Zhang et al. (2020)	Artificial Neural Network, RF, SVM	Anhui Province, Eastern China
	Yang et al. (2021)	CART, kNN, Linear Regression, Ridge regression, Stochastic gradient descent regressor, SVR, Extremely Randomized Trees, Gradient Boosting Decision Tree, RF, XGBoost	Worldwide
	Gomes et al. (2019)	Cu, RF, SVM, Generalized Linear Models	Brazil
	Ahriwal et al. (2021)	XGboost	Indian Himalayan Region
Wang et al. (2020)	RF, Land Use Regression	Tianjin City, China	

계학습을 토대로 다양한 종류의 토양오염물질 흡착 용량을 예측 및 평가하였는데 여기에서 기계학습 알고리즘은 기존 데이터의 패턴을 통해 미래를 예측하여 관련 연구 활용에 용이하다. 지질학, 임업, 수문학, 농업 및 기타 분야에서 기계학습을 채택하여 상당한 데이터를 수집하여 그 정보를 기반으로 최적의 통합, 분석 및 의사결정을 내린다. 기계학습은 대규모 데이터 세트에서 규칙과 패턴을 파악하는 데 도움이 되는 새로운 기술로 널리 사용되고 있다(Du & Jeffrey 2007).

위와 같이 과학적 근거가 있는 오염 측정 지점 선정 방법과 정량적 지표 개발하는 연구가 국내에는 아직 진행되지 않았으며 관련 사례도 여전히 미흡하다.

또한, 국외와 비교하여 토양오염물질 공간분포를 예측하는 연구도 전무한 실정이다. 국내에서도 토양에 대한 이해도를 높이고 오염물질을 저감할 수 있는 계획 수립에 대한 적극적인 검토가 필요할 것으로 판단된다. 따라서 본 연구는 대한민국 서울·경기권을 대상으로 환경영향평가정보지원시스템(EIASS: Environmental Impact Assessment Support System)의 도시개발 사업을 대상으로 토양오염물질 현황을 조사하고 토양 오염 데이터를 구축하였다. 대상 지역의 사업개발 과정에서 효율적인 토양관리 방안을 구축할 수 있도록 대상 물질 오염도를 예측할 수 있는 모델을 개발하고자 하였다.

II. 연구방법

1. 연구대상지 및 분석범위

대한민국 수도권은 인천광역시, 그리고 경기도를 포함하여 수도 서울을 중심으로 형성되는 광역 대도시권이다. 대상 지역의 평균 연간 기온은 12.5°C, 평균 연간 강우량은 1,450mm이다(KMA 2020). 또한, 11,865.7km²의 면적에 2022년 1월 기준 약 2,603만 명의 인구가 거주하여 면적 상 전국의 약 11.2%에 불과하지만 전 국민의 절반 이상이 수도권에 밀집되어 있다. 과도하게 집중된 집약적 도시인 수도권은 산업 중심 지역으로 주거, 교통 및 각종 인프라의 유치로 수도권 집중화 현상이 일어난다.

이에 본 연구는 도시화가 활발하게 이루어지며 토지이용 변화가 많은 대한민국의 수도권을 연구 대상으로 선정하였다. 또한, 분석 범위는 토지피복분류에 따른 영향을 정량화하기 위해 대상 사업의 토양 추출 지점에서 최소 100m 이상의 버퍼(buffer area)를 포함하도록 하였다(Figure 1). 데이터의 가공 및 구축은 지리정보시스템(ArcGIS 10.5)을 이용하였다.

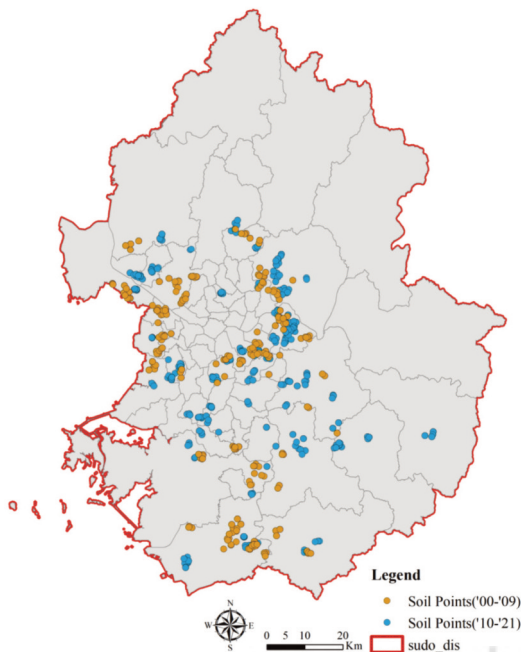


Figure 1. Location of the study area and the distribution of sampling points.

