



머신러닝 모델을 활용한 산업단지 화재 재산피해 크기 예측

Predicting the Size of Fire Property Damage in an Industrial Complex Using a Machine Learning Model

이종호* · 최규진** · 박초롱*** · 이재욱**** · 손동욱*****

Jongho Lee* · Kyujin Choi** · Chorong Park*** · Jaewook Lee**** · Dongwook Sohn*****

* Main author, Research Specialist, Korea Institute of Civil Engineering & Building Technology, South Korea (leejongho@kict.re.kr)

** Coauthor, MS. course, Dept. of Architecture & Architectural Engineering, Yonsei Univ., South Korea (choikj331@yonsei.ac.kr)

*** Coauthor, Ph.D. course, Dept. of Architecture & Architectural Engineering, Yonsei Univ., South Korea (parkchorong@yonsei.ac.kr)

**** Coauthor, Professor, Dept. of Architecture & Architectural Engineering, Gachon Univ., South Korea (juklee@gachon.ac.kr)

***** Corresponding author, Professor, Dept. of Architecture & Architectural Engineering, Yonsei Univ., South Korea (sohndw@yonsei.ac.kr)

ABSTRACT

Purpose: This study proposes a novel approach to reduce the severe damages caused by factory fires in South Korea. The current fire risk assessment system faces limitations in providing detailed evaluations for factory buildings. This research utilizes public data and machine learning to swiftly and accurately predict fire risks in factories and seeks methods to identify and manage high-risk areas within industrial complexes. **Method:** The research process encompasses data collection, preprocessing, model prediction, and the integration of spatial data using GIS. It leverages building information provided by the national data portal and fire scenario data set as control variables. Data preprocessing includes the simplification of categorical variables, creation of derived variables, and the conversion of string data into numeric data. The predictive outcomes are integrated with spatial data using GIS, and industrial complexes are subdivided into blocks for risk level grading. This method aims to make a practical contribution to the management and prevention of fire risks in industrial complexes. **Result:** This study classified and analyzed the characteristics of factory buildings in aged three industrial complexes, assessing regional differences. Utilizing the Random Forest model, fire risks were categorized into low, medium, and severe levels, and regression analysis was employed to evaluate the impact of factors on fire risk. A five-tier grading system based on GIS visualization comprehensively represents the fire risk by region, offering valuable information for fire risk management. This research contributes to the development of policies aimed at enhancing safety in industrial complexes and minimizing property loss.

KEYWORD

기계학습
산업단지
화재 재산피해
공간지리정보

Machine Learning
Industrial Complex
Fire Property Damage
GIS

ACCEPTANCE INFO

Received Mar. 12, 2024

Final revision received Apr. 2, 2024

Accepted Apr. 8, 2024

© 2024. KIEAE all rights reserved.

1. 서론

1.1. 연구의 필요성

대한민국에서는 공장에서의 잦은 화재발생으로 인명피해와 재산 피해가 크게 발생하고 있다. 소방청 보도자료에 따르면 2016년에서 2020년까지 5년간 공장화재는 12,645건이 발생하였고, 사망 70명, 부상 830명인 900명의 사상자가 발생했다¹⁾. 9명 사망, 6명 부상을 입은 2018년 8월 21일 전자공장 화재를 비롯하여 38명 사망, 12명 부상을 입은 2020년에 발생한 물류센터 화재 사건 등 공장화재 사건들이 대형화재로 분류되고 있다²⁾. 도심 속 산업단지는 일반적으로 공장 건축물이 밀집 배치되어 있다. 이로 인해 화재 발생 시 인접 건축물로의 화재확산이 용이하여 화재 피해가 크고 복합적으로 나타난다³⁾. 이를 방지하기 위한 국내의 화재 관련 위험 평가제도를 살펴보면, 화재예방안전진단[1], 사전재난영향성검토[2], 지역안전지수[3] 등이 있다. 화재예방안전진단은 공공인프라시설을 대상으로 시설물 내부공간과 소방설비, 피난시뮬레이션 등을 세세하게 시행한다. 사전재난영향성검토는 초고층건축물을 대상으로 내부공간과

소방설비, 방화구획 등을 전문가 심의를 통해 정성적으로 평가한다. 두 가지 평가제도 모두 건축물 단위로 세세하게 평가하지만, 공장이 대상이 아닐뿐더러 건축물 내부의 세부 정보가 필요하고 인력과 시간이 소요된다는 단점이 있다. 지역안전지수의 경우는 지역단위의 통계자료를 가지고 정량적으로 평가하긴 하지만 건축물 단위의 평가는 이루어지지 않는다는 한계가 있다.

이에 본 연구는 국가에서 제공하는 공공데이터와 머신러닝 모델을 활용하여 다수의 공장 건축물의 화재위험을 빠르고 정확하게 예측하는 방안을 제안하고자 한다. 이러한 정량적 평가 방법은 제한된 데이터를 활용해 신속하고 쉽게 평가를 나타낼 수 있는 장점이 있어 다양한 형태를 가지는 산업단지에 적용하는 것이 적절할 것으로 판단된다. 또한 산업단지 내에 화재 피해가 클 것으로 예상되는 지역을 도출한 후 등급화하여 해당 지역의 화재 위험도를 관리할 수 있도록 하는 것을 연구의 목적으로 한다.

1.2. 연구의 범위

본 연구에서는 「건축법 시행령」 별표 1에 규정된 공장 건축물만을 분석한다. 공장별로 세부 업종이 있으며, 분석 범위는 한국표준 산업분류 체계를 활용하였다. 공간적 범위로는 산업단지 내 공장 건

축물로 한정하며, 건축물 간 거리나 소방서까지의 거리 등 주변 요인은 고려하지 않는다. 또한 본 연구는 국가에서 제공하는 공공데이터만을 활용하고 예측하는 것을 목적으로 하므로, 건축물 내에서 얻을 수 없는 세부적인 정보는 분석 대상에서 제외한다.

2. 이론적 고찰

본 장에서는, 화재위험 관련 분야에서 머신러닝 기법이 사용된 연구를 선정하여 학습데이터와 예측데이터에 대한 내용을 집중적으로 고찰하였다. 선행연구의 선정기준은 다음과 같다.

첫째, KCI(Korea Citation Index) 데이터베이스를 활용하여 “인공지능(AI)” 및 “화재” 키워드가 포함된 논문을 대상으로 검색을 수행하였다. 이러한 검색 과정을 통해, 본 연구의 핵심 주제인 ‘인공지능을 활용한 화재 예측’에 관련된 연구들을 포괄적으로 수집하고자 하였다.

둘째, 수집된 논문들 중에서 데이터를 기반으로 한 머신러닝 모델을 개발하거나 활용한 연구들만을 선별하였다. 이 기준은 본 연구의 핵심 목표인 간단한 데이터만을 활용하여 화재를 예측하는 모델 개발에 집중하기 위함이다. 이 과정을 통해, 본 연구와 직접적으로 관련된, 실제 적용 가능성이 높은 연구들에 초점을 맞추고자 하였다.

셋째, 해당 연구들의 접근 방법, 사용된 데이터의 유형 및 범위, 모델의 성능, 그리고 실제 적용 사례 등을 종합적으로 고려하여 본 연구의 이론적 토대를 마련하였다. 이를 통해, 본 연구의 차별성과 실용성을 강화할 수 있는 연구들을 체계적으로 분석하고자 하였다.

