

논문

머신러닝을 활용한 악취물질의 농도와 민원 발생 예측

Prediction of Odor Concentration and Complaint Occurrences Using Machine Learning

김동한, 오용환, 서서희, 천소현, 손윤석*

국립부경대학교 지구환경시스템과학부

Dong-Han Kim, Yong-Hwan Oh, Seo Hee Seo, So-Hyeon Cheon, Youn-Suk Son*

Division of Earth Environmental System Science, Pukyong National University,
Busan, Republic of Korea

접수일 2025년 10월 9일
수정일 2025년 10월 31일
채택일 2025년 11월 2일

Received 9 October 2025
Revised 31 October 2025
Accepted 2 November 2025

*Corresponding author
Tel : +82-(0)51-629-6522
E-mail : sonys@pknu.ac.kr

Abstract With the advancement of industries and the densification of residential areas, odor issues have become increasingly persistent, leading to frequent complaints from nearby residents. This study aims to develop predictive models for estimating the concentration of odor substances emitted and dispersed from steelworks and forecasting the occurrence of related complaints based on these predictions. Four machine learning models commonly used for time-series data analysis were selected as candidates. The predicted concentration data were then used as new input to train classification models to predict the occurrences of odor complaints. For this classification task, five models were evaluated, and the one with the highest predictive accuracy was adopted as the final model. As a result, the XGBoost regression model was selected for predicting odor concentrations. This model achieved an improvement rate of over 57% in RMSE across all three sites, and an R^2 value of 0.738 or higher, indicating excellent performance. For predicting the occurrence of complaints, the XGBoost classifier was chosen, demonstrating a ROC-AUC value of 0.867 or higher, along with both reliability and accuracy exceeding 75%, showing stable performance. Based on these findings, odors generated from steelworks can be effectively managed, potentially resulting in a significant reduction in related complaints.

Key words: Odor, Complaint occurrences, Machine learning, XGBoost model, Prediction

1. 서론

산업의 발전으로 인하여, 산업단지 인근 주민들은 직·간접적으로 다양한 형태의 대기오염물질에 노출되고 있다(Lee *et al.*, 2013). 특히, 악취는 2016년부터 환경부에서 지정한 국가환경분야 5대 난제(악취, 미세 먼지, 가뭄, 녹조, 지반침하) 중 하나로, 삶의 질을 저해시키는 대표적인 요인으로 평가받고 있다. 악취는 사람의 후각을 자극성 물질이 자극하여 불쾌감 및 혐오감을 주는 냄새로 환경부에서는 2010년 이후 황화수소, 암모니아, 메틸메르캅탄 등 22종을 지정악취물질로 지정하여 관리하고 있다. 또한, 악취물질은 낮은 농

도에서도 주민들에게 심리적 영향을 미칠 수 있다(Rincón *et al.*, 2019). 이에, 정부 및 지방자치단체는 기술지원, 원인 규명 등 다양한 개선 노력을 시도하고 있으나, 민원 발생은 여전히 높은 수준을 기록하고 있으며, 실질적인 제어 방안은 미비한 실정이다.

특히, 제강 사업소에서 발생하는 악취에 대한 연구는 과거부터 꾸준히 진행되고 있다(Nordin and Lidén, 2006; Fehnel, 1942). 국내 비철금속 공업은 1970년대 초기부터 본격화되었으며, 급격한 성장을 바탕으로 국제 규격을 만족할 수 있는 설비를 갖추기 시작하였다. 그러나, 비철금속 제련 및 정련 등 제조 사업장에서는 공통적으로 쇠 냄새가 감지되었으며, 공정 특성

에 따라 타는 냄새와 기름 냄새가 인지되는 특성을 지니고 있다(Paeng *et al.*, 2006). 이러한 사업장에 발생된 악취문제는 단순히 제강 사업소에 국한되지 않고, 인근에 위치한 주거지, 학교 등에 영향을 미칠 수 있다. 이러한 주변 지역에 미치는 영향을 파악하기 위하여 악취의 확산을 모델링하는 연구가 활발히 진행되고 있다(Qu *et al.*, 2025; Choi *et al.*, 2022; Xu *et al.*, 2022; Kang *et al.*, 2020). 하지만 위 연구들은 단순히 악취물질의 농도를 예측하거나(Xu *et al.*, 2022; Kang *et al.*, 2020), 인공지능을 활용해 악취물질의 종류를 분류하거나(Qu *et al.*, 2025), 악취물질의 발생원을 찾는 등의 연구에 한정되어 있다(Choi *et al.*, 2022). 이와 더불어, 보다 최근에는 악취물질의 농도를 기반으로 민원 발생을 예측하는 연구도 수행되었지만, 예측 기간이 시간 단위로 매우 짧고, 이미 확산된 악취에 대해 발생하는 민원을 예측했다는 한계점을 내포하고 있다(Xiao *et al.*, 2024). 따라서, 본 연구에서는 악취물질의 농도를 장기간 예측하고, 이를 활용해 시간 단위 가 아닌 일 단위 민원 발생 여부를 예측하고자 한다.

본 연구에서는 악취물질의 농도(Odor Unit, OU) 데이터를 기반으로 머신러닝을 이용하여 제강 사업소에서 배출되는 악취 물질의 확산에 따른 악취물질의 농도를 예측하고 예측한 농도를 활용해 인근 주거지에서 악취로 인해 발생하는 민원을 예측할 수 있는 모델을 구축하고자 연구를 수행하였다.

2. 연구 방법

본 연구에서는 크게 악취물질의 농도와 민원이라는 두 가지 측면에 대해서 나누어 평가를 진행했다. 이를 위하여 악취물질의 농도를 예측하는 경우에는 회귀모델을 이용하였고, 민원 자료의 경우 발생 여부를 예측하기 위하여 분류모델을 이용하여 예측을 진행하였다.

2.1 데이터 수집

본 연구에서는 악취물질의 농도를 예측하기 위하여

2019년 12월 1일부터 2025년 4월 1일까지의 1시간 복합 악취물질 농도(OU) 자료를 이용했다. 해당 자료는 사업장에서 직접 센서를 사용하여 측정한 자료를 사용하였다. 자료는 제강사업소 주변에서 측정한 결과값을 사용했다. 측정 위치는 총 세 곳으로, 악취물질이 배출되는 배출구 주변 외벽(Site #1), 제강사업소 내 배출구 반대편의 외벽(Site #2), 그리고 인근 주거지(Site #3)에 설치되어 있다.