2. 연구 방법

토양오염물질 농도 예측을 위한 연구 흐름 기본 접근 방식은 다음 3가지 주요 단계로 구성된다. 1) 데이터 수집 및 전처리 과정을 거친 후, 2) 다양한 기계학습 알고리즘의 정확도 분석을 수행한다. 그리고 3) 최종적으로 매핑을 통한 토양오염 가능성을 예측한다. 먼저 ArcGIS 10.5을 이용하여 대상 지역인 수도권 영역의 지도 및 각 포인트를 구축하였으며 해당 데이터를 가공 및 추출하였다. 그리고 잠재적 토양오염의 변동 예측 및 평가를 위한 3가지 기계학습 기법 NB (Naive-bayes), kNN (K-nearest neighbor), RF (Random Forest)를 활용하여 최적의 성능을 가진 알고리즘을 최종 선정하였다. 대상 토양오염물질은 총 9가지로 카드뮴(Cd), 구리(Cu), 비소(As), 수은(Hg), 납(Pb), 크롬(Cr6+), 아연(Zn), 니켈(Ni), 불소(F)에 대해 공간분포를 예측하고 매핑하는 단계를 거쳤다(Figure 2).

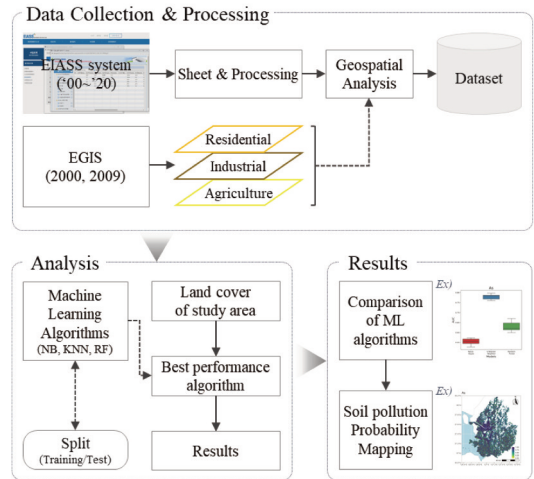


Figure 2. Research flow for the prediction of soil pollutants.

1) 데이터베이스 구축

주요 입력 자료는 환경영향평가정보지원시스템 (<https://eiass.go.kr/>)으로부터 제공되는 국가 인벤토리 데이터를 활용하였다. 17개의 사업유형 중 도시의 개발 평가 구분에 따른 도시개발사업 271건을 대상으로 토양오염 평가서 추출 정보로부터 데이터를 수집하였다. 본 연구에서 설정한 기간인 2000년부터

Table 2. Descriptive statistics of soil pollutants of the literature review

Author (Year)	Value	Cd	Cu	As	Hg	Pb	Cr6+	Zn	Ni	F
Azizi et al. (2022)	Min		<u>8</u>						34	
	Max		42						153	
	Mean		21.6						91.82	
Ma et al. (2016)	Min	0.025		<u>0.55</u>		<u>9.64</u>				
	Max	0.164		9.61		33.53				
	Mean	0.079		3.19		23.11				
Jia et al. (2019)	Min	<u>0.003</u>			<u>0.008</u>					
	Max	114			7					
	Mean	0.3			0.2					
Zhang et al. (2020)	Min	51	15.1	3.5	10	17.5	35.8	<u>34.5</u>	<u>11.5</u>	
	Max	<u>791</u>	92.7	<u>40.2</u>	<u>1867</u>	90.8	112.2	421	59.4	
	Mean	123.4	28.7	10.2	76.8	29.4	69.5	64.6	26.7	
Zhang et al. (2021)	Min	0.066	8.3	4.32	0.009	13.8	53	37.1	18	
	Max	0.304	324	22	0.187	30.2	102	137	44.3	
	Mean	0.14	21.1	9.86	0.025	19.4	68	61.5	26.4	
Wang et al. (2020)	Min	0.11		3.8	0.02	16.71	52.82	54.23		
	Max	1.8		15.84	2.99	81.12	<u>230.3</u>	368.8		
	Mean	0.32		10.39	0.24	33.14	93.5	111.8		
Li et al. (2019)	Min	2.0	33			92	<u>32</u>	129	37	<u>342</u>
	Max	36.7	<u>1359</u>			<u>1113</u>	177	<u>4774</u>	<u>206</u>	<u>2664</u>
	Mean	14.9	354			369	58	1180	74	883

2021년까지의 대상 물질 추출 포인트 지점 총 2,443 개를 구축하였다. 본 데이터는 각 사업에 대한 식별 코드·사업 시작 및 종료 연도·조사 지점 좌표·중금속 요소별 결과 농도 및 추진내용을 포함하며, 토양오염 물질 총 17가지 항목 각각의 농도와 해당 추출 지점 좌표 데이터를 시계열적으로 정리하였다. 본 연구에서는 환경부 토양환경보전법 토양 오염 규제물질로 지정된 항목 중 중금속 및 불소 등 총 9가지(Cd, Cu, As, Hg, Pb, Cr6+, Zn, Ni, F)를 분석 및 평가 대상 항목으로 선정하였다.