도시 분야의 연구로 서민송 외(2021)는 화재발생 건수가 가장 높은 경기도를 대상으로 SVM(Support Vector Machine), RF(Random Forest), GBRT(Gradient Boosting Regression Trees) 방법을 활용하여 화재발생위험요인 예측 분석을 실시하였다. 화재조사 데이터 외 시군구 면적, 인구밀도, 유동인구 등을 추가한 총 25개의 변수를 학습시켰으며 그 결과 발화장소가 화재 발생에 가장 큰 영향을 주는 위험요인으로 예측되었다[4].

정보학 분야의 연구로 김태준, 김한준(2018)은 DNN(Deep Neural Networks) 방법을 활용하여 총 7,350건의 화재 건축물 데이터셋을 정규화 후 모델에 학습시켰으며 정밀한 화재 예방을 위해 건축물 수준에서의 화재 위험 군집을 예측하였다[5]. 고정석 외(2018)는 대구광역시를 대상으로 전기안전점검데이터, 전기화재사고정보, 건축물 정보, 기상청 정보 등을 기반으로 전기화재를 예측하는 알고리즘을 개발하였으며 이를 통해 전기화재에 영향을 주는 요인과 예측 모델을 도출하였다[6]. 고정석 외(2020)는 MLP(Multi-Layer Perceptron) 신경망 모델을 활용하여 화재예측 모델을 개발하였다. 전국 건축물화재 데이터와 건축물 속성을 융합한 데이터를 학습 데이터로 사용하였으며 그 결과 화재위험도를 5가지 등급으로 분류해 주는 예측 알고리즘을 제시하였다[7].

방재 분야의 연구로 윤홍식 외(2022)는 건축물 화재에서 열적 환경을 상층부와 하층부로 구분해 예측해주는 화재모델인 CFAST를 활용하여 단일 및 다층 건축물의 화재 현상에 대한 예측 성능을 평가하였다[8]. 김도형, 조병완(2021)의 연구에서는 머신러닝 알고리즘의 성능 테스트를 진행하여 정확도가 가장 높게 나온 랜덤 포레스트

Table 1. Summarizes the research using machine learning

Field	#	Aurthor	Prediction Data
Urban Engineering	[4]	M.S. Seo et al. (2021)	Fire occurrence factors
	[5]	T.J. Kim, H.J. Kim (2018)	Fire hazard building clusters
Informatics	[6]	K.S. Ko et al. (2018)	Electrical fire accident
	[7]	K.S. Ko et al. (2020)	5 Levels of fire risk
Disaster prevention	[8]	H.S. Yun et al. (2022)	Fire growth and temperature, etc.
	[9]	D.H. Kim, B.W. Jo (2021)	Fire risk
Chemical Engineering	[10]	S.H. Ahn et al. (2015)	Number of monthly fire incidents in 2012
	[11]	C.W. Kim, D.I. Shin (2020)	3 Levels of fire risk

알고리즘을 이용한 상관성 분석을 진행하였다. 64,647건의 화재에 대한 자료를 분석하여 나온 216개의 데이터를 학습데이터로 활용하였으며 화재위험도를 정의하고 이를 예측하는 화재현장 위험도 예측 모델을 개발하였다[9].

마지막으로 화공 분야의 연구로 안상훈 외(2015)에서는 시간에 따라 변하는 시계열 모델과 모델의 타당성을 확인할 수 있는 ARIMA 모델 방법을 사용하여, 2012년 이전 5년간의 화재 데이터로부터 2012년 화재 발생 건수를 예측하는 모델 검증과정을 통해 화재 발생 빈도 예측 모델을 제안하였다[10]. 김창완, 신동일(2020) 연구에서는 머신러닝 기법인 K-평균 클러스터링을 이용하여 2017-2012년 국가에서 발생한 화재 사고를 등급별로 분류하고 주기성 분석을 통해 초대형 화재사고의 발생 시점 예측 가능성을 판단하였다[11]. Table 1.은 앞서 언급한 논문들을 요약한 것이다.

상기 연구들을 종합해보면, 다양한 방법론을 활용하여 화재 예측 및 위험 요인 분석에 대한 연구가 활발하게 진행되고 있음을 알 수 있다. 하지만 선행연구 고찰을 통해 분석한 화재 예측 연구는 대체로 복잡한 데이터셋과 다양한 변수를 필요로 한다. 예를 들어 도시, 정보학, 방재, 화학 등 각 전문분야에서 집중하고자 하는 데이터를 예측하고자 하여, 유동인구, 안전점검데이터, 건축 도면 등 굉장히 세부적인 정보를 필요로 한다. 이는 본 연구에서 다루고자 하는 산업단지 내 수 천개의 불특정 다수의 건축물의 화재위험을 예측 하는 것의 어려움을 초래한다. 이 경우에는 실용적인 적용에 있어 상당한 제약을 가지게 된다. 반면, 본 연구는 국가 데이터 포털에서 제공하는 공공데이터와 건축물의 기본적인 정보만을 활용하여 화재를 예측하고자 한다. 이는 건축물의 위치, 구조, 업종 등 제한된 데이터를 통해 빠르고 정확한 화재 예측이 가능함을 시사한다. 이러한 접근법은 지방 자치단체나 관할 소방본부에서 많은 수의 건축물과 지역을 대상으로 화재 위험 평가를 신속하게 수행할 수 있도록 하여, 예방 조치를 취하는 데 큰 도움이 될 것으로 예상된다.

3. 방법론 및 대상지 선정

3.1. 방법론

본 연구는 기 개발된 머신러닝 모델을 사용하여 산업단지의 건축물의 화재 재산 피해 크기 등급을 예측하고, 블록단위의 위험도를 도출하는 것이 목표이며, 다음과 같은 방법론이 사용된다[12].

첫째, 예측하고자 하는 산업단지의 데이터를 Table 2.와 같이 수집하고 구축한다. 필요한 예측 데이터셋은 구득 가능한 데이터와 구득할 수 없는 데이터로 분류할 수 있다. 구득 가능한 데이터는 정부가 제공하는 국가 데이터 포털을 통해 수집할 수 있는 건축물 정보이다. 반면에 화재 시나리오로 지칭한 화재에 영향을 주는 요인(발화열원, 발화요인 등)은 획득할 수 없기 때문에, 이들은 통계 변수로 설정되어야 하며, 데이터셋 내에서 임의의 값이 구성되어야 한다. 통제변수에 대한 상세한 내용은 3.5절에서 서술한다.

둘째, 데이터셋을 구축한 후에는 다음과 같은 데이터 전처리 단계를 수행한다.

