

# 군집화를 통해 추출한 AI 학습용 분광 스펙트럼 데이터 기반 유해화학물질 판독 모델

유성민\*, 김연진\*, 조숙경\*\*, 백성하\*\*\*, 김경배°

## AI-Based Model for Detection of Hazardous Chemicals Using Spectral Data Extracted Through Clustering

Seong-Min Ryoo\*, Yeon-Jin Kim\*, Sook-Kyung Cho\*\*, Sung-Ha Baek\*\*\*, Gyeong-Bae Kim°

### 요약

유해화학물질 사고는 소량이라도 주변에 큰 영향을 미칠 수 있기에 사고 발생 시 해당 화학물질을 신속하게 식별하고 적절히 대응하는 것이 매우 중요하다. 기존의 화학물질판독 연구는 화학물질 센서 및 영상정보를 활용하였으나, 화학물질 센서를 기반으로 한 방법은 센서가 없는 경우 적용이 어려우며, 영상정보를 기반으로 한 방법은 물질의 색상이 동일하거나 무색인 경우에 탐지가 어려운 문제점이 있다. 따라서 본 논문에서는 동일한 색상이나 무색의 유해화학물질을 구분할 수 있도록 분광 스펙트럼 데이터를 이용한 새로운 AI 기반 유해화학물질 판독 시스템의 인공지능학습데이터 구축 기법 및 유해화학물질 판독 모델을 제안한다. 제안된 인공지능학습데이터 구축 기법은 원시 분광 스펙트럼 데이터에 군집화를 통한 데이터 추출 방법을 적용하여 데이터 추출의 정확도를 향상시켰다. 또한, 추출한 데이터를 기반으로 물질의 분광 스펙트럼 데이터의 특성을 반영해서 랜덤 포레스트 알고리즘을 활용해 구현한 유해화학물질 판독 모델을 제안하였고, 자체 실험을 통해 모델의 성능을 검증했다. 본 논문에서 제안된 인공지능기반 유해화학물질 판독 시스템은 시각적 특징이 없는 유해화학물질 사고 발생 시에도 신속한 판독이 가능함으로써 선제적 대응을 통해 화학 사고의 피해를 줄일 수 있을 것으로 기대된다.

**키워드** : 유해화학물질 판독, AI, 초분광, 재난재해, AI 학습데이터

**Key Words** : Detection of hazardous chemicals, artificial intelligence, hyperspectral imaging, disaster and hazard, artificial intelligence training data

### ABSTRACT

Hazardous chemical incidents can have a significant impact on the surroundings even with small quantities involved. Therefore, it is crucial to swiftly identify the chemical substance involved in an incident and respond appropriately when such incidents occur. Conventional studies on chemical substance identification have utilized chemical sensors and visual information. However, methods based on chemical sensors are challenging to apply when sensors are unavailable, and approaches relying on visual information face difficulties in detecting

※ 본 연구는 과학기술정보통신부와 정보통신산업진흥원 주관으로 소망청 컨소시엄에서 수행하는 “AI융합 유해화학물질 판독 시스템 구축사업” (2022-2024)의 지원을 받았다

• First Author : Woogyung IT, smin9830@gmail.com, 정회원

° Corresponding Author : Seowon University, Dept. of Software, gbkim@seowon.ac.kr, 종신회원

\* Seowon University, anne6497@naver.com, 정회원

\*\* Nestors Co.,Ltd., skyindb@naver.com, 종신회원

\*\*\* Inha Technical College, bshzeratul@gmail.com

논문번호 : 202403-049-0-SE, Received March 8, 2024; Revised April 11, 2024; Accepted April 13, 2024

substances with identical or colorless characteristics. Therefore, this paper proposes a novel artificial intelligence-based hazardous chemical detection system using spectral spectrum data to address the challenge of distinguishing hazardous chemicals with the same color or those that are colorless. The paper introduces a technique for constructing artificial intelligence training data and a hazardous chemical detection model based on spectral spectrum data. The proposed artificial intelligence training data construction method improved the accuracy of data extraction by applying clustering to raw spectral spectrum data for extraction. Additionally, based on the extracted data and reflecting the characteristics of material spectral spectrum data, we proposed a hazardous chemical detection model implemented using the random forest algorithm. We validated the performance of the model through in-house experiments. The AI-based hazardous chemical detection system proposed in this paper is expected to minimize the damage from chemical incidents through rapid identification, even in cases where hazardous chemicals lack visual characteristics, enabling proactive response measures.

## 1. 서 론

유해화학물질(hazardous chemicals)은 유독물, 관찰물질, 취급제한물질 또는 취급금지물질, 사고대비물질, 그 밖에 유해성 또는 위해성이 있거나 그러할 우려가 있는 화학물질로 다양한 산업 및 일상생활에서 사용된다. 최근 전세계적으로 유해화학물질의 유통량이 크게 늘어남에 따라 유해화학물질 관련 사고가 꾸준히 발생하고 있으며, 사고로 인한 인명 피해도 늘어나고 있다. 2022년도 국내 화학사고 발생 건수는 총 218건이며, 인명 피해는 241명(사망 13명, 부상 228명)으로 전년 대비 발생 건수는 감소하였지만, 인명 피해는 늘어나고 있다<sup>[1]</sup>. 유해화학물질 사고는 비교적 적은 양이라도 화학물질이 지닌 인화성, 산화성, 폭발성 및 독성과 같은 특성으로 인해 주변 환경, 주민, 재산에 큰 영향을 줄 수 있다. 따라서 사고 발생 시에 사고가 발생한 유해화학물질의 종류를 신속하게 판단하고 적절하게 대응하는 것이 매우 중요하다.