악취 민원 자료의 경우, 제강사업소를 관리하고 있는 구청에 접수된 동일 기간 동안의 민원 접수 현황 자료를 활용하였다. 이는 악취물질 농도 자료와는 다르게 1일당 접수된 민원의 횟수에 대한 자료이다.

악취물질의 비산을 정확히 예측하기 위한 주요 변수로 기상 자료를 사용했다. 기상 자료에는 강수량, 풍향, 풍속, 그리고 기온을 변수로 입력했다. 위 기상자료들은 기상자료개방포털의 방재기상관측자료를 이용했고 제강사업소와 인근 관측소의 거리는 약 1.5km였다(KMA, 2025).

2.2 자료 전처리

모델 학습과 평가의 재현성과 객관성을 확보하기 위해, 사전에 설정한 절차에 따라 전처리를 수행하였다. 또한, 모든 모델은 데이터를 학습할 때 모든 변수를 동일하게 적용시켜 학습시켰다. 우선적으로, 악취물질의 농도 자료는 시간당 자료이므로, 이 값을 일평균 값으로 변환하여 사용했다. 또한 자료 예측에 영향을 끼치지 않기 위해, 훈련 자료와 테스트 자료를 분류했다. 이를 위해, 훈련 자료의 기간은 2019년 12월 1일부터 2023년 12월 31일까지로, 테스트 자료는 2024년 1월 1일부터 2025년 4월 1일까지 자료로 설정했다. 또한, 본 연구에서 사용된 자료는 시계열 자료이다. 시계열 자료에서 가장 중요한 것은 시간 흐름에 따른 변화를 파악하는 것인데, 이를 위하여 본 연구에서는 이동창(Rolling window) 기법을 적용하였다. 이동창 기법은 일정한 길이의 분석 구간을 설정하고, 이를 시간 축을 따라 한 단계씩 이동시키며 통계적 특성을 분석하여 자료를 학습시키는 기법이다(Nam *et al.*, 2020; Li

et al., 2018; Liu et al., 2018). 이때, 이동창 기법의 기간으로 단기간 데이터를 학습하기 위하여 1일, 3일, 7일을 선정했고, 장기간 데이터 학습을 위하여 30일, 계절적 특성을 학습하기 위해 90일, 그리고 연간 데이터의 경향성을 학습하기 위해 365일을 기간으로 설정했다. 모델 성능을 검증하기 위해 훈련 데이터 내에서 5-겹 교차 검증(five-fold cross-validation)을 수행하였다.

2.3 모델 선택

본 연구에서는 악취물질의 농도를 예측하기 위하여 회귀모델을, 민원 발생 여부를 판단하기 위해서는 분류 모델을 사용했다. 회귀모델의 선택 기준은 Root Mean Square Error (RMSE)와 R^2 값을 바탕으로, RMSE 값이 기준선 RMSE, 즉 baseline RMSE와 비교했을 때 50% 이상 개선되고, R^2 값이 0.7 이상인 모델을 선정하였다. 분류모델의 경우, ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) 값을 기준으로 모델을 선정하였다. Baseline RMSE의 경우, 월별 학습 구간의 단순 평균값을 예측으로 사용하는 기준모형의 RMSE를 의미한다. ROC-AUC 값은 1에 가까울수록 적절하게 분류한다고 판단하는 값이므로 0.7 이상인 모델을 선정하였다. 하지만 ROC-AUC 값은 특정 임계값이 아닌 전체 임계값에 대한 수치이므로, 분류 후 다시 정확도(Accuracy)와 신뢰도(Precision) 값을 바탕으로 최종 모델을 선정하였다. 여기서 정확도는 전체 데이터 중 실제로 올바르게 민원이 발생했다고 판단한 비율을 의미하고, 신뢰도는 민원이 발생했다고 판단한 경우 중 실제로 민원이 발생한 경우를 평가한 비율이다. 본 연구는 PyCharm 개발 환경에서 Python 3.11를 사용하여 구현하였다. 사용한 주요 Python 패키지로는 데이터 조작을 위한 Pandas, 모델 구축 및 평가를 위한 scikit-learn, 데이터 시각화를 위한 matplotlib을 이용했다.

2.3.1 회귀모델

악취물질의 농도를 적절하게 예측하기 위해, 일반적으로 사용되는 다양한 머신러닝 알고리즘을 후보군

으로 선정한 후 기준에 맞는 최종 모델을 선정하였다. 후보군으로 선택한 모델은 Random Forest (RF), XG Boost (eXtreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), Support Vector Regression (SVR)이다.

RF 모델은 효과적인 지도 학습 앙상블 모델이다. 학습 과정은, 첫 번째는 학습 집합을 추출하는 것으로, 원래 학습에서 무작위로 선택한 부분 집합을 학습 데이터의 집합으로 설정한다. 다음으로, M 개 변수의 학습 집합에서 m 개의 특징을 무작위로 추출하여 의사결정 트리를 구축하는 것이다. 이후 분류는 최대 정보이득 특징을 선택함에 따라 노드를 분류하여 의사결정 트리를 완전히 성장시킨다. 이 두 단계를 반복하여 N 개의 의사결정 트리를 얻어 랜덤 포레스트를 형성한다. 분류 결과는 각 의사결정 트리의 개별 분류에 의해 결정된다. 수많은 의사결정 트리 투표로 인해 이상치에 대한 민감성이 본질적으로 낮아져 안정성이 향상된다(Shaikhina et al., 2019).

XGBoost는 그래디언트 부스트(Gradient Boost)를 기반으로 하는 의사결정트리 모델이다(Liang et al., 2020). 이 모델은 각 특징 분할마다 새로운 트리를 추가하여 마지막 예측의 잔차를 새로운 함수로 맞추는 것이다. 새로 학습된 트리의 리프 점수는 순차적으로 누적되어 해당 샘플의 예측값을 보정한다.

LightGBM은 새로운 의사결정 트리를 학습시켜 예측값을 실제 값에 가깝게 유지하는 방식으로, 그래디언트 부스팅의 의사결정 트리의 변형모델이다(Liang et al., 2020). 매 반복마다 기존 모델을 변경하지 않고 새로운 함수를 사용하여 모델에 추가하여 예측값을 실제 값에 가깝게 만든다(Shahhosseini et al., 2021).