- 1) 범주형 변수의 단순화, 2) 파생 변수 생성, 3) 문자열 데이터를 숫자 데이터로 변경

Table 2. Variable properties

Variable Name	Feature	Variable Type	Type of Use
Facility location information	Building information	Continuous	Independent
Number of floors	Building information	Continuous	Independent
Number of basement floor	Building information	Continuous	Independent
Total floor area	Building information	Continuous	Independent
Architecture area	Building information	Continuous	Independent
Structure	Building information	Categorical	Independent
Industry type	Building information	Categorical	Independent
Fire-resistant structure	Building information	Categorical	Independent
Season of fire	Fire scenario	Categorical	Independent
Time of fire	Fire scenario	Categorical	Independent
Ignition factor	Fire scenario	Categorical	Independent
Ignition material classification	Fire scenario	Categorical	Independent
Ignition point classification	Fire scenario	Categorical	Independent
Human casualties	Fire scenario	Categorical	Independent
Burnt area/TFA	Fire scenario	Continuous	Independent

데이터 전처리 후에는 머신러닝 모델을 통해 산업단지 공장 건축물들의 화재 재산 피해 크기 등급을 예측한다.

셋째, GIS를 통해 결과 데이터셋과 공간 정보와 결합한다. 지리 정보 시스템인 Esri의 ArcGIS Pro를 데이터 통합을 위해 사용한다. 지역 경계, 도로, 건축물과 같은 기본 공간 정보 이외의 데이터는 도로명이나 X, Y 좌표를 포함한 스프레드시트에서 제공되어, 이를 위치 정보로 변환할 필요가 있다. 따라서, ArcGIS Pro 프로그램의 지오코딩 도구를 활용하여 주소 데이터를 위도와 경도 좌표로 변환하고, 이를 포인트 데이터로 확보한다.

마지막으로 산업단지를 도로로 나누어 블록화 하고, 머신러닝 모델을 통해 예측된 결과를 총 건축물의 개수와 위험 건축물의 개수의 비율로 등급을 나누어 평가한다. 이를 지도를 통해 시각화하고 예측 모델 결과의 실용적인 적용을 제안한다.

3.2. 머신러닝 모델 개요

본 연구에서는 선행연구[13]를 통해 개발된 RF모델을 활용하며, 머신러닝 모델의 특징은 Table 3., 성능은 Table 4.와 같다.

RF모델은 화재 피해 예측의 복잡성을 다루기 위해 다수의 DT(Decision Tree)모델을 활용하는 앙상블 학습 방식을 채택하고 있다. 각각의 결정 트리는 데이터의 다양한 부분 집합에서 학습되며, 최종적으로 모든 트리의 예측을 종합하여 더 정확한 결과를 도출한다. 학습 데이터셋은 공공데이터포털을 통해 구득한 10년간(2009~2018)의 공장화재사고 데이터이며 품질을 보장하기 위해 철저한 데이터 전처리 과정을 거쳤다.

예측을 위한 입력 데이터는 건축물 정보와 화재 시나리오 정보로 분류되고, 이는 Table 2.에서 설명한 변수와 동일하다. 출력 데이터는 화재로 인한 재산 피해의 수준을 3등급(낮음, 중간, 심각)으로 분류한다. 등급의 기준이 되는 수치는 10년간(2009~2018)의 공장화재사고에 대한 재산피해액을 3분위로 나눈 값이다.

Table 3. Characteristics of machine learning models

Field	Output
Data Set	• 11,393 fire incidents over 10 years(2009-2018)
Test Data	• 7,975(70% of Data Set)
Validation Data	• 3,418(30% of Data Set)
Input Data	• Same as Table 2
Output Data	• Low : Property damage ≤ 3600 • Moderate : 3600 < Property damage ≤ 25000 • Severe : 25000 < Property damage

Table 4. Performance of machine learning models

Grade	Precision	Recall	F1-Score
Low	0.729	0.761	0.745
Moderate	0.644	0.605	0.624
Severe	0.841	0.860	0.851
Overall	0.738	0.742	0.740

모델의 성능은 정밀도, 재현율, F-1 스코어 등의 지표를 통해 평가되었다. 전체 데이터셋의 30%가 무작위로 검증데이터로 사용되었으며, 심각 등급에 대해 매우 높은 재현율(0.860)을 기록했으며 전체 F1 점수(0.740)도 높았다. 이는 상기 모델이 심각한 화재를 감지하는 데 효과적이며 전체적으로 균형 잡힌 성능을 제공한다는 것을 의미한다.

3.3. 대상지 선정 및 데이터 전처리

3.1절에서 언급한 데이터셋을 구축하는 과정에서, 공장의 업종 데이터를 얻기 위해서는 국가 산업 단지를 대상 지역으로 설정할 필요가 있다⁵⁾. 2022년 기준으로 총 47개의 국가 산업 단지가 있으며, 이 중 20개는 완공되었고 25개는 건설 중이며 2개는 미개발 상태이다. 이 연구는 20년 이상 운영된 노후 산업 단지를 선택하는 것을 목표로 하며, 노후단지는 17개가 존재한다.

서론에서 밝힌 것과 같이 도심 속 산업단지는 일반적으로 공장 건축물이 밀집 배치되어 있어 화재 발생 시 인접 건축물로의 화재확산이 빠르다. 그리고 주거 및 상업지구와 근접해 있을 경우, 이러한 화재사고는 지역 주민들에게 상당한 재산 및 인명피해를 야기할 수 있다. 주목할 점은, 대규모의 공장이나 플랜트같은 시설은 화학물질 등으로 인해 폭발성 화재의 특성을 보이는 경우가 많아, 더욱 정교한 연구가 요구된다는 점이다. 따라서 이러한 대규모 공장 및 플랜트 시설은 본 연구의 범위에서는 제외하였다.

따라서 연구의 대상지는 다음 조건에 따라 선정되었다. 첫 번째, 소규모 공장이 밀집된 단지를 구분한다. 두 번째, 도시 지역(주거 및 상업)에 인접한 단지를 구분한다. 세 번째, 수도권, 광역시, 지방에서 유사한 면적의 산업단지를 각각 한 곳씩 선정하며 분석 결과는 Table 5와 같다.

상기 조건에 따라 수도권에 위치한 인천 남동 산업 단지, 광역시에 위치한 부산 명지 녹산 산업 단지, 지방에 위치한 구미 1 산업 단지 세 곳이 선정되었다. 이 지역에 위치한 공장의 특성은 3.4절에서 자세히 논의한다.

화재 피해 예측을 위해 건축물 정보는 ‘디지털 트윈 국토 포털’에서 얻을 수 있으며, 이 파일⁶⁾은 산업 단지를 기반으로 공장 업종 정보를 제공하고 있다.

취합된 데이터는 머신러닝을 통해 예측하기전 3.1절에서 언급한 전처리 과정이 필요하다.

범주형 변수의 경우, 각 변수 수준의 빈도(또는 비율)가 낮을 때 모델 성능이 감소한다. 따라서 범주형 변수의 수준 수를 단순화 할 필요가 있다. 따라서 시설 위치 정보에 대해 226개 지자체로 구분되어 있는 공장 건축물의 위치를 수도권, 광역시, 지방 등 3개 수준으로 재분류한다.

또한 예측에 필요한 내화구조 정보는 공공데이터에서 제공하고 있지 않아, 기존 변수를 기반으로 새로운 변수를 생성하였다. 건축법에 따르면 5,000m² 이상의 연면적에는 내화구조가 의무화되어 있기 때문에 이를 기준으로 내화구조 여부의 파생변수를 생성한다.

마지막으로 문자열 데이터를 머신러닝 모델이 읽을 수 있게 숫자형 카테고리로 변경한다(레이블 인코딩).