데이터 구축 과정에서 EIASS 환경영향평가서 추출 정보와 원문 정보 좌표의 차이로 토양 현지 조사 지점의 재검토가 필요했다. 원문과 상이한 좌표를 가지는 데이터는 보고서의 토양 현지조사 지점의 주소를 수집했다. 그리고 구글맵(Google maps)에서 공간 좌표화하여 해당 데이터를 재구축하였다. 2000년부터 2009년까지의 데이터 1,092개 및 2010년에서 2021년까지의 데이터 1,351개의 값을 최종적으로 생

성하였다.

또한, 시계열 분석을 통해 대상 항목의 시간적 범위 측정 지수에 대한 상이성을 발견하였다. 국내 토양오염공정시험기준이 2010년 용출법(약산 추출)에서 전함량 추출법(왕수 추출)으로 변경되어 관련 환경기준 농도가 급격하게 상승했고, 동일한 시료를 분석하더라도 전함량 추출방식을 사용했을 경우 해당 항목의 농도가 높게 나타날 수 있다. 이는 구축된 초기 빅데이터의 17가지 토양 속 물질 농도의 대부분에서 시계열 경향 차이를 확인할 수 있었다. 분석 중 오류 데이터에 대한 우려에 따라 대상 물질별 적정 값에 대한 데이터 정제가 필요했다. 또한, 토양오염물질 중 중금속의 경우 토양 중 분포는 지질학 및 지리학 조건이 모두 관계되며, 지구화학적 측면에서 자연적으로도 상당한 수준의 중금속에 노출될 수 있다(KIMG 2000). 따라서 중금속에 대한 기준 설정을 위해서는 토양 중 농도 분포에 대한 조사가 필요하였으며, 연구 대상 지역 및 토지이용에 따른 분포 실태를

Table 3. List of input data

Category	Variables		Period	Source
Dependent	Soil	EIASS	2000-2021	EIASS
Independent	Land Cover	Residential area	2000, 2009	Environmental Geographic Information Service (EGIS)
		Cropland		
		Industrial area		

Table 4. Mean value by soil pollutants (Units: mg/kg)

	Cd	Cu	As	Hg	Pb	Cr6+	Zn	Ni	F
Mean	0.31	17.46	2.03	0.06	15.48	0.15	66.33	12.05	110.67

조사하여 기초 자료로 이용하고자 하였다. 따라서 지점별 토양오염물질 농도 수준을 알아보기 위해 관련 국내외 선행연구를 검토하였다(Table 2). 이에 따라 토양오염물질 농도가 연구된 국내외 논문에서 각 중금속 및 오염물질의 농도 값을 추출하였고 동시에 환경부 자료 토양측정망 운영 결과 보고서를 통해 토양오염물질 농도 적정 수준을 정리하였다. 그리고 최종 선택한 9가지의 항목별 최대, 최소, 그리고 평균값에 대한 범위를 설정하여 범위를 벗어나는 수치를 검토 대상에서 삭제하였다. 표에 기재된 이탤릭체 및 밑줄은 토양오염물질별 최소, 최대 수치를 나타낸다.

토양오염 공간분포를 예측하기 위해 주요 입력 매개변수로 토지피복자료를 활용하였다. EIASS 오염물질별 조사 지점 및 환경부에서 제공하는 토지피복 데이터 세트를 기반으로 주거지역·농업지역·공업지역을 추출해 각각의 토지피복 별 면적과 해당 퍼센트 데이터를 구축하였다. Li et al. (2014)은 토양의 중금속 오염은 주로 농업, 도시화, 주거 폐기물, 산업화 및 광업의 인위적인 활동이 원인임을 확인하였다. Table 3은 구축된 데이터의 정보를 나타낸다.

2) 토양 중금속 및 불소 분포 예측

토양오염물질 농도의 공간 예측을 위해 대상지를 200 m × 200 m 그리드를 생성하여 진행하였다. 토지피복지도에 기반하여 토양오염물질 조사 지점으로부터 반경 100 m의 피복 정보를 추출하였으며 피복 인자를 적용하여 사업 시행 주변부의 토양 오염도를 동시에 파악하고자 하였다. 토지 피복별 지점에 대한 버퍼 설정 및 해당 면적 계산 또한 ArcGIS 10.5를 통

해 수행되었다.