유해화학물질판독에 관한 기존연구는 화학물질 센서를 기반으로 한 방법과 영상정보를 기반으로 한 방법으로 진행되었다. 화학센서를 이용한 방법으로는 황화수소 센서, 일산화탄소 센서 및 이산화탄소 센서 등을 이용한 가스검출 연구<sup>[2]</sup>, VOCs 센서를 이용해서 VOCs 화학종 가스검출 연구<sup>[3]</sup>, 인듐-주석-산화물의 산화 환원 반응에 기반한 유해화학물질 검출 연구<sup>[4]</sup>, 그리고 식품 내의 화학물질 검출을 위한 바이오센서 개발 연구<sup>[5]</sup> 및 누액감지기와 가스 검출기 등을 사용한 방재 시스템 연구<sup>[6]</sup>가 수행되었다. 그러나, 기존의 화학물질 판독방식은 센서가 없는 경우 적용할 수 없다는 문제점을 지니고 있어 국내 화학물질을 취급하는 3만여 개의 사업장의 사고에 적용하는 데에 한계가 있다. 영상정보를 기반으

로 한 방법은 화학물질 영상의 IR 데이터와 RGB 데이터를 이용해 화재감지 모델을 학습시켜 화재를 감지하는 연구<sup>[7]</sup>와 이미지 기반의 화재감지 방법 Local Binary Pattern과 SVM(Supported Vector Machine)을 이용해 감지성능을 높이는 기법<sup>[8]</sup> 등이 연구되었다.

최근에는 유해화학물질의 영상 및 이미지를 인공지능기술에 적용한 연구가 진행되고 있다<sup>[9]</sup>. 인공지능기반의 판독 시스템은 영상 데이터를 통해 수집한 유해화학 물질에 대한 정보를 학습하여 머신러닝 기술을 적용한 인공지능시스템을 구축하고, 이를 이용하여 화학물질 사고가 발생한 물질에 대한 판독하는 시스템이다. 그러나, 영상정보에 기반한 기법은 유해화학물질이 동일 및 유사한 색상이거나 무색인 경우에 물질 탐색 및 구별이 어려운 문제점이 있다.

따라서 본 논문에서는 초분광카메라의 분광스펙트럼 데이터를 이용하여 동일 및 유사 색상이나 무색의 사고 유해화학물질을 구분할 수 있는 인공지능기반의 유해화학물질 판독 시스템을 제안한다. 초분광 카메라는 적외선을 이용하여 물질을 촬영하며, 물질에 따라 적외선의 흡수와 반사가 다르게 나타나는 특성을 이용하는 카메라이다. 초분광 카메라로 물질 촬영 시 분광스펙트럼 데이터가 출력되고, 이 데이터를 이용해서 물질을 구분하는 것이 가능하다. 초분광 카메라는 미술<sup>[10,11]</sup>, 농업<sup>[12]</sup>분야뿐만 아니라 환경분야의 수질오염 분석<sup>[13]</sup> 및 하천수 오염 연구<sup>[14,15]</sup>, 해양누출 사고 연구<sup>[16,17]</sup> 등에서 쓰이고 있다. 분광스펙트럼의 데이터를 인공지능기반의 유해화학물질 판독 시스템에 적용하기 위해서는 유해화학물질의 분광스펙트럼 데이터를 인공지능 학습용 데이터로 추출하기 위한 기법과 이를 이용한 인공지능기반의 유해화학물질 판독 모델이 필요하다. 본 논문에서는 데이터 구축 방안으로 군집화를 통한

데이터 추출 방법을 제시한다. 분광 스펙트럼 데이터는 촬영 시 물질이 아닌 데이터도 같이 촬영되어 물질의 데이터만을 추출하는 것에 어려움이 있다. 군집화를 사용하면 데이터의 특징에 따라서 영역이 분할되어 물질의 특징을 가진 데이터만을 추출할 수 있다.

또한, 유해화학물질 판독을 위한 모델로는 군집화를 통해 추출한 분광 스펙트럼 데이터에서 얻을 수 있는 특징을 기반으로 유해화학물질 판독을 위한 모델을 학습시켰다. 동일 및 유사 색상이나 무색의 유해화학물질을 구분할 수 있도록 제안된 모델은 랜덤 포레스트 알고리즘을 이용해 구현되었고, 분광 스펙트럼의 밴드별 반사율 데이터로 학습시킨 모델에 비해 판독 소요 시간이 더 적게 걸리는 장점이 있다.

본 논문의 구성은 다음과 같다. 2장에서 관련 연구에 관해 기술하고, 3장에서는 유해화학물질 판독 시스템에 관해 설명한다. 4장에서는 군집화를 통한 인공지능학습용 분광 스펙트럼 데이터 추출, 5장에서는 분광 스펙트럼 데이터를 통해 학습시킨 인공지능모델을 제시한다. 마지막으로 6장에서는 결론을 맺고 향후 연구를 제시한다.

## II. 관련 연구

초분광 카메라는 미술 분야에서 유화에 사용된 물감, 혹은 단청에 사용되는 안료를 분석하기 위해 사용되고 있다<sup>10,11</sup>. 단청 안료 분석 연구의 경우 한국전통 목조건물의 색소 분석에 초분광기법을 적용하여 단청의 제작시기와 복원시기를 추정하였다. 이를 위해 ENVI 프로그램을 사용해서 분석하고자 하는 안료로 칠해진 부분 중에 한 픽셀만을 선택해서 스펙트럼 데이터를 분석했으며, 유화 물감 분석의 경우는 물감이 존재하는 픽셀 중 5, 10 픽셀 분량의 데이터를 추출해 평균값을 사용했다. 해당 연구들의 데이터 추출 방법은 비교 분석을 위한 소량의 데이터만을 추출하기 때문에 인공지능학습에는 적절하지 않은 방법이다.

하천수 오염 연구의<sup>14,15</sup> 경우 초분광영상을 취득할 수 있는 드론을 이용하여 넓은 지역을 잦은 빈도로 조사하여 하천의 분광영상을 취득하고, 최적 밴드비 탐색 알고리즘을 적용하여 분석하여 하천의 변화를 추적하였다. 그러나 해당 연구는 분광 라이브러리 구축에 초점이 맞춰져 있고, 화학물질 사고 중 수계로 유출되는 유해화학물질 식별이 목적이라는 부분에서 차이가 존재한다. 데이터 수집 방법 또한 빛의 영향이 없는 부분에서 10\*10픽셀의 영역을 지정해서 추출하는 방법을 사용했다.

## III. 인공지능기반의 유해화학물질 판독 시스템

유해화학물질 판독 시스템은 화학사고 발생 시 현장에 출동하는 소방관들을 위해서 사고현장 도착 전에 사고발생 화학물질에 대한 정보를 신속하고 정확하게 제공하여 현장에 출동한 소방관들이 화학 사고에 적절하게 대응하여 피해를 줄일 수 있도록 도와주는 시스템이다.