Table 5. National industrial complexes more than 20 years after the start of construction

Name of Industrial Complex	Regional classification	Area	Dense in Small-scale	Neighborhood of City
Gunsan 2	Province	51,715	X	X
Daebul	Province	20,886	X	X
Gunsan	Province	13,702	X	X
Gumi (1st Complex)	Province	10,089	O	O
Gwangju	Metropolitan City	9,991	X	O
Namdong	Province	9,574	O	O
Myeongji and Noksan	Metropolitan City	8,841	O	O
Asan National	Province	6,156	X	X
Bupyeong National	Province	4,278	X	X
Samil Resource Reserve	Province	4,157	X	X
Wolsong Nuclear Plant	Province	3,693	X	X
Jisepo Resource Stockpile	Province	2,942	X	X
Korea Export	Metropolitan Area	3,711	O	O
Paju Publishing	Metropolitan Area	1,562	X	X
Iksan National History	Province	1,336	X	O
Daejuk Resource Reserve	Province	912	X	X
Paju Tanhyeon	Metropolitan Area	80	O	X

3.4. 대상지 데이터 특성

머신러닝을 기반으로 화재피해 위험등급을 예측하기 전에 Table 2.에서 언급했던 변수 중 공공데이터를 통해 수집되고 전처리된 건축물 정보에 대한 데이터 분석을 대상지별로 실시하였다. 공장 건축물의 위치 및 표본수, 지상 층수, 지하 층수, 업종, 내화 구조, 건축물 구조, 연면적, 건축면적과 같은 각 요소의 특성을 빈도 분석을 통해 검토하였다.

1) 위치 및 표본수

인천 남동 산업단지에는 총 4,051개의 공장이 있다. 부산 명지녹산 산업단지에는 총 3,890개의 공장이 있으며, 구미1 산업단지에는 총 5,091개의 공장이 있다. 종합해보면 분석 대상 지역에 위치한 공장의 합계는 13,032개이다.

2) 건축물 지상 층수

대상지 별로 산업단지 내 건축물의 층수를 Fig. 1.과 같이 나타낸

다. 인천은 1층과 2층의 건축물이 전체 4051개 중에서 1506개(약 37.18%), 1397개(약 34.49%)으로 유사하게 가장 많은 것으로 나타났다. 그다음 3층, 4-5층, 6층 이상 순으로 816개(약 20.14%), 308개(약 7.6%), 24개(약 0.59%)로 나타났으며 비교적 완만한 기울기로 감소하는 것을 확인할 수 있다. 부산도 전체 3890개의 건축물 중에서 단층이 2602개(약 66.89%)으로 가장 많았고 2층 931개(약 23.93%), 3층 286개(약 7.35%), 4-5층 70개(약 1.8%), 6층 이상 1개(0.03%) 순으로 층수가 높아질수록 건축물 수가 줄어드는 것을 확인할 수 있다. 구미는 전체 5091개의 건축물 중에서 단층이 4418개(약 86.78%)로 다른 지역들과 비교했을 때 가장 높은 비율을 차지하는 것으로 나타났으며 그 외 2층부터 501개(약 9.84%), 108개(약 2.12%), 56개(약 1.1%), 8개(약 0.16%)으로 급격하게 줄어드는 것을 확인할 수 있다.

전체적으로 층수가 높아질수록 건축물 수가 줄어드는 것을 확인할 수 있었는데 구미지역 단층 건축물이 가장 많았으며 부산지역의 6층 이상의 건축물이 1개로 가장 적은 것으로 나타났다.

3) 건축물 지하 층수

지역별 산업단지 내 건축물의 지하 층수 현황을 Fig. 2와 같이 나타낸다. 인천은 지하층이 없는 건축물이 전체 4051개 중에서 3263개(약 80.55%)로 가장 높은 비율을 차지하고 있는 것으로 나타났으며 그다음 지하 1층, 지하 2층 순으로 784개(약 19.35%), 4개(약 0.10%)로 나타났다. 부산도 전체 3890개의 건축물 중에서 지하층이 없는 건축물이 3859개(약 99.2%)로 가장 높은 비율을 차지하고 있었으며 지하 1층이 30개(약 0.77%), 지하 2층이 1개(약 0.1%)로 지하층이 많아질수록 낮은 비율을 차지하고 있는 것을 확인할 수 있다. 구미 또한 전체 5091개의 건축물 중에서 지하층이 없는 건축물이 4891개(약 96.7%)로 가장 높은 비율을 차지하고 있었으며 그 외 지하 1층과 지하 2층이 각각 197개(약 3.87%), 3개(약 0.06%)로 나타났다.

전체적으로 모든 지역에서 지하층이 없는 건축물이 80% 이상을 차지하고 있어 대부분의 건축물이 지하층이 없는 것을 확인할 수 있었으며 특히 부산지역의 경우 약 99% 이상이 지하층이 없는 건축물로 이루어져 있는 것을 알 수 있다.

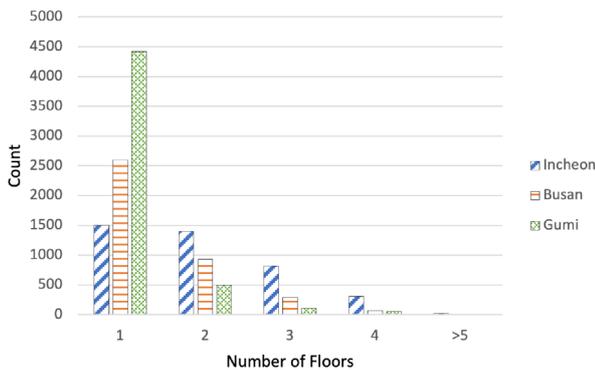


Fig. 1. Distribution of building floors by region

4) 건축물 업종 현황

지역별 산업단지 내 건축물 업종 현황은 Fig. 3과 같다. 인천은 전기전자 산업(EEI)에서 가장 많은 기업을 보유하고 있어, 전자제품의 제조 및 기술 개발에서 강점이 있는 것으로 확인되었다. 또한, 인천은 금속기계 및 장비 산업(MMEI)에서도 높은 기업 수를 보여, 이 분야에서 강력한 산업 기반을 가지고 있음을 나타낸다.

부산은 목재 가공 및 목공업(WPCI)에서 높은 기업 수를 보여, 이 분야에서 견고한 산업 기반을 가지고 있는 것으로 나타났다. 또한, 화학 산업(CI)과 펄프 및 종이 산업(PPI)에서도 상대적으로 높은 기업 수를 보여, 이 분야에서 활발한 산업 활동이 이루어지고 있음을 알 수 있다.

반면 구미는 다른 두 지역에 비해 전반적으로 낮은 기업 수를 보이나, 화학 산업(CI)과 전기전자 산업(EEI)에서 일정 수준의 기업을 유지하고 있는 것으로 확인되었다. 이는 구미가 이 분야에서 안정적인 산업 기반을 가지고 있음을 나타낸다.