본 연구는 토양에서의 오염물질 농도의 공간적 분포를 예측을 위해 세 가지 기계학습 모델을 이용하였다. NB, kNN 및 RF 세 가지 기계학습 기법의 성능을 비교하였으며 최고의 성능을 가진 알고리즘을 선정하여 최종 토양오염 예측 매핑을 진행하였다. 분석은 Python 'sklearn' 패키지를 활용하였으며, 모델의 성능을 평가하기 위해 학습(70%, n=737) 및 테스트 데이터 세트(30%, n=316)로 나누어 8회 반복 진행되었다.

토양오염물질의 공간 예측을 위한 값을 산출해 두 가지 민감도로 분류할 때 해당 값에 대한 판정 기준 값이 필요하며 본 연구에서는 이를 물질별 평균값(Mean)으로 설정하였다(Table 4). 많은 연구에서 원소의 평균값을 평가 기준으로 활용하고 있었다(Varol, 2011). 오염 물질별 평균값을 기준으로 관찰된 농도가 해당 수치보다 높으면 상대적으로 환경에 부정적 영향을 미칠 수 있으며, 낮은 범위에 속할 때 심각한 우려로 해석하였다.

III. 연구 결과 및 고찰

1. 기계학습 기법의 성능 비교

본 연구에서는 각 모델의 성능을 평가하기 위해 Receiver Operating Characteristic (ROC) 곡선을 사용하였는데, 대부분의 기계학습을 이용한 모델링 연구에서 ROC가 모델의 성능을 평가하는 도구로써 사용된다(Pham et al. 2016b; Shahabi & Hashim 2015). 그리고 ROC의 필수적 부분으로서 모델의 성

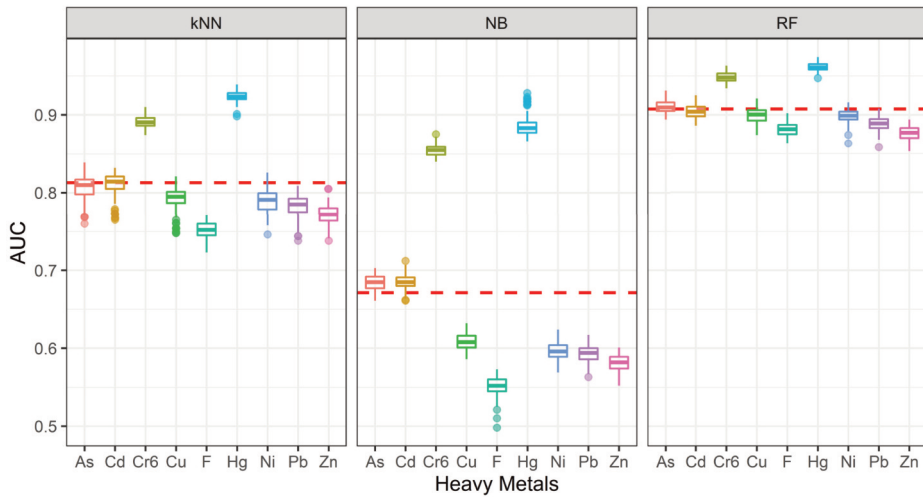


Figure 3. Comparison of performance of each machine learning algorithms (NB, kNN, RF) for pollutants in possibility of soil contamination (The red dotted line: the average AUC value of each machine learning algorithms).

Table 5. Validation results of the machine learning algorithms with average AUC value 8 repetitions of model training

Element	NB (Naive-bayes)	kNN (K-nearest neighbor)	RF (Random Forest)
Cd	0.69	0.81	0.90
Cu	0.61	0.79	0.90
As	0.68	0.81	0.91
Hg	0.89	0.92	0.96
Pb	0.59	0.78	0.89
Cr6+	0.85	0.89	0.95
Zn	0.58	0.77	0.88
Ni	0.60	0.79	0.90
F	0.55	0.75	0.88

능을 평가하는 데 일반적으로 사용되는 정량적 측정 기준인 Area Under the ROC Curve (AUC) 성능지표를 사용하여 AUC값 비교를 통한 모델 정확도를 평가하였다(Figure 3). 여기에서 AUC 수치가 높다는 것은 클래스를 구별하는 모델의 예측 성능이 높다는 것을 의미한다. 중금속에 대한 NB, kNN, RF 모델 결과에 따르면, 9가지 모든 중금속에서 RF 모델이 평균적으로 가장 높은 예측 성능을 달성하였다. 그리고 kNN 및 NB 순으로 일정한 모델 성능을 보여주었다(Table 5).