SVR은 Support Vector Machine (SVM) 프레임워크에서 파생된 회귀 알고리즘이다(Alfasanah et al., 2025). SVR은 비선형적 매핑을 이용해 저차원 입력 공간의 데이터를 고차원 특징 공간으로 투영한다(Zhang and O'Donnell, 2020; Toivonen et al., 2007). 이러한 방식을 활용해 SVR은 복잡한 관계를 선형 모델로 표현할 수 있도록 할 수 있는 알고리즘이다(Zhang

and O'Donnell, 2020).

RMSE와 R^2 값에 따라 선정된 최종 모델은 XGBoost Regression 모델이 선정되었다.

2.3.2 분류모델

회귀모델로 예측한 악취물질의 농도를 이용해, 민원의 발생 여부를 예측하기 위해 이진분류 기법을 사용하였다. 이진분류는 민원이 발생했을 때를 1로, 발생하지 않았을 때를 0으로 판단한다. 이 데이터를 분류하기 위해, 후보군으로 총 5개의 모델을 선정했다. 후보군으로 선정된 모델은 RF Classifier, XGBoost Classifier, LightGBM classifier, SVM, Logistic Regression을 선정했다.

Logistic Regression은 사건이 일어난 확률과 일어나지 않을 확률의 비의 자연로그를 예측 변수들의 선형함수로 모델링하는 이진분류 기법이다. 회귀 계수의 지수를 취하면 입력 변수의 단위 변화가 사건의 승산에 미치는 비율을 얻을 수 있어 해석이 용이하다. 여러 예측 변수를 포함하면 다른 변수의 영향을 보정한 상태에서 각 변수의 승산비를 해석할 수 있으며, 변수 선택은 교란을 줄이고 추정의 정확도를 높이기 위해 중요하다. 이때, 최적의 계수는 반복적 근사 과정을 통해 추정된다(LaValley, 2008).

이렇게 선택된 후보군 중, ROC-AUC 값을 기준으로 모델을 선정했다. 그 결과 XGBoost Classifier 모델과 Logistic Regression 모델이 선정되었다. 이후 신뢰도와 정확도 두 값을 이용해 더 적절한 모델을 판단하고자 연구를 진행하였다.

2.3.3 스택킹 기법

스택킹 기법의 경우, 단일 모델 자체로 결과를 도출하는 것이 아닌, 여러 모델들을 계층적으로 학습시키는 앙상블 학습 프레임워크이다(Kim *et al.*, 2021). 스택킹 기법은 기초 학습기의 다양성을 활용하여 개별 모델 편향과 분산을 완화한다. 이러한 기법을 머신러닝 모델에 적용하면 예측모델의 성능을 높일 수 있다(Pavlyshenko, 2018). 이에 따라, 본 연구에서는 악취물질의 농도를 가장 잘 예측했던 XGBoost Regression

모델을 이용하여 악취로 인한 민원 발생 횟수를 우선 예측했고, 민원 발생간의 관계를 잘 예측한 XGBoost Classification 모델과 Logistic Regression 모델을 각각 스택킹하여 민원 발생의 정확도를 더욱 향상시키고자 하였다.

3. 결과 및 고찰

3.1 악취물질 농도 예측 결과

그림 1은 각 지점에서 측정된 악취물질의 농도와 해당 일에 발생된 민원의 횟수를 나타낸 그래프이다. 이 그래프에서 나타나는 것과 같이, 악취물질의 농도가 높은 날임에도 민원이 접수되지 않은 경우가 있고, 반대로 악취물질의 농도가 낮게 측정된 날에도 민원이 접수된 경우가 있었다. 따라서, 단순한 비교만으로는 악취 농도와 민원 발생 간의 명확한 상관성은 확인할 수 없었다.

그림 2는 후보군으로 선정했던 모델들에 대한 예측 성능을 보여준다. 지점 #1의 baseline RMSE는 1.211이었고, RF 모델에서는 RMSE는 0.905로, 매우 낮은 개선율을 기록했고, 뿐만 아니라 R^2 역시 0.429로 낮은 수치를 보였다. LightGBM의 경우, 0.867, 0.476으로 RF보다 개선된 결과를 보였지만, 여전히 적절하지 않은 수치를 기록했다. 반면 SVR의 경우, RMSE가 0.765로 다른 두 모델보다 더 높은 개선율을 보였지만, R^2 값이 0.587로 여전히 낮은 값을 기록했다

가장 높은 개선율과 R^2 값을 기록한 XGBoost Regression의 경우, 그림 3에서 나타낸 것과 같이, 지점 #1에서 RMSE가 0.519로 57%, R^2 값은 0.816으로 매우 높은 수치를 나타냈다. 뿐만 아니라, 지점 #2, #3의 baseline RMSE는 각각 1.039, 1.578이었고, 각각 65%, 61%의 개선율을 보여 최종 RMSE는 0.363, 0.614를 나타냈다. 더욱이, R^2 값 역시 각각 0.797, 0.738로, 앞의 3가지 모델보다 높은 값을 나타냈다. 이에 복합 악취물질을 예측하는 모델로 XGBoost Regression 모델을 선정했다.