5) 건축물 구조

세 지역 모두에서 건축물 구조 유형은 Fig. 4와 같으며, 철근 콘크리트 구조(RC)가 총 10,279개 건축물로 가장 높은 비율을 나타냈다. 구미가 4,499개로 가장 많고, 부산이 3,303개, 인천이 2,477개로 뒤를 이었다. 다음으로, 철골 구조(SFC)가 총 1,977개 건축물로 두 번째로 높은 비율을 나타냈으며, 인천이 1,124개로 가장 많고, 구미가 448개로 가장 낮았다. 마지막으로, 벽돌 구조(BC)에서는 지역 간에 상당한 차이가 있는 것으로 나타났다. 인천은 총 308개 건축물 중

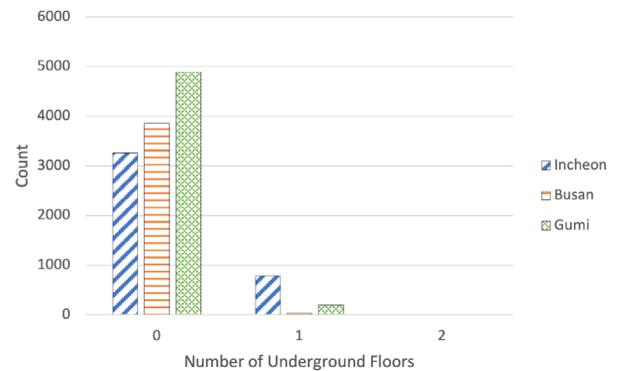


Fig. 2. Distribution of building underground floors by region

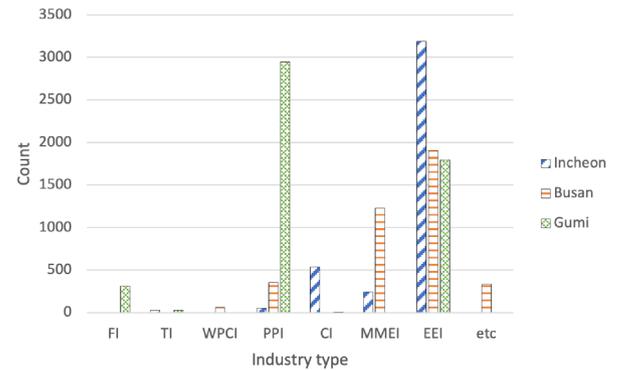


Fig. 3. Distribution of industry type by region

203개로 높은 비율을 차지하고 있는 것에 비하여, 부산과 구미는 각각 34개, 71개로 낮은 비율을 차지하고 있는 것으로 집계되었다.

6) 건축물 내화구조

지역별 산업단지 내 건축물의 내화구조 현황을 분석해보면 Fig. 5와 같이 나타난다.

모든 지역에서 평균 약 95.97%의 비율로 대부분의 건축물이 내화구조로 되어있지 않은 것으로 확인되었다. 먼저 인천의 경우 전체 건축물 4051개 중에서 3851개(약 95.05%)의 건축물이 내화구조로 되어있지 않았고 부산은 3757개(약 96.58%), 구미는 4901개(약 96.27%)의 건축물이 내화구조로 되어있지 않은 것으로 나타났다.

본 연구에서 내화구조 정보는 공장 건축물의 연면적 정보만을 가지고 생성한 파생변수 이기 때문에, 이 점은 유의할 필요가 있다.

7) 연면적

연면적을 Fig. 6과 같이 살펴보면, 가장 작은 연면적 범위(0-499m²)가 가장 높은 비율을 차지하고 있는 것으로 확인되었으며, 7,480개의 공장이 이 범위에 속하였다. 이는 작은 규모의 공장이 주를 이루고 있음을 나타낸다. 또한, 연면적이 증가함에 따라 공장 수는 감소하는 추세를 보였는데, 이러한 추세는 세 지역 모두에서 일관되게 나타났다.

또한, 구미는 인천과 부산에 비해 가장 큰 연면적 범위(5000m² 이상)에 속한 공장 수가 더 많은 것으로 확인되었는데, 이는 구미에 큰 규모의 산업 운영이 많을 수 있다는 것을 나타낸다. 반면, 인천은 가

장 작은 연면적 범위(0-499m²)에 속한 공장 수의 비율이 높아, 비교적 소규모 사업체 운영이 많을 수 있는 것으로 나타났다.

8) 건축면적

Fig. 7과 같이, 건축면적 분석결과로 가장 작은 건축 면적 범위(0-499m²)가 총 8,113개의 공장으로 가장 높은 비율을 차지하고 있으며, 이 범위에 인천이 1,935개, 부산이 2,414개, 구미가 3,764개가 속하였다. 또한, 건축 면적이 커질수록 공장 수는 감소하는 경향을 보였으며, 가장 큰 건축 면적 범위(5000m² 이상)에는 총 260개의 공장이 속하였다.

구미는 가장 작은 건축 면적 범위(0-499m²)에 공장 수가 많은 것으로 확인되었다. 이는 앞서 연면적에서 나타난 경향과 일치하며, 구미에 소규모 공장이 집중되어 있음을 나타낸다.

부산은 중간 건축 면적 범위에 걸쳐 비교적 균등한 비율의 공장 수를 보였다. 연면적과 함께 고려하면, 부산은 다양한 크기의 공장이 균형있게 분포된 산업 구조를 보이는 것으로 분석된다.

인천은 큰 건축 면적 범위의 공장 수가 비교적 많은 것으로 나타났다. 이는 연면적과 일관된 결과로, 대규모의 공장 비율이 높은 것을 나타내며, 인천이 대규모의 공장이 많은 산업 환경을 가지고 있을 가능성을 나타낸다.