본 논문에서 제안하는 인공지능기반의 유해화학물질 판독 시스템<sup>18</sup>의 구성은 그림 1과 같다. 제안 시스템의 전체적인 구성은 스마트폰, 카메라 등을 이용해서 유해화학물질 사고현장에서 영상을 전송하는 부분, 전송된 사고 영상을 받아 현장의 상황을 판단하고 대응하기 위한 119 종합 상황실 시스템(소방청 상황판), 그리고 사고현장의 영상 및 관련 정보를 통해서 해당 물질이 어떤 물질인지 판독하는 인공지능기반의 유해화학물질 판독 시스템으로 구성되어 있다. 유해화학물질 사고현장에서 소방청의 119 종합 상황실로 전송된 영상 및 정보는 사고 발생 화학물질의 종류를 판독하기 위해서 유해화학물질 판독 시스템으로 전송되고 인공지능기반의 유해화학물질 판독 시스템은 영상을 위한 CNN 기반의 이미지 분류 방법과 초분광 카메라를 통해 얻은 분광 스펙트럼 데이터를 기반으로 한 분류 기법을 적용하여 정확한 유해화학물질에 대한 판독을 수행하도록 설계 및 개발되었다.

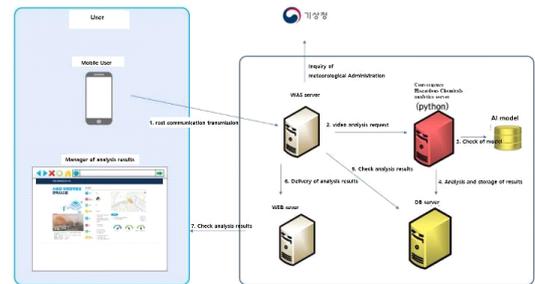


그림 1. 유해화학물질 판독 시스템<sup>18</sup>  
Fig. 1. Hazardous Chemical Detection System<sup>18</sup>

## IV. 군집화를 통한 분광 스펙트럼 데이터 추출

인공지능모델은 학습에 사용하는 데이터가 많을수록 판독의 정확성 및 성능이 향상되므로 대량의 학습데이터 구축이 필요하다. 분광 스펙트럼 데이터를 추출하기 위해서 촬영된 원시 초분광 카메라데이터는 분광 스펙트럼 데이터의 각 파장대별 반사율이나 흡수율 등의

정보를 포함하고 있다. 촬영된 분광데이터는 많은 오류 및 노이즈를 포함하고 있어 데이터의 수집 및 전처리과정에서 데이터의 품질을 높이고 모델 학습에 적합한 형태로 변환하기 위한 스펙트럼 데이터의 추출 기법이 매우 중요하다.

분광 스펙트럼 데이터는 그림 2에 표시된 영역 내의 픽셀마다 존재한다. 학습을 위한 화학물질은 유리병 내부에만 존재하므로 인공지능학습을 위한 데이터의 정확도 및 품질을 높이기 위해서는 그림 3의 영역처럼 유리병 내의 물질 데이터만을 추출해야 한다. 또한, 초분광카메라의 특성으로 인해서 유리병의 곡면에 의한 빛의 산란현상으로 인해 노이즈가 포함되는 현상이 발생한다. 유해화학물질의 분광데이터 정확도 및 품질을

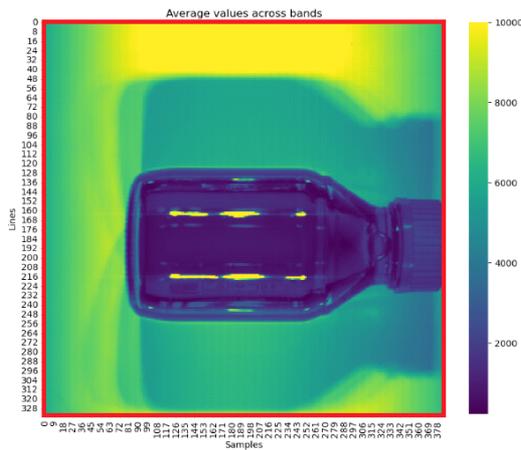


그림 2. 전체 분광 스펙트럼 데이터 영역  
Fig. 2. Whole Spectral Spectrum Data Range

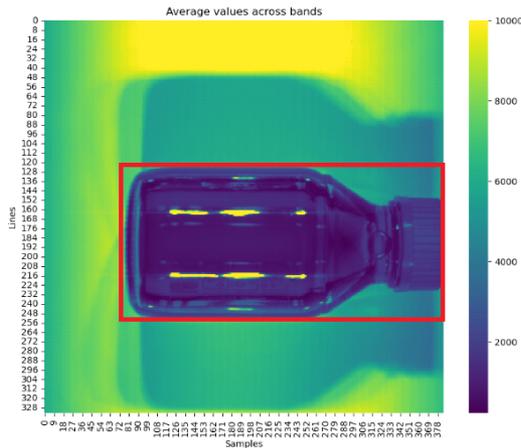


그림 3. 등유 분광 스펙트럼 데이터 영역  
Fig. 3. Kerosene Spectral Spectrum Data Range

높이기 위해서는 정확한 유해화학물질의 영역 추출과 노이즈를 제거하기 위한 기법이 필요하다. 이를 해결하기 위해서 본 장에서 군집화를 통해 정확한 유해화학물질 데이터 영역 추출과 유리병 외부 및 내부의 노이즈 데이터를 제거하는 기법을 적용하였다.

#### 4.1 군집화를 통한 데이터 추출 및 노이즈 제거 기법

분광 스펙트럼 데이터는 측정 장비의 센서 오차, 촬영 길이, 환경적 요인, 측정 과정에서 발생하는 외부 간섭 등으로 인해 노이즈를 제거하기 위해 군집화를 통한 데이터 추출 및 노이즈를 제거한다. 유해화학물질은 물질의 특성으로 인해서 비슷한 분광특성의 값들을 중심으로 군집화를 이루게 된다. 제안된 군집화 기법은 촬영된 분광 데이터에서 주어진 데이터들의 특성을 고려해 데이터의 집단을 정의하고 데이터 집단의 대표할 수 있는 대표점을 찾아 군집화를 구성한다. 군집화를 하기 위해서는 데이터의 어떤 특징을 비교해 군집화를 할 것인지 결정해야 한다. 본 논문에서는 유해화학물질에 따라 분광스펙트럼 데이터가 지니는 특성인 극대점 및 극소점의 수, 평균·최대·최소 신호 값(반사율)의 강도, 1, 2차 도함수의 평균값, 1차 도함수의 평균값을 적용하여 군집화를 결정한다. 그림 4는 스펙트럼 그래프 파형에서 추출하는 8가지 특징의 일부를 표시한 것으로 붉은색으로 표시된 최대 신호 값(반사율)과 극대점 및 극소점의 개수이다. 이 특징들을 각 픽셀별로 추출하게 되고, 서로 비교해서 군집화를 진행하게 된다.