2. 토양오염물질 농도 공간 예측

3가지 기계학습의 성능 비교를 통해 가장 우수한

성능을 보여준 RF 모델을 사용하여 수도권 지역의 중금속과 불소 농도의 공간 예측분석을 진행하였다. 공간분포의 값은 0에서 1 사이의 값을 가지며, 1과 가까울수록 상대적으로 중금속의 농도가 높고 0에 가까울수록 상대적으로 중금속의 농도가 낮다고 해석할 수 있다(Figure 4).

결과를 보면 대체로 서울과 같이 주거, 농업 및 공업지역이 밀집해 있는 지역에서 토양오염물질 농도가 상대적으로 높은 경향을 가지는 것을 확인할 수 있다. 특히 Zn, F 및 Cd에서 이러한 현상이 두드러진다. 하지만 농도가 기준 이하로 검출된 중금속 Hg 및 Cr6+의 경향 또한 분명했다. 토양환경센터로부터 제공되는 오염 규제 물질 및 기준에서 오염물질 리스트 중,

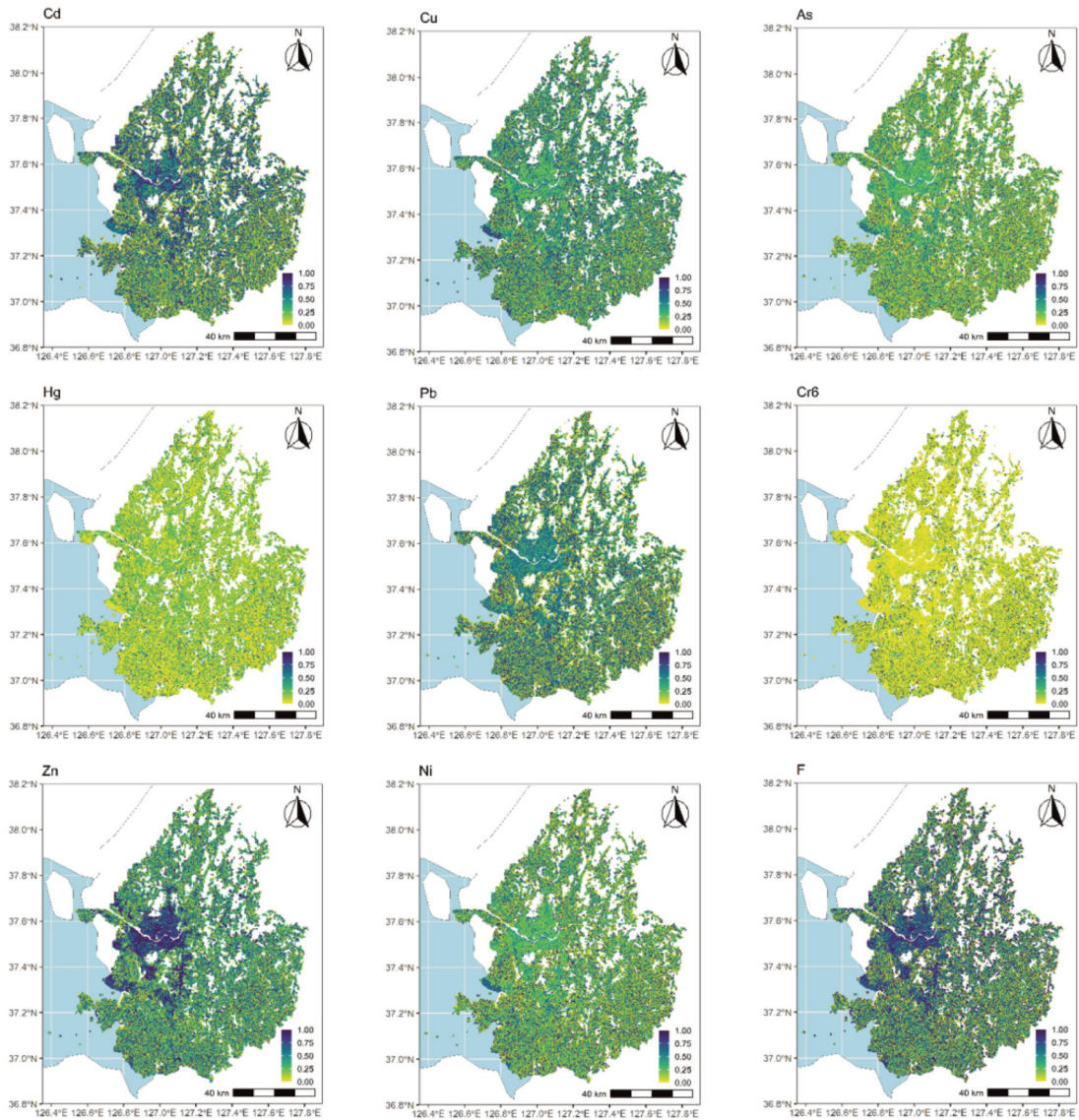


Figure 4. Spatial prediction of each heavy metals and F in soil.