3.5. 통제변수의 설정

선행 연구[13]에서 ‘화재 시나리오 정보’를 고려하지 않고 건축물

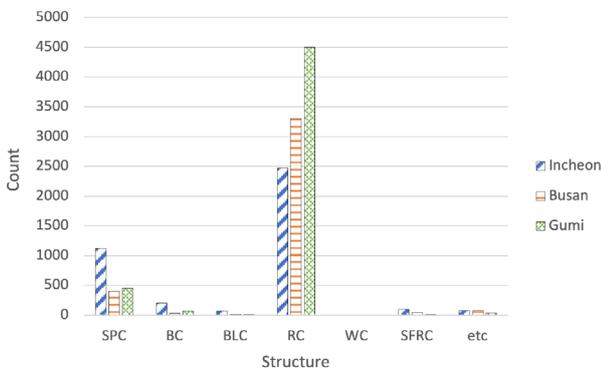


Fig. 4. Distribution of building structure region

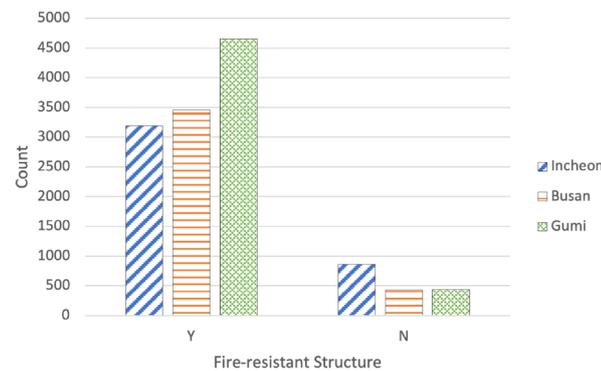


Fig. 5. Distribution of fire-resistant structures by region

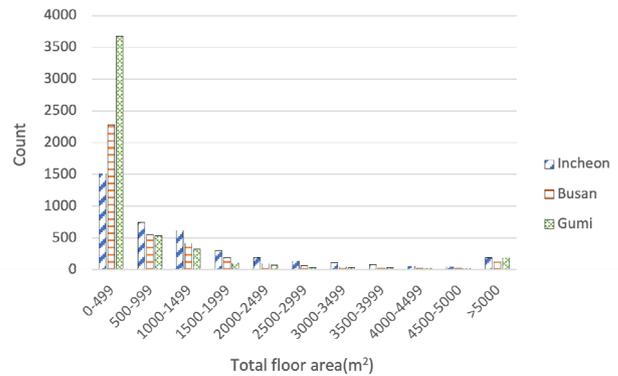


Fig. 6. Distribution of total floor area by region

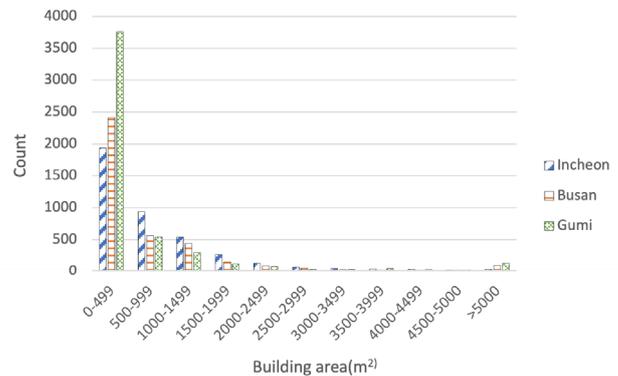


Fig. 7. Distribution of architecture floor area by region

정보만을 사용하여 개발한 모델은 제한적인 예측 성능을 보이며, 반면에 건축물 정보에 ‘화재 시나리오 정보’를 포함하여 개발한 모델은 더욱 향상된 예측 성능을 보이는 것을 확인하였다.⁷⁾

이에 본 연구에서는 화재 발생에 대한 예측의 정확성을 높이고자 ‘건축물 정보와 화재 시나리오 정보’를 통해 학습한 화재위험을 예측하고자 하였다. 하지만 예측을 위해 필요한 화재 발생계절, 시간대, 발화 재료, 발화 지점, 화재 원인, 인명 피해 및 소실 면적 등의 변수들은 화재가 발생하기 전에는 생성되지 않는 정보이기 때문에, 화재가 발생하지 않은 건축물에 가상의 변수를 설정해야만 한다. 따라서 본 절에서는 이를 통제 변수로 설정하는 방법에 대해 논의한다. 이는 데이터 전처리 과정에서 누락된 값을 처리하는 과정과 유사하다. 가장 일반적으로 사용되는 방식은 수치 데이터의 경우 누락된 값을 평균으로 대체하고, 범주형 데이터의 경우에는 최빈값으로 대체한다. 이 방법을 따라 ‘화재 시나리오 변수’를 설정할 수 있다. 자세한 정보는 Table 6.과 같다.

10년간의 화재발생 데이터 11,393건을 분석한 결과, 계절적 요인으로 겨울철에 화재 발생률이 가장 높은 것으로 나타났다. 시간에 따라서는 오후 12시부터 오후 6시 사이에 화재가 가장 자주 보고되었다. 발화 재료는 대부분 알 수 없는 것으로 나타났는데, 이는 화재 후 남은 잔해에서 확실한 발화 지점을 식별하기 어려운 것으로 유추된다. 발화 공간은 주로 기능적 공간에서 발생하였다. 기능적 공간은 제조와 같은 특정 작업이 이루어지는 공간을 의미한다. 또한 화재요인으로는 전기적 요인이 화재의 가장 많은 원인으로 나타났다. 인명 피해는 대부분의 화재 사건에서 ‘없음’으로 기록되었다. 소실면적의 평균 비율은 31.29%로 기록되었다.

본 연구에서는 화재에 영향을 미치는 요인에 대한 통제 변수로 최빈값과 평균값이 설정되었다. 하지만 이 변수들은 화재 피해 예측의 목적, 관리 주체, 공장의 특성에 따라 사용자가 맞춤으로 설정할 수 있으며, 목적과 환경에 맞는 화재피해크기를 예측하는 데 사용될 수 있다.

4. 화재재산피해 크기 예측

4.1. 예측결과

랜덤 포레스트(RF) 모델을 사용한 테스트 데이터셋의 예측 결과는 Fig. 8.과 같다.

Table 6. Setting of control variables

No	Factors	Characteristic	Value	
1	Season	Categorical	Mode	Winter
2	Time	Categorical	Mode	12:00~18:00
3	Ignition Material	Categorical	Mode	Unknown
4	Ignition Space	Categorical	Mode	Functional
5	Causes	Categorical	Mode	Electric Cause
6	Human Injury	Categorical	Mode	No
7	Burnt Area (%)	Numerical	Average	31.29%

인천, 부산, 그리고 구미에 위치한 세 개의 산업단지에서의 화재 위험 분포는 낮은 등급, 중간 등급, 그리고 심각 등급의 세 단계로 분류된다. 총 데이터 수는 13,032건으로, 인천이 4,051건, 부산이 3,890건, 구미가 5,091건을 차지한다. 낮은 위험등급은 구미에서 1,930건(전체의 약 14.8%)으로 가장 높으며, 그 다음으로 부산이 926건(약 7.1%), 인천이 530건(약 4.1%)이다. 중간 위험등급은 구미에서 1,553건(약 11.9%)으로 가장 높고, 부산이 1,137건(약 8.7%), 인천이 593건(약 4.6%)이다. 심각 위험등급은 인천에서 2,928건(약 22.5%)으로 가장 높으며, 그 다음으로 부산이 1,827건(약 14.0%), 구미가 1,608건(약 12.3%)이다. 전반적으로, 심각한 위험등급이 가장 큰 비율을 차지하며, 전체 예측치의 약 48.8%인 6,363건을 차지하고 있다. 반면에, 낮음과 중간 위험등급은 각각 약 26.0%와 25.2%를 차지한다.

분석된 결과에 따르면, 연면적은 공장 건축물의 위험 수준을 결정하는 데 중요한 요인이다. 구체적으로 살펴보면, 연면적이 59m² 미만인 공장 건축물은 일관되게 낮은 위험등급으로 분류되었다.

연면적이 59m²에서 107m² 사이인 건축물의 경우, 구조형식에 따라 위험 수준이 달라졌으며, 콘크리트 구조나 다른 유형의 구조는 낮은 위험으로 분류되었다. 연면적이 267m²에서 337m² 사이의 건축물은 주로 중간 위험등급으로 분류되었고, 417m² 이상인 건축물은 대부분 심각한 위험등급으로 분류되었다.

특히 주목할 만한 범위는 277m²에서 287m² 사이였으며, 이 구간이 중간 위험등급과 심각한 위험등급의 차이가 가장 적었다. 이 범위 내 83건 중 39건이 심각한 위험등급으로 분류되었다.