본 논문은 5개의 영역으로 데이터 군집화를 실행했다. 그림 2의 히트맵 이미지와 같은 데이터에 군집화를 적용했고, 그림 5과 같이 5개의 영역으로 특징이 비슷한 데이터들끼리 영역을 이루었다.

그림 3과 그림 5를 비교하면 물질이 존재하는 부분은 노란색 영역과 진한 청록색 영역인 것을 확인할 수 있다. 유리병 곡면에 의한 빛의 산란으로 생긴 노이즈

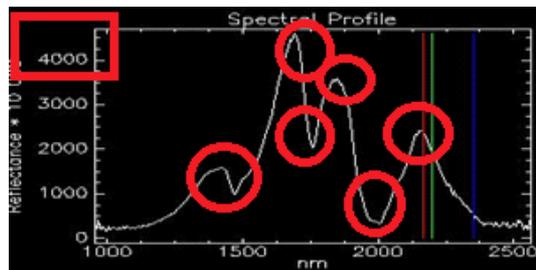


그림 4. 유해화학물질의 분광스펙트럼 그래프 특징 추출  
Fig. 4. Extraction of Spectral Graph Features of Hazardous Chemicals

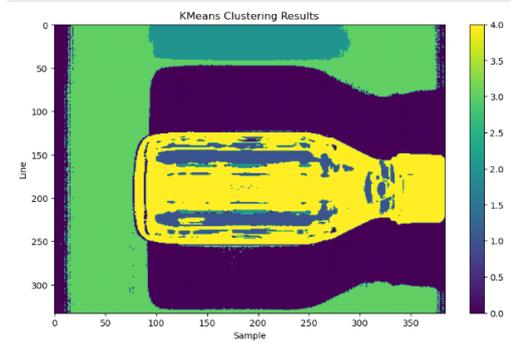


그림 5. 등유의 군집화 히트맵 예  
Fig. 5. Example of a spectral clustering heatmap for kerosene

데이터는 유리병 내부에서 연한 청록색으로 물질 영역과는 다른 영역에 속하고 있다. 물질 영역 2개만을 지정해서 데이터를 추출하게 되면 유리병 내부의 노이즈 데이터는 추출되지 않기에 유해화학물질의 분광 스펙트럼 데이터만을 추출할 수 있어 데이터의 정확도 및 품질이 향상된다.

### V. 분광 스펙트럼 데이터를 통한 인공지능 유해화학물질 판독 모델 및 성능평가

본 장에서는 군집화하여 추출한 분광 스펙트럼 데이터를 이용하여 유해화학물질을 판독하기 위한 유해화학물질의 분광 스펙트럼 데이터의 특성에 기반한 4가지 판독 모델을 제안하고 제안 모델에 대한 성능평가를 수행하였다. 제안된 모델은 논문에서 구현한 인공지능 기반 유해화학물질 판독 시스템의 머신러닝 알고리즘(랜덤 포레스트)에 적용하여 판독 모델의 정확성을 실험하였다. 랜덤 포레스트 알고리즘은 검출, 분류, 회귀 분석 등에 사용되는 학습 방법으로 신속하고 정확하게 모델을 훈련시키며, 대량의 데이터를 처리를 지원한다는 장점이 있어서 선택되었다. 판독 모델은 특징의 개수를 달리해 4개의 모델을 학습시킨다. 해당 4가지 모델들은 학습에 동일한 분광 스펙트럼 데이터 셋을 사용했으며, 황산(sulfuric acid), 암모니아(ammonia), 등유(kerosene), 플루오라이드칼륨 수용액(potassium fluoride solution), 톨루엔(toluene), 메탄올(methanol)의 6가지 무색 물질의 분광 스펙트럼 데이터를 사용해 학습을 진행했다. 훈련 데이터 셋과 검증 데이터 셋은 7 : 3의 비율로 학습을 진행하였다. 실험에 사용된 PC 환경은 표 1과 같다.

표 1. 실험 PC 환경  
Table 1. Experimental PC Environment

CPU	Intel Core i7 12700H
GPU	NVIDIA GeForce RTX 3050TI Laptop GPU
RAM	DDR5 32GB

#### 5.1 유해화학물질의 분광 스펙트럼 데이터 특성에 기반 한 판독 모델

본 연구에서는 유해화학물질 분류를 위해 유해화학물질의 분광 스펙트럼 데이터 특성을 기반으로 4가지 유형의 발된 판독 모델을 개발하였다. 개발된 모델은 분광 스펙트럼 데이터의 극대점 및 극소점 개수, 100단 위 범위별 반사율의 평균 강도, 전체 반사율의 평균과 표준편차, 그리고 1차 및 2차 도함수의 평균값, 1차 도함수의 절댓값 합 등을 적용한다. 이러한 특성들은 스펙트럼 파형 데이터에서 추출하여 극점의 개수, 반사율 관련 특성, 그리고 기울기 관련 도함수 특성으로 분류하여 판독에 사용한다. 모델 1은 극점의 개수를 기본 특성으로 선정하였고, 모델 2는 여기에 반사율 관련 세 가지 특성을 추가하여 총 다섯 가지 특성을 사용하였다. 모델 3은 기울기 관련 특성 세 가지를 더해 총 여덟 가지 특성을 사용하였다. 이 모델들은 특성의 개수와 유형에

표 2. 모델별 학습 특징 요약표  
Table 2. Summary Table of Learning Characteristics by Model

	model1	model2	model3	model4
local maxima	O	O	O	X
local minima	O	O	O	X
spectral reflectance	X	O	O	X
total reflectance	X	O	O	X
standard deviation	X	O	O	X
first derivative	X	X	O	X
second derivative	X	X	O	X
absolute value of the first derivative	X	X	O	X
band feature	X	X	X	O

다른 성능 비교를 위해 선택되었으며, 마지막으로 모델 4는 특성 추출 없이 원천 데이터를 직접 사용하여 학습된 모델로, 스펙트럼 데이터의 Band별 반사율 수치를 직접 학습 데이터로 사용하였다.