Hg와 Cr6가 비교적 낮은 농도를 보였던 결과와 유사하다고 볼 수 있다. 동시에 공장 지역 및 폐기물 처리 시설과 같이 Hg 오염농도에 영향을 미치는 오염원의 부족이 해당 중금속 농도가 기준 이하의 결과로 도출되는 것을 예측할 수 있다. 서울을 포함한 경기 남부 및 수도권 전역에 비교적 높은 농도로 고르게 분포하고 있는 Cd는 토양오염 우려기준에 유의성을 보였다. 유사한 결과를 보인 As, Ni 및 Cu는 오히려 서울보다 경기지역에 높은 농도를 보였는데 이는 공업

단지에서 높게 검출되는 중금속으로 해당 토지피복은 서울지역보다 경기도에 조성되어 있음을 연계할 수 있다. 산업단지에서 주로 검출된다고 보고되는 중금속 Pb 경우에도 서울 전역에 밀집된 중위 농도와 경기도 북부지역에서 비교적 높은 오염도를 보였다.

특히, 주목해야 할 부분은 서울지역에 상대적으로 토양오염 우려기준의 농도가 나타나는 Zn 및 F이다. 두 토양 중금속 농도 모두 유사한 경향을 보였다. 먼저 수도권 전역에서 높은 농도로 분포함과 동시에 서

울권 수치는 1과 근접한 값으로서 토양오염 우려기준에 해당하는 수치를 보였다. 또한, 경기도 전체 면적 중 서북지역의 농도가 짙은 것을 확인할 수 있는데, 이러한 결과는 먼저 한국산업단지공단 지역의 위치 정보 및 현황 지도 자료를 기반으로 검토하였을 때 산업단지의 영향을 받은 것으로 해석할 수 있다. Zn은 산업 및 공업단지에 높은 배출량이 주로 감지되며 특히 과거 배출의 침강으로 인해 생태계에 위협을 줄 수 있다고 알려져 있다(Govil et al., 2008). 이는 서울시 내 국가산업단지 조성 비율이 높은 것과 연계할 수 있다. 정태욱 외(2015)의 연구에서도 국가산업단지에서 Zn의 농도가 높게 나타났으며 이는 금속산업과 자동차 등에서 발생하는 오염물질의 영향을 받은 것으로 해석하였다. 또한, 한국지질자원연구원에 따르면 F에 의한 오염 토양은 일반적으로 토양이나 암석에 함유되어 있는 형석 광물이 건축 및 공사와 같은 개발 사업에 의해 지표면이나 수계에 노출되면서 발생한다고 알려져 있다(KIGAM 2007). 따라서 이전 연구에서 발표된 바와 같이 토양특성 및 토지이용 유형이 토양 금속 및 오염물질과의 상관관계에 어느 정도 영향을 미칠 수 있다고 유추하였다(Li et al., 2015). 이러한 영향의 결과로 급격한 택지개발사업에 따른 도시화 지역의 토양오염물질의 농도가 높게 나타난 것으로 평가할 수 있다.

IV. 결론

토양오염은 환경에 심각한 문제로 대두되고 있다. 토지이용 별 토양오염물질 농도 증가 기여도에 따라 육상 생태계 및 토양 환경을 정량적으로 모니터링하고 평가하기 위해서는 동태를 파악할 수 있는 정보가 필요로 한 시점이다. 본 연구는 토양 속 유해 중금속 및 불소 총 9가지에 대한 공간분포 예측을 통해 기초 자료를 제공함으로써 관리 방안을 마련하고자 하였으며 환경영향평가 제도에 추가로 제공할 수 있는 기반을 재고하였다. EIASS에서 수집된 국가 인벤토리를 기반으로 토양오염물질 데이터를 구축하고, 급속하게 도시화가 진행된 지역에서 대상 물질의 공간분포

를 예측했다. 그리고 잠재적 오염 변동 예측을 위해 3가지 기계학습 기법의 성능 비교를 통해 선정된 모델을 활용하여 공간분포 예측을 하였고, 이는 국내 최초로 토양 중금속의 공간분포 매핑의 연구라는 점에서 의의가 있다. 연구 결과는 주거 및 농공업 지역의 비율이 높은 도시화된 지역일수록 토양오염이 높게 노출되었으며, 이는 도시 토양 중금속은 일반적으로 농업 및 공업지역에서 유래된다는 점에서 국외 선행 연구 결과와도 유사함을 시사한다.