분석결과 층수, 위치(수도권 내), 식품 산업 등의 요소가 심각 위험등급에 주로 영향을 주었다. 한편, 중간 위험등급의 44건은 펄프 및 종이 산업, 기타 구조, 기타 업종 등의 요소가 영향을 준 것으로 분석되었다.

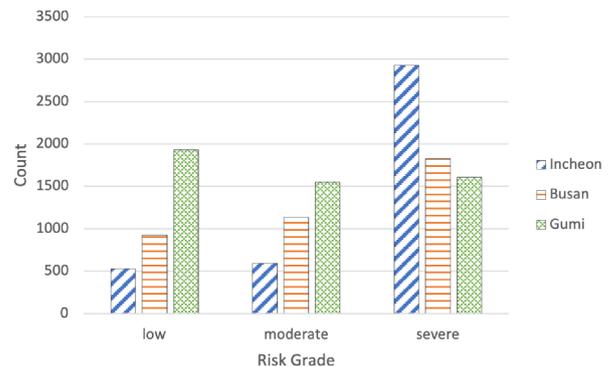


Fig. 8. Predicted fire damage outcomes

Table 7. Distribution of total floor area according to prediction results

Risk Grade	Average	Gumi	Busan	Incheon
Low	4~107m ²	4~93m ²	4~107m ²	4~104m ²
Moderate	59~498m ²	59~414m ²	59~498m ²	59~396m ²
Severe	267m ² ~	273m ² ~	267m ² ~	267m ² ~

모든 연면적 범위에서 층수, 위치, 업종, 구조에 따라 위험 등급이 분류되는 일관된 패턴을 보인다. 이러한 발견은 층수, 위치, 업종, 구조가 건축물 위험 평가 및 관리에 신뢰할 수 있는 지표로 작용 가능하다는 것을 뜻한다.

Table 8.은 건축물 정보에 대한 회귀 분석 결과를 나타낸다. 건축물 위치, 층수, 지하 층수, 연면적, 건축 면적, 구조, 업종, 내화구조 등의 다양한 입력 변수가 화재피해 등급인 출력 변수에 미치는 영향을 β 계수, 표준 오차(SE), F-값, p-값을 사용하여 보여준다.

분석 결과, 층수 ($\beta = 0.371, p < .001$)와 지하 층수 ($\beta = 0.111, p < .001$)는 통계적으로 유의하며 양의 관계를 가지고 있다. 이는 층수와 지하 층수가 증가하면 화재 피해 등급도 증가한다는 것을 의미한다. 또한, 구조 ($\beta = 0.119, p = 0.004$)와 내화구조 여부 ($\beta = 0.108, p < .001$)가 결과 변수에 유의한 영향을 미친다. 특히 내화구조 여부는 매우 중요한 요소이다. 업종도 의미 있는 영향을 미친다 ($\beta = 0.079, p = 0.011$). 연면적은 유의한 결과를 보이지만 ($p < .001$ 및 $p = 0.015$) 건축면적은 통계적으로 유의하지 않다 ($p = 0.074$).

전반적으로, 이 회귀 모델은 R^2 가 0.249로 선택된 입력 변수가 결과 변수의 변동성의 약 24.9%를 설명한다. 모델 전체적으로 통계적으로 매우 유의하다 ($F = 11.326, p < .001$).

4.2. GIS를 통한 가시화

머신러닝을 통한 화재피해 크기의 예측 결과의 활용성을 높이기 위해서 본 장에서는 산업단지를 도로의 최소 단위인 4미터 도로폭 기준으로 나누어 블록을 생성하고, 5단계의 등급을 부여하여 시각화한다. 그 절차와 방법은 다음과 같다.

- (1) 벡터 분석을 통해 처리된 데이터를 통합하기 위해, 폴리곤 데이터를 래스터 데이터로 변환한다.
- (2) Raster Calculator 도구를 사용하여 각 래스터 값의 합계를 계산하여 픽셀 당 총 값(Value)을 산출한다.
- (3) 공장 건축물의 화재 피해 예측 결과를 지역별 화재 위험 레벨로 등급을 매기기 위해, Zonal Statistics 도구를 사용하여 지역별로 픽셀 값의 평균을 계산한다.
- (4) 각 블록 내에서 심각 등급의 건축물 수를 블록 내 전체 건물 수로 나누어 다섯 등급(0~100%)의 화재 위험 지도를 구성한다.

Table 8. Regression analysis on input variables for output variables

Category	β	SE	F	p-value
Location	.140	.061	5.324	.005
Number of Floors	.371	.067	30.690	<.001
Number of Underground Floors	.111	.033	11.528	<.001
TFA	.185	.114	2.659	.015
Building Area	-.063	.035	3.209	.074
Structure	.119	.041	8.480	.004
Industry Type	.079	.048	2.763	.011
Fire-resistant Structure	.108	.031	1.059	<.001

$R^2=.249, F=11.326, p<.001$

그 결과는 통계 결과(Table 9.)와 지도(Table 10.)와 같다.

인천 남동 산업단지의 지도를 검토해보면, 짙은 빨간색으로 표시된 지역일 수록 건축물 화재안전에 대한 우려를 나타낸다. 특히 지도의 남동부와 일부 중앙 지역에 고위험 건축물이 집중되어 있는 것을 관찰할 수 있다. 반대로, 밝은 색으로 표시된 지역은 일반적으로 공원과 같은 공장 건축물이 적은 개방된 공간에 해당하여, 건축물에 대한 위험이 낮음을 시사한다. 전반적으로 지도는 고위험 및 저위험 지역의 분포를 명확하게 구분한다. 통계 데이터에 따르면 인천의 건축물 중 3등급과 4등급 위험을 가진 건축물이 각각 44.43%, 42.86%를 차지하여, 지도상에 관찰된 광범위한 짙은 빨간색 지역과 일치합니다. 이 분포는 인천에 중간에서 고위험 수준의 건축물이 상당수 존재함을 시사한다.

부산 명지녹산 산업단지의 위험 지도를 살펴보면, 지도의 오른쪽과 중앙 부분에 위치한 블록이 짙은 빨간색으로 표시되어 위험한 건축물의 높은 밀도를 나타낸다. 그 외 대부분의 지역은 밝은 색으로 표시되어 있으며 상대적으로 안전한 것으로 분석되며, 이 지역들에서는 건축물의 화재 안전성이 전반적으로 더 높은 것을 시사한다. 일부 중앙 블록에서 위험 수준의 변동성을 관찰할 수 있으며, 이는 지역 내 건축물 상태의 비균일성을 나타낸다. 통계를 검토하면, 부산의 3등급 위험 건축물이 60.34%로 가장 높은 비율을 차지하며, 이는 지도상에 보이는 많은 3등급 지역과 일관성을 보인다. 또한, 4등급 건축물의 비율은 6.9%로 상대적으로 낮으며, 2등급 건축물의 비율은 28.45%로 구미 및 인천에 비해 상대적으로 높다.