위의 4개의 모델을 비교해서 유해화학물질 판독 시스템에 적합한 모델을 찾기 위해 성능평가 및 분류실험을 진행한다. 모델들의 학습종료 후 검증 데이터 셋을 통한 모델들의 성능평가를 진행한다. 성능평가 지표는 검증정확도(validation accuracy), 물질별 정밀도(precision), 재현율(recall), f1-score로 한다.

첫 번째 실험은 대량의 데이터에 대한 판독처리 성능을 비교하기 위해 데이터의 양의 변화에 따른 실험을 진행하였다. 화학사고 발생 시 현장에서 촬영한 데이터는 실험실 환경에서 촬영한 데이터와는 다르게 어느 부분에 물질이 존재하는지 알 수 없기에 입력된 데이터에서 모든 픽셀의 데이터를 확인해야 한다. 판독 모델은 신속하게 사고가 발생한 유해화학물질을 판독해서 물질 관련 정보를 제공해야 하므로 대량의 데이터를 신속하고 정확하게 분류해야 한다.

두 번째 실험은 실험실 환경이 아닌 외부환경에서의 사고 시 판독율을 확인하기 위한 노이즈 데이터의 분류 실험을 진행한다. 실제 화학사고 발생 시 현장의 데이터는 기상 상황, 일광 등에 의해서 영향을 받아 노이즈가 많이 삽입된다. 노이즈가 섞인 데이터들을 추출해 실험을 진행함으로써 사고현장에서의 성능에 대해 유해화학물질 판독의 정확성을 평가한다.

5.2 테스트 데이터 셋에 대한 물질별 정확도 비교

모델들의 성능을 측정하기 위해 각 모델의 검증정확도, 물질별 정밀도, 재현율, 그리고 f1-score값을 비교한다. 데이터를 분류하는 데에 사용하는 특징이 2개인 모델 1이 검증정확도를 비롯한 물질별 정확도 점수가 가장 낮다. 이와는 반대로 가장 많은 288개의 특징으로 학습한 모델 4의 경우는 가장 정확도가 높다. 검증정확도는 모델별로 각각 0.39342, 0.94053, 0.95133, 0.98231의 수치를 기록했으며, 학습시키는데 사용한 데이터의 특징의 수가 많을수록 검증정확도가 높다.

위의 그림 6, 그림 7, 그리고 그림 8은 모델들의 물질별 정밀도, 물질별 재현율, f1-score이다. 정밀도는 높을수록 모델이 예측한 물질이 정답일 확률이 높다는 것을 의미하며, 재현율은 높을수록 모델의 물질 판독률이 높다는 것을 의미한다. 모델 1의 경우 정밀도, 재현율, f1-score가 다른 모델들보다 낮다. 모델 1은 등유와 메탄올의 지표가 매우 낮는데, 재현율의 경우는 0.1보다 낮다. 모델 1은 등유와 메탄올을 판독하지 못한다. 그

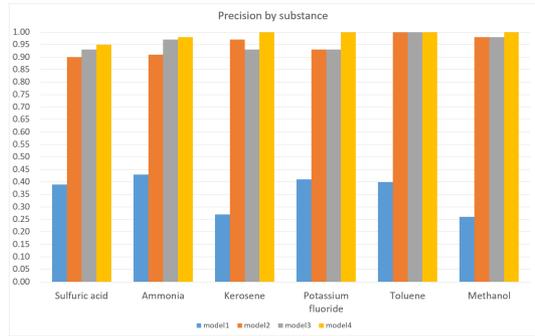


그림 6. 테스트 데이터 셋에 대한 물질별 정밀도 비교  
Fig. 6. Comparison of Material-specific Precision for the Test Dataset

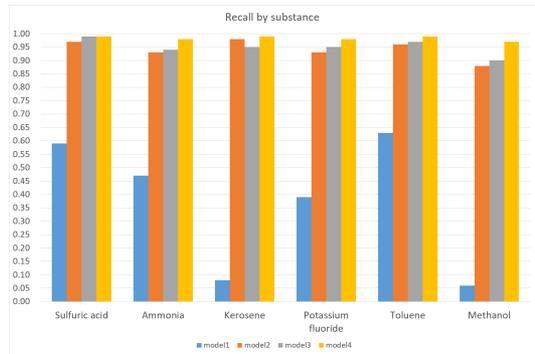


그림 7. 테스트 데이터 셋에 대한 물질별 재현율 비교  
Fig. 7. Comparison of Material-specific Recall for the Test Dataset

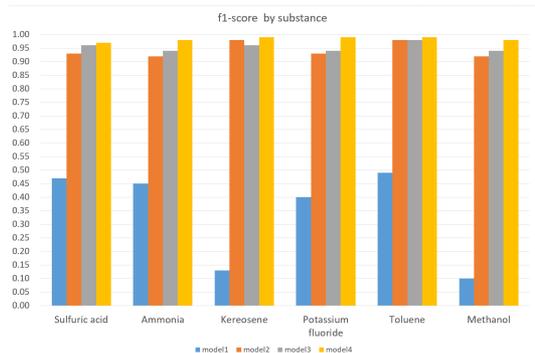


그림 8. 테스트 데이터 셋에 대한 물질별 f1-score 비교  
Fig. 8. Comparison of Material-specific f1-score for the Test Dataset

위의 물질들도 다른 모델들보다 수치가 낮다. 모델 1은 물질 판독 성능이 다른 모델들보다 낮다. 모델 2는 메탄올의 재현율이 0.88로 다른 지표보다 낮으며, 메탄올을 찾아내는 성능이 다른 물질들을 찾아내는 성능보다 부족하다. 모델 1을 제외한 세 모델은 모든 물질에 대하여

3개의 지표가 고루 높은 것을 확인할 수 있으며, 검증 데이터의 6종의 물질을 분류하는 데에 0.9 이상의 성능을 보였다.