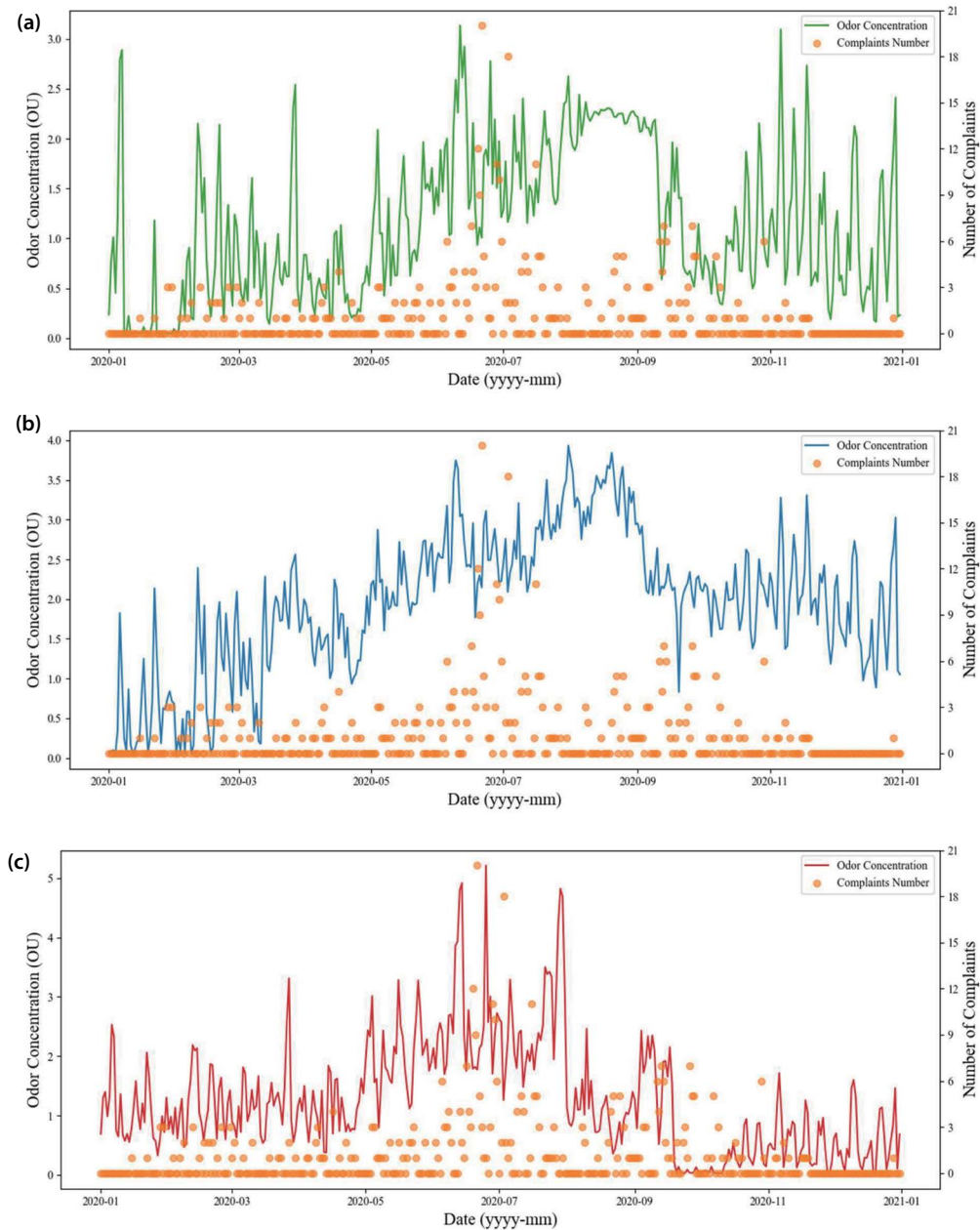


Fig. 1. Daily variations in the odorous compound concentrations and the complaint number in 2020 ((a) Site #1, (b) Site #2, and (c) Site #3).

3.2 민원발생 예측 결과

민원 발생에 대한 예측의 경우, 다섯 가지의 후보군 모델을 선정했고, 그 중 두 가지의 최종 모델을 선정

했다. 그 결과로, 그림 4에 나타난 것과 같이 Logistic Regression, XGBoost Classifier 두 모델이 선정되었다. 두 모델의 ROC-AUC는 각각 0.726, 0.764로, 높은 수

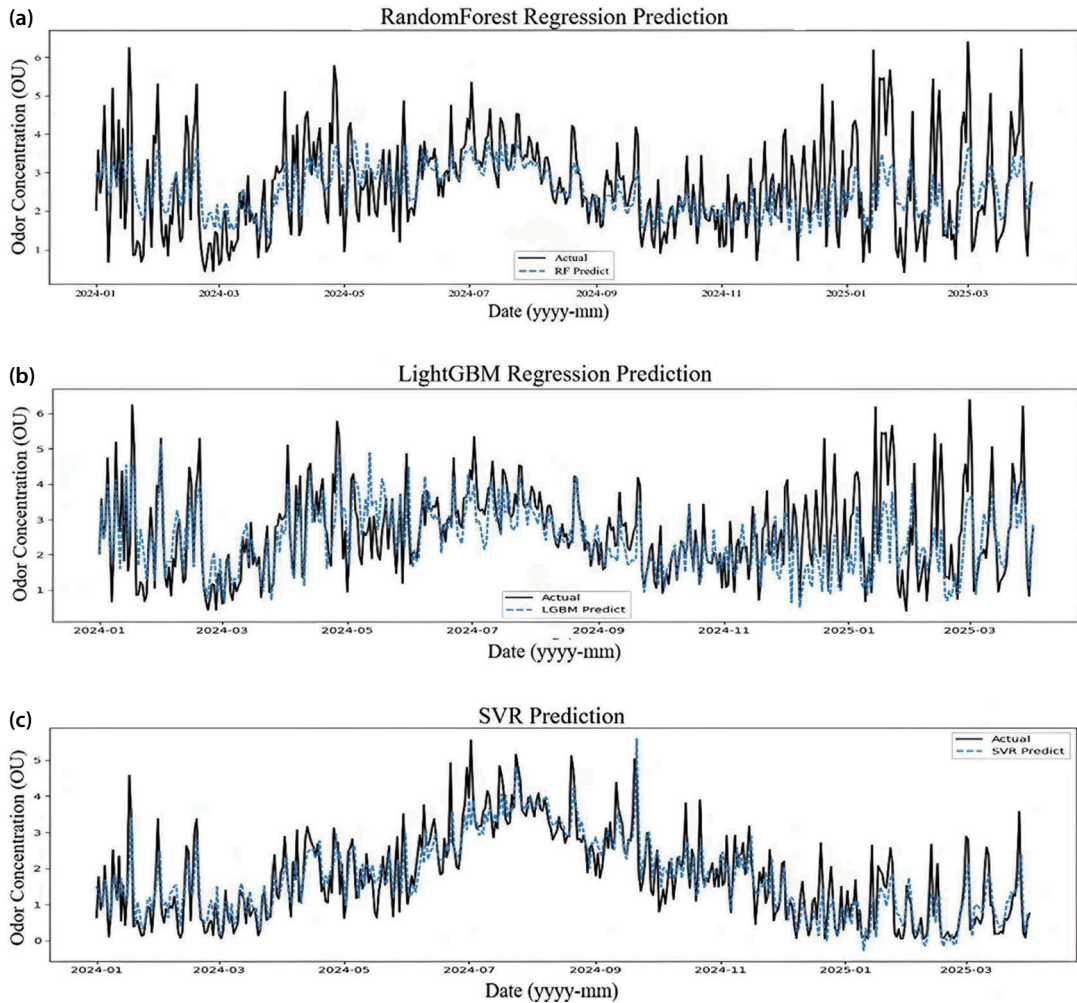


Fig. 2. Daily measured and predicted odorous compound concentrations using various models ((a) Random Forest, (b) LightGBM, and (c) SVR).