그럼에도 불구하고 본 연구는 몇 가지 한계점을 지닌다. 첫째, 대상 물질 중 중금속은 오염원에 따라 그 농도와 범위가 상이하기 때문에 본 연구 결과를 향후 범용적으로 적용하는 것은 한계가 있을 수 있다. 둘째, 입력 자료에서 토지피복 중 주거, 농업, 산업지역 세 가지만 선택하여 분석을 진행하였는데, 이는 토양의 불균질성을 고려하지 못할 수 있다. 마지막으로 데이터베이스 구축에서 설정한 시간적 범위 내 오염물질 농도에 오류가 있을 가능성이 있다. 국내 토양오염 공정시험기준의 변경에 따라 관련 환경기준 농도가 높게 나타나기 때문인데, 향후 연구에서는 2010년 이후의 농도 수치를 시계열적으로 비교 및 검토하는 것이 필요하다.

결론적으로 토양오염은 다른 환경오염에 비해 복원하는 데 많은 시간과 비용이 드는 뿐만 아니라 지하수를 통한 2차 오염까지 야기할 수 있기 때문에 사전에 예방하는 것이 필요하다. 이에 토양 중금속 및 기타 오염물질 공간분포를 매핑하는 것은 도시의 개발에 따른 토양오염 우려기준을 초과하는 지역에서 특히 유용하며, 환경영향평가에 있어 신뢰할 수 있는 전략을 제공할 수 있을 것으로 사료된다.

사사

본 결과물은 환경부의 재원으로 한국환경산업기술원의 ICT기반 환경영향평가 의사결정 지원 기술개발 사업의 지원을 받아 연구되었습니다(No. 202000299 0009).

References

- Ahirwal J, Nath A, Brahma B, Deb S, Sahoo UK, Nath AJ. 2021. Patterns and driving factors of biomass carbon and soil organic carbon stock in the Indian Himalayan region. *Science of the Total Environment* 770: 145292.
- Azizi K, Ayoubi S, Nabiollahi K, Garosi Y, Gislum R. 2022. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. *Journal of Geochemical Exploration* 233: 106921.
- Bae J, Ryu Y. 2015. Land use and land cover changes explain spatial and temporal variations of the soil organic carbon stocks in a constructed urban park. *Landscape and Urban Planning* 136: 57-67.
- Cambou A, Shaw RK, Huot H, Vidal-Beaudet L, Hunault G, Cannavo P, Nold F, Schwartz C. 2018. Estimation of soil organic carbon stocks of two cities, New York City and Paris. *Science of the Total Environment* 644: 452-464.
- Yoon DH, Hong YK, Kim JW, Kim SH, Song GH, Lee KM, Kim SC. 2019. Evaluating Heavy Metal Pollution in Soil Using Pollution Index Model. *Korean Society of Soil Science and Fertilizer*, 168-168. [Korean Literature]
- Feng Y, Chen S, Tong X, Lei Z, Gao C, Wang J. 2020. Modeling changes in China's 2000-2030 carbon stock caused by land use change. *Journal of Cleaner Production* 252: 119659.
- Forkuor G, Hounkpatin OK, Welp G, Thiel M. 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PLoS One* 12(1): e0170478.
- Golia EE, Diakouloukas V. 2021. Soil parameters affecting the levels of potentially harmful metals in Thessaly area, Greece: a robust quadratic regression approach of soil pollution prediction. *Environmental Science and Pollution Research*, 1-18.
- Gomes LC, Faria RM, de Souza E, Veloso GV, Schaefer CEG, Fernandes FEI. 2019. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma* 340: 337-350.
- Govil PK, Sorlie JE, Murthy NN, Sujatha D, Reddy GLN, Rudolph-Lund K, Krishna AK, Rama MK. 2008. Soil contamination of heavy metals in the Katedan industrial development area, Hyderabad, India. *Environmental Monitoring and Assessment* 140(1): 313-323.
- Hengl T, Mendes de Jesus J, Heuvelink GB, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Antonio Guevara M, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* 12(2): e0169748.
- Jeong TU, Cho EJ, Jeong JE, Ji HS, Lee KS, Yoo PJ, Kim GG, Choi JY, Park JH, Kim SH, Heo JS, Seo DC. 2015. Soil contamination of heavy metals in national industrial complexes, Korea. *Korean Journal of Environmental Agriculture* 34(2): 69-76. [Korean Literature]
- Jia X, Hu B, Marchant BP, Zhou L, Shi Z, Zhu Y. 2019. A methodological framework for identifying potential sources of soil heavy metal pollution based on machine learning:

- A case study in the Yangtze Delta, China. *Environmental Pollution* 250: 601-609.
- Jun YJ. 2011. Study on characterization of air and soil pollution in Daejeon area using GIS. Master's thesis, Hanbat National University, pp. 1-44. [Korean Literature]
- Kim SM, Choi Y, Yi H, Park HD. 2017. Geostatistical prediction of heavy metal concentrations in stream sediments considering the stream networks. *Environmental Earth Sciences* 76(2): 1-18. [Korean Literature]
- KIMG. 2000. Natural geological mapping for natural environment.
- Knoll L, Breuer L, Bach M. 2019. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Science of the Total Environment* 668: 1317-1327.
- Korea Institute of Geoscience and Mineral Resources. 2007. A method for treating fluorine-contaminated soil induced by fluorite. [Korean Literature]
- Korea Meteorological Administration. 2021. 2020 Metropolitan Meteorological Administration Climate Data Collection, 11-1360619-000006-10. [Korean Literature]
- Li F, Huang J, Zeng G, Liu W, Huang X, Huang B, Gu Y, Shi L, He X, He Y. 2015. Toxic metals in topsoil under different land uses from Xiandao District, middle China: distribution, relationship with soil characteristics, and health risk assessment. *Environmental Science and Pollution Research* 22(16): 12261-12275.
- Li Y, Wang S, Nan Z, Zang F, Sun H, Zhang Q, Huang W, Bao L. 2019. Accumulation, fractionation and health risk assessment of fluoride and heavy metals in soil-crop systems in northwest China. *Science of the Total Environment* 663: 307-314.
- Li Z, Ma Z, van der Kuijp TJ, Yuan Z, Huang L. 2014. A review of soil heavy metal pollution from mines in China: Pollution and health risk assessment. *Science of The Total Environment* 468-469: 843-853.
- Liang Z, Chen S, Yang Y, Zhou Y, Shi Z. 2019. High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling. *Science of The Total Environment* 685: 480-489.
- Ma W, Tan K, Du P. 2016. Predicting soil heavy metal based on Random Forest model. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4331-4334, IEEE.
- Mahmoudzadeh H, Matinfar HR, Taghizadeh-Mehrjardi R, Kerry R. 2020. Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Regional* 21: e00260.
- Ministry of Environment. 2010. Summary of the results of the 2009 soil measurement network and soil pollution survey. [Korean Literature]
- Pham BT, Pradhan B, Bui DT, Prakash I, Dholakia MB. 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software* 84: 240-250.
- Shahabi H, Hashim M. 2015. Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Scientific Reports* 5(1): 1-15.
- Shen LQ, Amatulli G, Sethi T, Raymond P, Domisch S. 2020. Estimating nitrogen and phosphorus concentrations in streams and rivers, within a machine learning framework. *Scientific*

- Data 7(1): 1-11.
- Varol M. 2011. Assessment of heavy metal contamination in sediments of the Tigris River (Turkey) using pollution indices and multivariate statistical techniques. *Journal of Hazardous Materials* 195: 355-364.
- Vasenev VI, Stoorvogel JJ, Leemans R, Valentini R, Hajiaghayeva RA. 2018. Projection of urban expansion and related changes in soil carbon stocks in the Moscow Region. *Journal of Cleaner Production* 170: 902-914.
- Viscarra-Rossel RA, Lee J, Behrens T, Luo Z, Baldock J, Richards A. 2019. Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls. *Nature Geoscience* 12(7): 547-552.
- Wang H, Yilihamu Q, Yuan M, Bai H, Xu H, Wu J. 2020. Prediction models of soil heavy metal (loid) s concentration for agricultural land in Dongli: A comparison of regression and random forest. *Ecological Indicators* 119: 106801.
- Yang H, Huang K, Zhang K, Weng Q, Zhang H, Wang F. 2021. Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities. *Environmental Science & Technology* 55(20): 14316-14328.
- Yoon JK, Kim DH, Kim TS, Park JG, Chung IR, Kim JH, Kim H. 2009. Evaluation on natural background of the soil heavy metals in Korea. *Journal of Soil and Groundwater Environment* 14(3): 32-39.
- Zhang D, Tsai JJ. (Eds.). 2006. *Advances in machine learning applications in software engineering*. Igi Global.
- Zhang H, Yin A, Yang X, Fan M, Shao S, Wu J, Wu P, Zhang M, Gao C. 2021. Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. *Ecological Indicators* 122: 107233.
- Zhang H, Yin S, Chen Y, Shao S, Wu J, Fan M, Chen F, Gao C. 2020. Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. *Journal of Cleaner Production* 273: 122858.
- Zhou T, Geng Y, Ji C, Xu X, Wang H, Pan J, Bumberger J, Haase D, Lausch A. 2021. Prediction of soil organic carbon and the C: N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. *Science of the Total Environment* 755: 142661.