### 5.3 데이터 개수에 따른 성능 비교

제안한 인공지능모델은 화학사고 발생 현장에서 전송하는 데이터 내의 모든 분광 스펙트럼 데이터를 확인해서 사고물질을 판독해야하므로 대량의 데이터를 빨리 처리할 수 있어야만 한다. 이번 실험은 데이터의 개수를 변화시키면서 정답률과 분류시간을 측정해 모델별 성능을 비교한다. 데이터는 물질별 100개, 500개, 1,000개, 2,000개로 실험을 진행하며 분광 데이터 1개당 127,872개의 데이터를 가지고 있기에 마지막 분류실험의 데이터 개수는 12만 개 이상으로 한다. 실험에 사용하는 검증 데이터는 같은 날에 촬영한 6종의 물질 데이터로 데이터 개수는 총 125,662개이다. 다섯 번째 실험에서는 이 데이터들을 전부 사용한다. 모델별 정확도와 분류시간은 다음 표와 같다.

실험별 정확도는 표 3과 같고, 실험별 분류시간은 표 4와 같다. 데이터 개수의 변화에 따른 각 모델의 정확도는 모델 4가 가장 높고 모델 1이 가장 낮다. 이유는 학습시킨 4개의 모델은 학습에 사용된 특징의 개수가 많을수록 모델의 정확도가 높아지기 때문이다. 모델 4 다음으로 정확도가 높은 모델은 모델 3인데, 모델 4와 정확도 차이가 모든 실험에서 3%p 이하다. 모델 1의 정확도는 모든 분류실험에서 40% 미만으로 유해화학물질들을 판독하는 정확도가 낮으며, 90% 이상의 정확도를 기록한 모델 2 ~ 4보다 판독 성능이 떨어진다. 분류시간은 모델 1 ~ 3의 경우 모든 실험에서 거의 동일한 시간이 소요되었고, 가장 시간 소요가 적은 모델이 실험 때마다 변경되었다. 반면, 모델 4의 경우는 모든

표 3. 데이터 개수 변화에 따른 모델별 분류실험 정확도  
Table 3. Model-specific Classification Accuracy with Changes in Data Quantity

Unit : %	model1	model2	model3	model4
classification experiment 1	38.5	93.5	95	98
classification experiment 2	37.5	95.57	95.77	98.7
classification experiment 3	38.22	95.52	95.93	98.78
classification experiment 4	37.89	95.02	95.7	98.58
classification experiment 5	34.33	95.54	95.87	98.74

표 4. 데이터의 개수 변화에 따른 모델별 분류실험 분류시간  
Table 4. Model-specific Classification Time in Classification Experiments with Changes in Data Quantity.

Unit : sec	model1	model2	model3	model4
classification experiment 1	1.978	1.959	1.926	4.776
classification experiment 2	9.330	9.341	9.416	21.232
classification experiment 3	19.407	19.219	18.640	43.769
classification experiment 4	37.307	37.635	37.807	84.918
classification experiment 5	391.362	399.706	396.879	451.570

실험에서 판독 시간이 가장 많이 소요되었다.

모델 4와 실험 별 가장 속도가 빠른 모델 간의 시간 차이를 비교해보면 첫 번째 실험에서 2.85초, 두 번째에서 11.902초, 세 번째에서 25.129초, 네 번째에서 47.61초, 다섯 번째 실험에서 54.691초 차이로 317%, 111%, 89.4%, 14.8%의 꾸준한 시간 차이의 증가를 볼 수 있다. 데이터의 양이 많아질수록 모델 4와 다른 모델과의 속도 차이는 계속 증가하기에 다른 모델들에 비해 모델 4의 분류 속도는 부족하다.

모델들의 물질별 정밀도, 재현율, f1-score 결과는 그림 9 ~ 그림 11과 같다. 정밀도, 재현율, f1-score의 경우 모델 2 ~ 4가 모든 물질에 대해서 높은 결과를 보였으며, 이는 모델 2 ~ 4는 학습에 사용한 6개의 물질에 대해서 매우 높은 정확도로 분류하고 있음을 확인할 수 있다. 모델 1의 경우 등유와 메탄올을 이번 실험에서도 거의 분류하지 못하고 있고, 전체적으로 모든 지표에서 낮은 점수를 기록했다. 모델 1은 유해화학물질을 분류함에 있어서 적합하지 않은 모델이다.

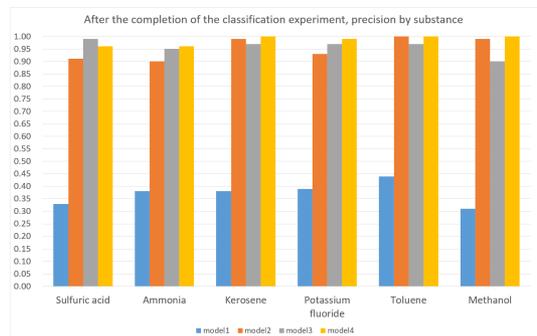


그림 9. 분류실험 종료 후 물질별 정밀도  
Fig. 9. Precision by Material After Classification Experiment Completion

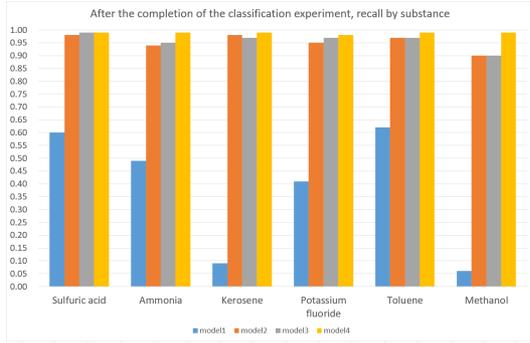


그림 10. 분류실험 종료 후 물질별 재현율  
Fig. 10. Recall by Material After Classification Experiment Completion

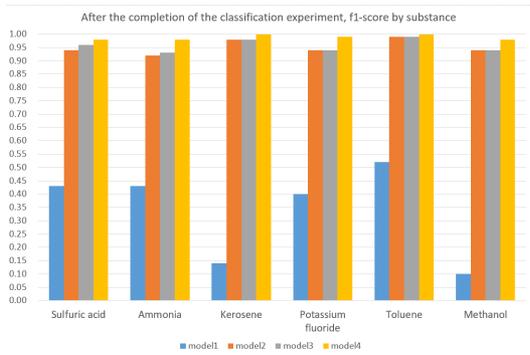


그림 11. 분류실험 종료 후 물질별 f1-score  
Fig. 11. F1-score by Material After Classification Experiment Completion

#### 5.4 노이즈가 섞인 데이터에 대한 분류 정확도

이번 실험은 실제 야외 유해화학물질 누출현장이나, 화재 발생 현장처럼 다른 요인에 의해 영향을 많이 받을 수밖에 없는 환경에서의 데이터에 대한 성능을 확인하기 위해 노이즈가 들어간 데이터를 이용해 실험한다.