치를 기록했다.

ROC-AUC값은 두 모델 모두 충분히 신뢰성 있는 값을 기록했지만, 분류모델 평가 지표에서는 좋지 못한 성능을 기록했다. Logistic Regression 모델의 신뢰도는 0.214, 정확도는 0.736이었고, XGBoost Classifier의 경우는 신뢰도가 0.266, 정확도가 0.547로 매우 낮은 값을 기록했다. 이러한 성능을 향상시키기 위해, XGBoost Regression 모델로 우선 회수 예측을 한 후, 그 결과를 다시 이용하여 분류모델에 이용하는 스테

킹 기법을 사용하였고, 그 결과는 그림 5에 나타났다.

이때, Logistic Regression의 경우는 신뢰도가 0.214에서 0.283으로 소폭 상승했고, 정확도 역시 0.736에서 0.812로 상승한 결과를 보였다. XGBoost Classifier의 경우는 신뢰도가 0.266에서 0.258로 오히려 소폭 하락한 결과를 보였고, 정확도는 0.547에서 0.575로 상승했다. 그러나, 두 모델 모두 여전히 신뢰도가 낮았기 때문에, 예측모델에 이용하기에는 적절하지 않다고 판단했다.

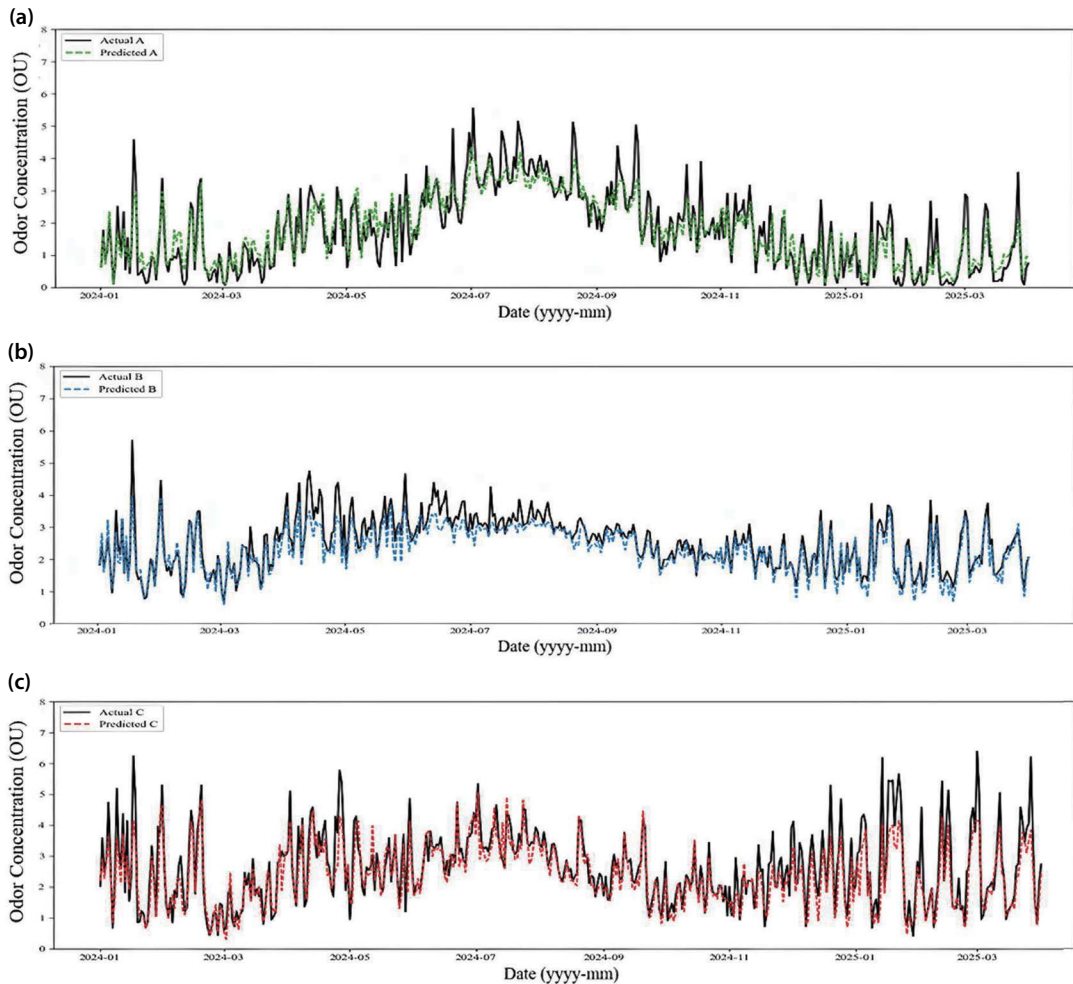


Fig. 3. Daily measured and predicted odorous compound concentrations at three sites using XGBoost ((a) Site #1, (b) Site #2, and (c) Site #3).

민원 예측의 경우, 기상 요인과 같이 단순히 수치적으로 표현되는 값뿐만 아니라, 주변에 거주하는 주민들의 심리적 요인과 같은 특별한 변수들이 작용할 것으로 판단된다. 이러한 심리적 변수들을 잘 반영하기 위하여 기간을 8일 예측 단위로 변경해 예측을 진행하였다. 예측 기간을 8일로 축소하여 민원 발생을 예측한 결과, XGBoost Classifier의 경우가 Logistic Regression 모델보다 더 나은 정확도와 신뢰도를 보였다. 상대적으로 민원 발생 횟수가 많았던 계절인 여름에 랜덤으로 두 기간을 선정해 신뢰도와 정확도를 평가했

고, 그 결과를 그림 6에 나타냈다.

두 기간의 ROC-AUC 값은 각각 0.867, 0.875로 매우 높은 값을 보였다. 첫 번째 기간의 경우, 신뢰도는 75%로, 1일을 제외하고 모두 발생한 경우를 올바르게 예측했고, 정확도는 100%로, 발생한 경우에 대해 모두 발생했다고 적절하게 예측했음을 알 수 있다. 두 번째 기간의 경우는 신뢰도가 100%로 발생했다고 예측한 경우 모두 민원이 발생했고, 정확도는 75%로 발생한 경우에 대해 1일을 제외하고 모두 올바르게 예측했다는 것을 알 수 있다.