검증 데이터 셋은 그림 12에 표시한 영역처럼 물질과 외부와의 경계 및 유리병 산란의 영향이 포함된 부분에서 추출한 등유의 분광 스펙트럼 데이터 2,464개를 사용해서 분류를 진행하며, 분류 정확도와 분류시간을 지표로 성능을 평가한다.

노이즈 데이터 분류실험 결과를 그림 13에 그래프로 나타냈다. 모델별 왼쪽의 막대가 정확도, 오른쪽 막대가 분류시간을 의미한다. 모델들의 정확도를 보면 모델 1을 제외하고 나머지 3개의 모델은 비슷한 정확도를 보이고 있다. 실제로 모델 2, 3의 정확도는 83.16%를 기록했고, 모델 4는 83.85%를 기록했다. 정확도의 차이는 0.69%p로 많은 차이를 보이지는 않았다. 분류시간의 경우 모델 1 ~ 3의 분류시간이 거의 비슷하게 7초대의

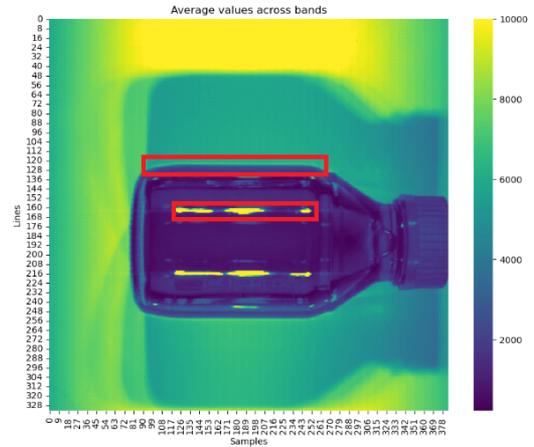


그림 12. 노이즈 데이터 분류실험 시 데이터 추출 영역 히트맵

Fig. 12. Noise Data Classification Experiment - Heatmap of Classification Data Extraction Area

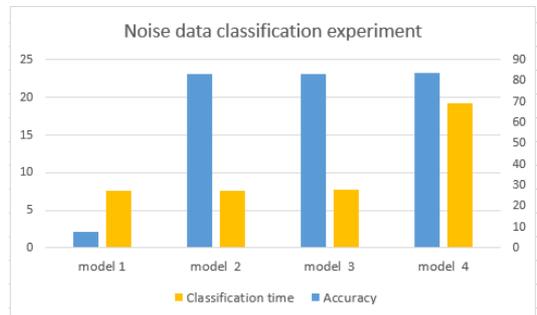


그림 13. 노이즈 데이터 분류실험 결과  
Fig. 13. Heatmap of Classification Data Extraction Area during Noise Data Classification Experiment

시간이 소요되었으며 분류시간의 차이가 0.1초 미만이었다. 모델 4만 분류시간이 19초 이상의 결과가 도출되었다.

#### 5.5 실험 결과 및 성능에 대한 고찰

노이즈 데이터 분류 실험에서 모델 2와 모델 3은 정확도에서 모델4와 0.69%p의 차이가 있었음에도 불구하고, 분류 시간이 2배 이상 빨랐기 때문에 속도와 정확도 면에서 가장 우수했다. 대량 데이터 분류 실험에서는 모델 4가 가장 높은 정확도를 보였지만, 모델 3과는 큰 차이가 없었던 반면, 분류 시간 측면에서는 마지막 실험에서 모델 4는 처리능력의 부족으로 모델 3보다 1분 이상 더 소요되었다. 6종의 물질에 대한 판독률은 모델 2, 모델 3, 모델 4가 모두 높게 나타났고, 시각적 특징이 부족한 유해화학물질을 높은 정확도로 판별할 수 있음

을 증명하였다. 6번의 실험 중 2번에서 모델 3이 유해화학물질 분류에 가장 적은 시간을 소요되었으며, 다른 실험에서도 가장 빠른 모델과 큰 차이를 보이지 않았다. 모든 실험에서 모델 3의 정확도는 가장 높았던 모델 4와 비교해도 3%p 이하의 차이로 매우 정확하게 물질을 분류했다. 이는 노이즈 데이터가 혼합된 데이터 셋을 분류할 때도 동일했다.

따라서, 본 논문에서는 제안된 네 가지 판독 모델 중에서 특성을 효과적으로 적용한 모델 3이 유해화학물질 판독 시스템의 인공지능 모델로서 우수한 성능을 보였다. 모델 3은 모델 2의 학습 특성에 그래프 기술기와 관련된 세 가지 추가 특성을 포함하여 학습되었다. 이 추가된 특성으로 인해 모델 3은 모델 1과 모델 2보다 더 높은 정확도를 달성할 수 있었다. 학습된 이 특성들은 분광 스펙트럼 데이터를 효과적으로 표현하여 모델 3이 기본적으로 높은 분류 정확도를 보였다.

## VI. 결 론

본 논문에서는 인공지능기반의 판독 시스템을 위해 동일한 색상이나 무색의 사고 화학물질을 구분하기 위한 방법으로 분광 스펙트럼 데이터 기반 판독기법과 효과적인 인공지능학습데이터를 구축하기 위한 방법과 유해화학물질의 판독을 위해서 분광 스펙트럼 데이터의 특성에 기반한 4가지 판독 모델을 제안하였다. 제안된 모델을 논문에서 구현한 인공지능기반 유해화학물질 판독 시스템의 머신러닝 알고리즘에 적용하여 제안된 4가지 판독 모델의 정확성을 실험하였다. 판독 모델들의 성능평가를 위해서 사용한 데이터는 황산, 암모니아, 등유, 플루오라이드칼륨 수용액, 톨루엔, 메탄올의 6가지 무색 물질의 분광 스펙트럼 데이터를 사용했다. 성능평가에서 가장 우수한 성능을 보인 모델 3을 적용함으로써 시각적 특징이 없는 유해화학물질 사고 발생 시에도 신속한 판독이 가능해짐에 따라 선제적 대응을 통해 화학 사고의 피해를 줄일 수 있을 것으로 기대된다.

향후 연구로는 기존의 영상을 통한 물질 판독 시스템과의 연계, 논문에서 학습한 6종의 물질 외의 다른 시각적 특징이 없는 물질의 데이터를 통한 판독 가능 물질 확대, 다른 파장대의 분광 스펙트럼 데이터를 통한 기존의 물질 데이터 보강 등이 필요하다.

## References

[1] National Fire Agency, “In 2022, there were 218 chemical accidents in South Korea.

*Efforts are being made to enhance the response capabilities to rare chemical incident,”* [https://www.nfa.go.kr/nfa/news/press-release/press/?boardId=bbs\\_000000000000010&mode=view&cntId=1731&category=&pageIdx=31&searchCondition=all&searchKeyword=](https://www.nfa.go.kr/nfa/news/press-release/press/?boardId=bbs_000000000000010&mode=view&cntId=1731&category=&pageIdx=31&searchCondition=all&searchKeyword=), Feb. 2024.

- [2] S.-B. Han, “A study on the remote measurement system for the detection of harmful gases in enclosed space,” Ph.D. dissertation, Pukyong National University, 2021.
- [3] N.-W. Cho, et al., “Development of remote sensing & monitoring technology of hazardous gases(2015),” Retrieved Feb. 2024.
- [4] J. Choi, “A study on the fabrication of printed Indium-Tin-Oxide thin film for the application electrochemical sensor to detect Hazardous and Noxious Substances,” Ph.D. dissertation, National Korea Maritime & Ocean University, 2021.
- [5] M. J. Lee, “Development of target-specific receptors and sensitive biosensors for the on-site detection of disease and chemical hazards detection,” Ph.D. dissertation, Inha University, 2022.
- [6] S. W. Lim, “Establishment of integrated disaster prevention system for hazardous materials and its optimal operation: Focused on semiconductor parts cleaning process,” Ph.D. dissertation, Seoul National University of Science and Technology, 2020.
- [7] D.-H. Kim, “A study on the improvement of image-based fire detection performance of ships using composite channel(RGB+IR),” Ph.D. dissertation, Chungnam National University, 2023.
- [8] A. U. Russo, et al., “Smoke detection method based on LBP and SVM from surveillance camera,” *2018 Int. Conf. Comput., Commun., Chem., Mater. and Electr. Eng. (IC4ME2)*, pp. 1-4, Rajshahi, Bangladesh, Feb. 2018. (<https://doi.org/10.1109/IC4ME2.2018.8465661>)
- [9] S. K. Cho, et al., “A prototype system for

detecting hazardous chemicals using artificial intelligence visualization techniques,” in *Proc. Symp. KICS*, pp. 101-102, JeJu Island, Korea, Jun. 2023.

- [10] C. Jung, “A study on pigment analysis of dancheong using hyperspectral technique - focusing on green colors -,” Ph.D. dissertation, Korea National University of Cultural Heritage, 2019.
- [11] M. J. Park, “Analytical study of oil paints using hyperspectral technique,” Ph.D. dissertation, Korea National University of Cultural Heritage, 2020.
- [12] ZHANGRUIRUI, et al., “Searching damaged pine trees by wilt disease based on deep learning using multispectral image,” *J. KICS*, vol. 45, no. 11, pp. 1823-1830, 2020. (<https://doi.org/10.7840/kics.2020.45.11.1823>)
- [13] Y. Gwon, et al., “Development of tracer concentration analysis method using drone-based spatio-temporal hyperspectral image and RGB image,” *J. Korea Water Resour. Assoc.*, vol. 55, no. 8, pp. 623-634, 2022.
- [14] Y. Gwon, D. Kim, and H. You, “A standardized procedure on building spectral library for hazardous chemicals mixed in river flow using hyperspectral image,” *J. Korea Water Resour. Assoc.*, vol. 53, no. 10, pp. 845-859, 2020.
- [15] Y. Gwon, “Development of spectral library for hazardous chemicals using hyperspectral image,” Ph.D. dissertation, Dankook University, 2020.
- [16] J.-J. Park, et al., “Detection of toluene hazardous and noxious substances (HNS) based on hyperspectral remote sensing,” *J. Korean Earth Sci. Soc.*, vol. 42, no. 6, pp. 623-631, Dec. 2021.
- [17] S.-M. Lee, et al., “Hyperspectral imaging (HSI) application for detection of organic compounds in water,” *J. Korean Soc. for Environ. Anal.*, vol. 15, no. 3, pp. 179-187, 2012.
- [18] Y. J. Kim, et al., “Research and development

of AI convergence hazardous chemicals detection system,” *Int. Conf. ICT Convergence 2023*, pp. 3-4, Jeju Island, Korea, Oct. 2023.

유 성 민 (Seong-Min Ryoo)



2022년 2월 : 숭실대학교 전기 공학과 졸업  
 2023년 1월~현재 : (주)우경정 보기기술 연구원  
 <관심분야> 인공지능, 초분광, 재난재해

김 연 진 (Yeon-Jin Kim)



2021년 2월 : 경북대학교 나노 소재공학부 신소재공학전공 졸업  
 2023년 8월 : 서원대학교 정보통신공학 석사  
 <관심분야> 빅데이터, 인공지능, 재난재해

조 속 경 (Sook-Kyoung Cho)



1990년 2월 : 인하대학교 전자계산학과 학사  
 1994년 2월 : 인하대학교 컴퓨터공학과 석사  
 2002년 2월 : 인하대학교 컴퓨터공학과 박사  
 2024년 2월~현재 : (주)네스토즈 연구소장

<관심분야> 유해화학물질, 영상 딥러닝, 빅데이터

**백 성 하 (Sung-Ha Baek)**



2005년 8월: 인하대학교 수학과 졸업  
2007년 8월: 인하대학교 컴퓨터공학과 석사  
2021년 3월: 인하대학교 컴퓨터공학과 박사  
<관심분야> 인공지능, 빅데이터

**김 경 배 (Gyoung-Bae Kim)**



1992년 2월: 인하대학교 전자계산공학과 졸업  
2000년 2월: 인하대학교 컴퓨터공학 박사  
2000년 3월~2004년 2월: 한국전자통신연구원 선임연구원  
2004년 3월~현재: 서원대학교 소프트웨어학부 교수  
<관심분야> 빅데이터, 인공지능, 재난재해, 클라우드 컴퓨팅