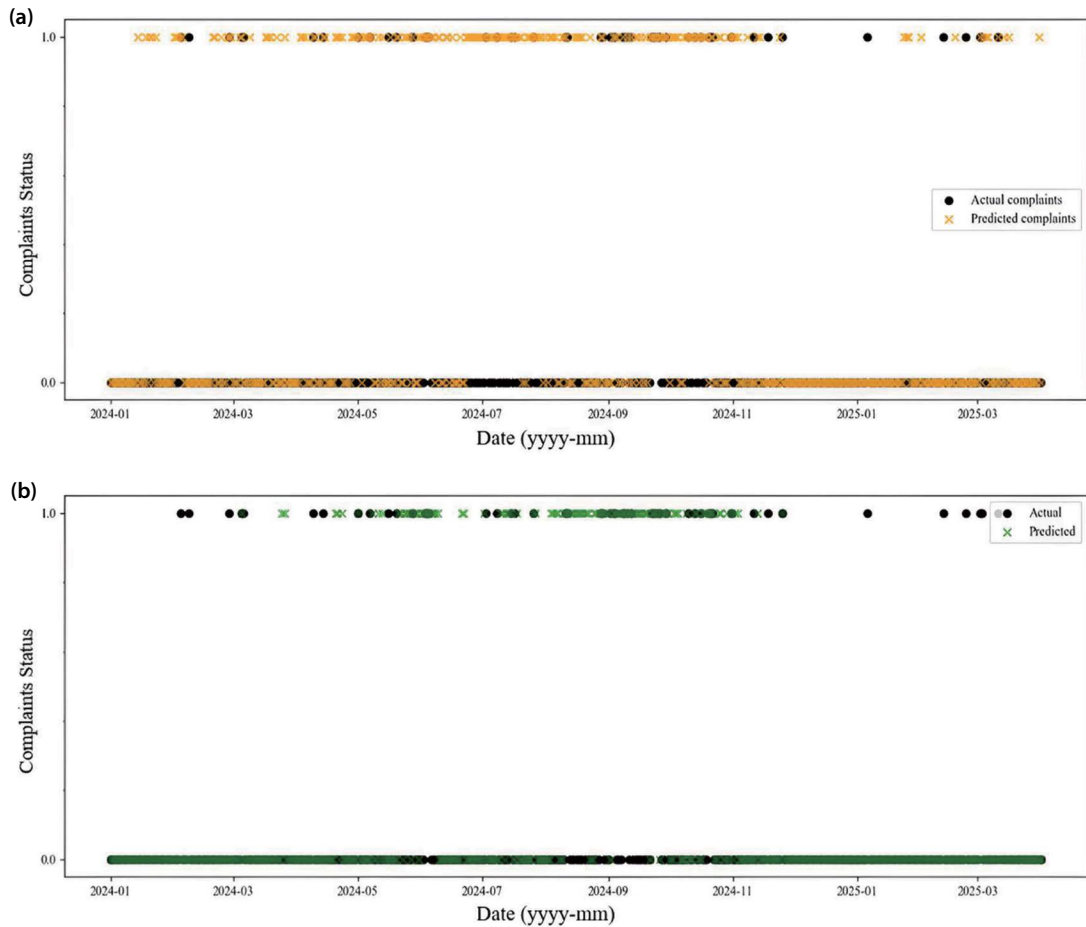


Fig. 4. Comparison in observed and predicted complaints occurrences and those predicted from odorous compound concentrations ((a) XGBoost Classifier and (b) Logistic Regression).

4. 결 론

본 연구는 다양한 머신러닝 모델을 이용하여 제강 사업소에서 발생하는 복합 악취물질의 농도를 예측하고, 예측한 결과를 이용해 대규모 주거지에서 발생하는 악취에 대한 민원을 예측하는 모델을 구축하고자 하였다. 그 결과, XGBoost Regression 모델이 가장 효과적으로 악취물질의 농도를 예측했다. 분류 모델의 경우, 1년 단위 예측에서 신뢰도와 정확도가 매우 낮은 수치를 보였다. 이에 대해, 성능 향상을 기대하며 XGBoost Regression 모델과 XGBoost Classifier

모델을 스택킹 기법을 이용해 모델을 구축했다. 그러나, 성능에 대해 눈에 띄는 상승은 없었다. 민원 데이터의 특성인 심리적 요인을 반영하고자, 예측 기간을 8일로 줄여 진행하였고, 무작위로 선택한 두 기간에 대해 모두 75% 이상의 신뢰도와 정확도를 기록했다.

현재까지 본 주제에 대한 다양한 추가적인 연구가 요구된다. 특히, 수치로 표시할 수 없는 여러 변인들, 즉 심리적 영향, 거주자들의 이주 등의 변인들을 더 효과적으로 모델에 학습시킬 수 있어야 한다. 또한, 주어진 자료가 복합 악취 물질이었기 때문에, 주요 물질

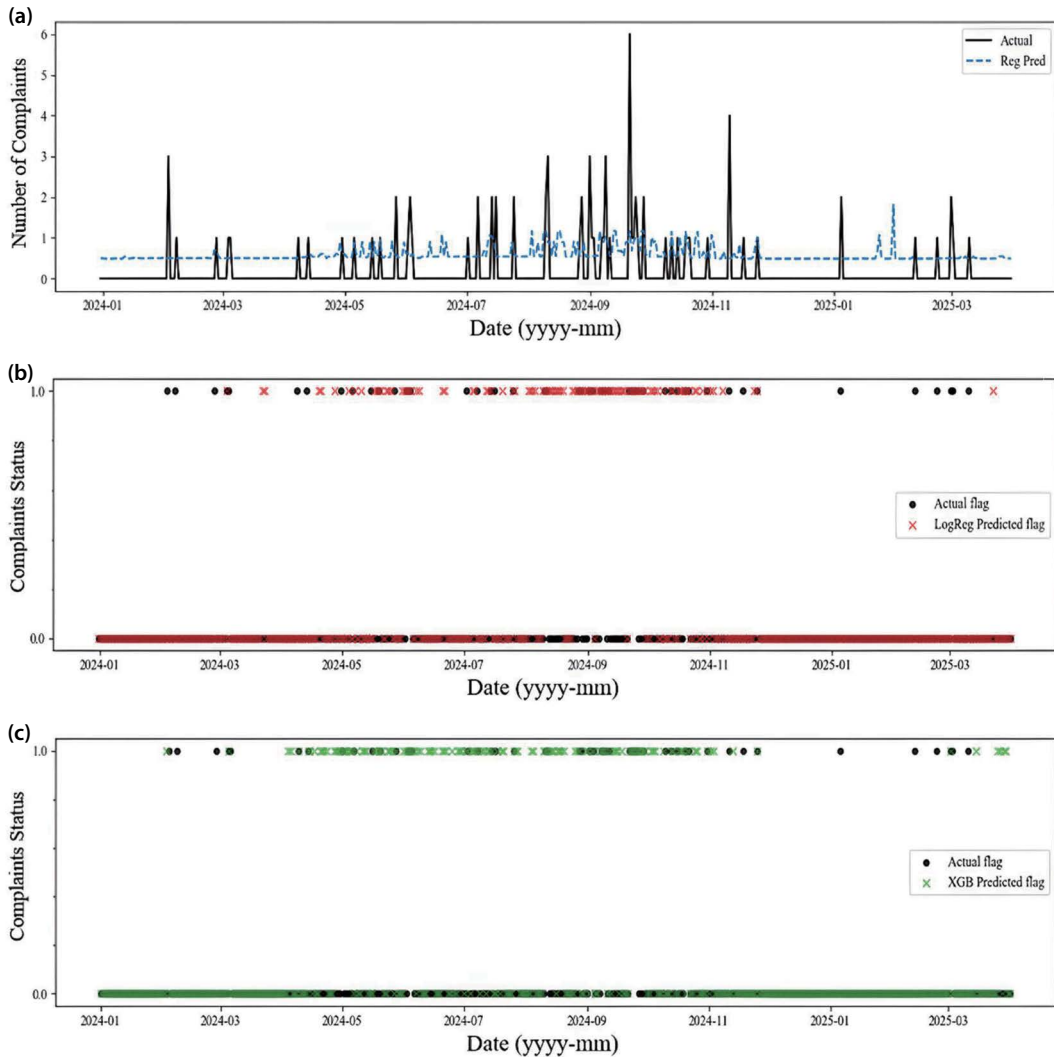


Fig. 5. Comparison in observed and predicted complaints occurrences using stacking ensemble ((a) complaints occurrence predicted from the XGBoost regression, (b) stacking result from XGBoost and logistic regression, and (c) stacking result from XGBoost and XGBoost classifier).

에 대한 개별의 농도 영향을 예측하는 데 한계점이 있었고, 최종적으로 지형적 특성을 모델에 학습시키지 못한 한계점이 있었다. 더욱이 이번에 사용된 데이터는 1950개가량의 가공된 데이터였고, 이에 따라 신경망을 이용하거나 더 많은 모델을 사용하기에는 한계가 있었다. 그럼에도 불구하고, 이번 연구는 적은 데이터로도 복합 악취물질에 대한 농도 영향을 효과적

로 예측할 뿐만 아니라, 이로 인해 인근 주거지에서 발생할 수 있는 민원 역시도 효과적으로 예측할 수 있었다. 향후 추가적인 양질의 데이터를 더 확보한 후 머신러닝뿐만 아니라 다양한 신경망 모델들도 사용하여 평가하고, 더 많은 모델을 사용해 스택킹하여 정확도 및 신뢰도를 더욱 높이는 연구가 필요하다. 이러한 기술이 성공적으로 개발된다면 향후 사업장에서 배출

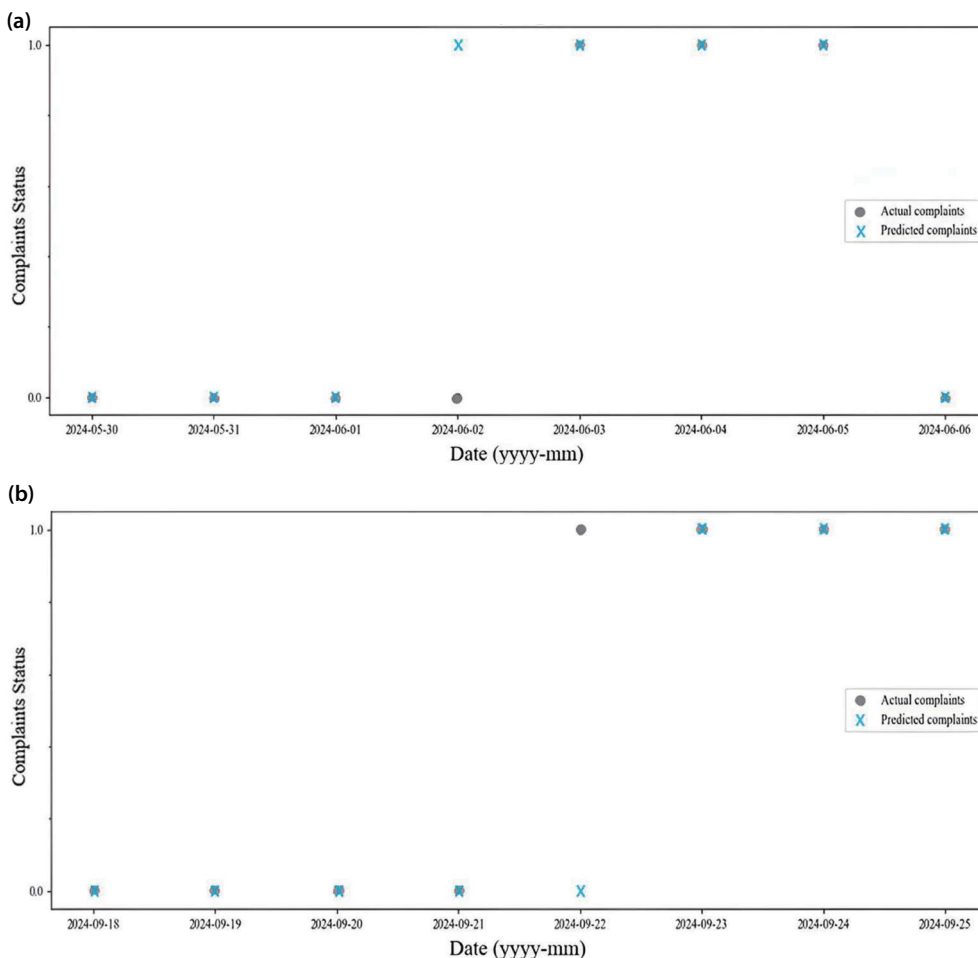


Fig. 6. Comparison between observed and predicted complaints occurrences over eight days using the XGBoost classifier model ((a) period from May 30 to June 6 and (b) period from September 18 to 25).

하는 악취물질의 농도를 효율적으로 관리할 수 있고, 주변 주민들에게 악취 발생에 대한 정보를 전송하여 미리 상황을 알림으로서 인근 주민에게 피해를 최소화할 수 있을 것이다.

감사의 글

이 논문은 국립부경대학교 자율창의기술연구비 (2025년)에 의하여 연구되었음.

References

- Alfasanah, Z., Niam, M.Z.H., Wardiani, S., Ahsan, M., Lee, M.H. (2025) Monitoring air quality index with EWMA and individual charts using XGBoost and SVR residuals, *MethodsX*, 14, 103107. <https://doi.org/10.1016/J.MEX.2024.103107>
- Choi, Y., Kim, K., Kim, S., Kim, D. (2022) Identification of odor emission sources in urban areas using machine learning-based classification models, *Atmospheric Environment: X*, 13, 100156. <https://doi.org/10.1016/J.AEAOA.2022.100156>
- Fehnel, J.W. (1942) Health Hazards in Steel Mills, *American Indus-*

- trial Hygiene Association Quarterly 3, 358-360. <https://doi.org/10.1080/00968204209344046>
- Kang, J.H., Song, J.H., Yoo, S.S., Lee, B.J., Ji, H.W. (2020) Prediction of Odor Concentration Emitted from Wastewater Treatment Plant Using an Artificial Neural Network (ANN), *Atmosphere*, 11(8), 784. <https://doi.org/10.3390/ATMOS11080784>
- Kim, C., You, S.C., Reys, J.M., Cheong, J.Y., Park, R.W. (2021) Machine-learning model to predict the cause of death using a stacking ensemble method for observational data, *Journal of the American Medical Informatics Association* 28, 1098-1107. <https://doi.org/10.1093/JAMIA/OCAA277>
- Korea Meteorological Administration (KMA) (2025) Automatic Weather Station. <https://data.kma.go.kr/data/grnd/selectAwsRltmList.do>
- LaValley, M.P. (2008) Logistic regression, *Circulation*, 117, 2395-2399. <https://doi.org/10.1161/CIRCULATIONAHA.106.682658>
- Lee, H.D., Jeon, S.B., Choi, W.J., Lee, S.S., Lee, M.H., Oh, K.J. (2013) A novel assessment of odor sources using instrumental analysis combined with resident monitoring records for an industrial area in Korea, *Atmospheric Environment*, 74, 277-290. <https://doi.org/10.1016/J.ATMOSENV.2013.04.001>
- Li, Y., Wu, H., Liu, H. (2018) Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction, *Energy Conversion and Management*, 167, 203-219. <https://doi.org/10.1016/J.ENCONMAN.2018.04.082>
- Liang, W., Luo, S., Zhao, G., Wu, H. (2020) Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms, *Mathematics*, 8(5), 765. <https://doi.org/10.3390/MATH8050765>
- Liu, H., Mi, X.-w., Li, Y.-f. (2018) Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network, *Energy Conversion and Management*, 156, 498-514. <https://doi.org/10.1016/J.ENCONMAN.2017.11.053>
- Nam, K.J., Hwangbo, S., Yoo, C.K. (2020) A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea, *Renewable and Sustainable Energy Reviews*, 122. <https://doi.org/10.1016/j.rser.2020.109725>
- Nordin, S., Lidén, E. (2006) Environmental odor annoyance from air pollution from steel industry and bio-fuel processing, *Journal of Environmental Psychology*, 26, 141-145. <https://doi.org/10.1016/J.JENVP.2006.05.002>
- Paeng, J.I., Cho, S.J., Kim, H.M. (2006) Odor Characteristics of Non-ferrous Metal Manufacture Factory at Sihwa · Banwol Complex, *Journal of Korea Society of Environmental Administration*, 12, 171-179.
- Pavlyshenko, B. (2018) Using Stacking Approaches for Machine Learning Models, *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP, 2018*, 255-258. <https://doi.org/10.1109/DSMP.2018.8478522>
- Qu, C., Zhang, Z., Liu, J., Zhao, P., Jing, B., Li, W., Wu, C., Liu, J. (2025) Multi-scenario adaptive electronic nose for the detection of environmental odor pollutants, *Journal of Hazardous Materials*, 489, 137660. <https://doi.org/10.1016/J.JHAZMAT.2025.137660>
- Rincón, C.A., De Guardia, A., Couvert, A., Wolbert, D., Le Roux, S., Soutrel, I., Nunes, G. (2019) Odor concentration (OC) prediction based on odor activity values (OAVs) during composting of solid wastes and digestates, *Atmospheric Environment*, 201, 1-12. <https://doi.org/10.1016/J.ATMOSENV.2018.12.030>
- Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S.V. (2021) Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt, *Scientific Reports*, 11, 1-15. <https://doi.org/10.1038/s41598-020-80820-1>
- Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., Khovanova, N. (2019) Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation, *Biomed Signal Process Control*, 52, 456-462. <https://doi.org/10.1016/J.BSPC.2017.01.012>
- Toivonen, H.T., Tötterman, S., Åkesson, B. (2007) Identification of state-dependent parameter models with support vector regression, *International Journal of Control*, 80, 1454-1470. <https://doi.org/10.1080/00207170701378673>
- Xiao, H., Tian, J., Chen, Y., Wang, C., Zhang, Y., Chen, L. (2024) Uncovering the features of industrial odors-derived environmental complaints and proactive countermeasures by using machine-learning, *Journal of Environmental Management*, 370, 122900. <https://doi.org/10.1016/J.JENVMAN.2024.122900>
- Xu, A., Li, R., Chang, H., Xu, Y., Li, X., Lin, G., Zhao, Y. (2022) Artificial neural network (ANN) modeling for the prediction of odor emission rates from landfill working surface, *Waste Management*, 138, 158-171. <https://doi.org/10.1016/J.WASMAN.2021.11.045>
- Zhang, F., O'Donnell, L.J. (2020) Support vector regression, *Machine Learning: Methods and Applications to Brain Disorders*, 123-140. <https://doi.org/10.1016/B978-0-12-815739-8.00007-9>

Authors Information

김동한(국립부경대학교 지구환경시스템과학부 학사과정)
(kdh381643@gmail.com)
오용환(국립부경대학교 지구환경시스템과학부 박사과정)
(gloryforest@naver.com)

서서희(국립부경대학교 환경연구소 연구원)
(seoseoheee@naver.com)
천소현(국립부경대학교 지구환경시스템과학부 학사과정)
(sohyeoncheon75@gmail.com)
손윤석(국립부경대학교 지구환경시스템과학부 교수)
(sonys@pknu.ac.kr)