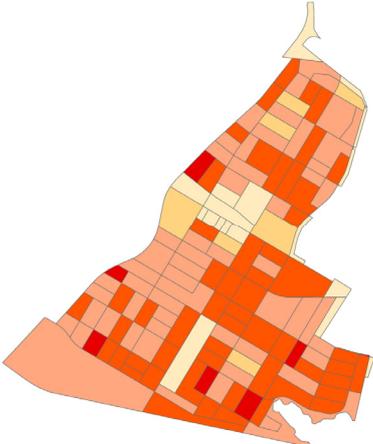
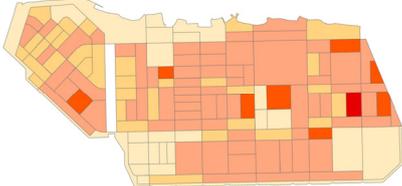
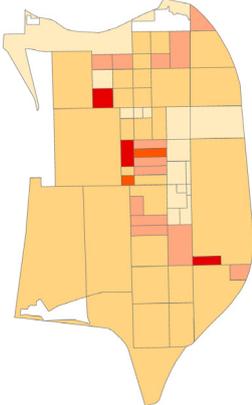
구미1 산업단지는 저위험 건축물이 지배적으로, 지도상에서 밝은 색으로 표현되며, 고위험 건축물이 있는 지역은 짙은 빨간색으로 명확하게 표시된다. 특히 중앙 지역, 오른쪽 하단, 왼쪽 상단 모서리가 건축물 안전성 향상을 위한 중요 관리 지역으로 식별된다. 통계 분석에 따르면 구미에서 2등급 건축물이 55.2%로 가장 높은 비율을 차지하며, 이는 지도상에 보이는 많은 밝은 색 지역에 해당한다. 구미는 다른 두 도시에 비해 3등급 건축물의 비율이 26.53%로 낮으며, 이는 전반적으로 저위험 건축물이 우세함을 나타낸다.

공통적으로 고위험 지역은 화재 안전예방을 위한 재건축, 보강, 또는 철거를 포함한 다양한 개선 조치를 우선적으로 고려해야 하는 지역으로 간주할 수 있다. 본 연구에서 도출한 통계 및 시각 데이터는 소방본부나 지방자치단체가 위험 감소 노력을 우선적으로 계획하고 실행하는 데 필수적인 정보를 제공한다. 또한, 특정 지역의 위험 관리 및 감소를 위한 정책을 수립하는 데 있어 중요한 기반을 제공할 수 있다.

Table 9. Statistical result

Grade	Incheon		Busan		Gumi	
	N	%	N	%	N	%
0-20	2	1.59	4	3.45	4	8.16
21-40	7	5.56	33	28.45	27	55.20
41-60	56	44.43	70	60.34	13	26.53
61-80	54	42.86	8	6.90	2	4.08
81-100	7	5.56	1	0.86	3	6.12
SUM	126	100.0	116	100.0	49	100.0

Table 10. Result maps of target areas

Area	Grade of Risk Area on GIS		
Incheon Nam Dong			
Busan Myeongji Noksan			
Gumi 1			
Grade	Color	%	
1	Light Yellow	0-20	
2	Yellow	21-40	
3	Orange	41-60	
4	Dark Orange	61-80	
5	Dark Red	81-100	

5. 결론

본 연구는 공장 건축물에서의 화재 피해를 예측하기 위한 공간 분석에서 머신러닝과 지리 정보 시스템(GIS)을 통합하는 분석적 접근 방식을 제안한다. 연구의 주요 목적은 화재 피해를 예측하기 위해 머신러닝 모델을 사용하는 것으로, 예측을 위해서는 구득 가능한 정보와 구득이 불가능한 데이터로 분류된 데이터 셋의 구축이 필요하다. 구득 가능한 데이터는 정부가 제공하는 국가 데이터 포털에서 상세한 건축물 정보를 포함하며, 화재에 영향을 미치는 구득 불가능한 데

이터는 통계 변수와 가상 값으로 대체된다. 데이터셋을 구축하는 과정에는 GIS를 사용하여 건축물 데이터를 공간 정보와 병합하는 것을 포함하고, 이 과정은 예측된 결과를 시각화하는 것에도 연계가 되므로 매우 중요하다.

화재피해 예측을 위한 대상으로 인천 남동 산업 단지, 부산 명지 녹산 산업 단지, 구미1 산업 단지와 같은 20년 이상 된 산업 단지를 선정하였으며, 각 지역의 공장 건축물 특성을 분류하는 작업을 수행하였다. 이를 통해 공장 건축물의 지역 차이를 분석하고자 하였다. 그런 다음 랜덤 포레스트 모델을 사용하여 화재 위험을 낮음, 중간, 심각의 세 가지 범주로 분류하였다. 또한, 건축물 특성이 화재 위험에 미치는 영향을 평가하기 위해 회귀 분석을 수행하였으며 이를 통해, 층 수, 건축물 구조, 업종과 같은 요소의 중요성을 강조하였다. 마지막으로 다섯 단계 등급체계를 사용한 GIS 기반 화재 위험 시각화는 다양한 지역의 화재 위험을 포괄적으로 나타내어 지자체와 관찰 소방본부가 화재위험 관리를 위한 지점으로 활용할 수 있다.

이 연구가 화재 위험 예측 및 관리 분야에 기여하는 바는 다음과 같다. 데이터 분석 기술과 공간 시각화를 결합함으로써, 산업 단지에서의 화재 위험을 분석하고 완화하기 위한 정량적이고 체계적인 접근 방식을 제공한다. 그리고 본 연구 방법론을 산업 환경에서의 안전성 향상과 화재로 인한 재산 손실을 최소화 하기 위한 정보에 근거한 의사 결정 및 정책 개발에 적용할 수 있다. 예를 들어 건축 계획 측면에서 GIS 데이터를 연결하여 새 공장의 위치와 주변 건축물의 화재 위험을 분석하고, 새로운 공장 설립을 제한하거나 제안할 수 있다. 또한, 연면적에 기반한 특정 범위를 설정함으로써 더 상세한 정책을 개발할 수 있다.

본 연구의 한계점은 구득 불가능한 데이터의 통계변수 대체가 실제 조건을 완벽히 반영할 수 없다는 점이다. 이는 국가에서 제공하고 있는 공공데이터의 한계이며, 추후 더 상세한 정보가 공공데이터를 통해 제공된다면 더 상세하고 정확한 예측 모델을 적용 할 수 있을 것으로 기대한다.

Acknowledgement

본 연구는 과학기술정보통신부 한국건설기술연구원 연구운영비 지원(주요사업)사업으로 수행되었습니다(과제번호 20240189-001, 산업공단 초고속 화재 예측·제어·대응기술 개발).

References

- [1] 소방청, 소방안전 특별관리시설물 화재예방 안전진단 세부 절차 및 평가방법 등에 관한 규정, [별표 7], 화재위험성평가 분야 평가기준. // (National Fire Agency, Regulations on detailed fire prevention safety diagnosis procedures and evaluation methods for fire safety special management facilities, [Appendix 7], Evaluation standards in the field of fire risk assessment.)
- [2] 소방청, 초고층 및 지하연계 복합건축물 재난관리에 관한 특별법, 제6조(사전재난영향성검토협의). // (National Fire Agency, Article 6 (Preliminary disaster impact review consultation) of the special act on disaster management of high-rise and basement-connected complex buildings.)
- [3] 행정안전부, 재난 및 안전관리 기본법, 제66조의10(안전지수의 공표). //

- (Ministry of the Interior and Safety, Article 66-10 of the Framework Act on Disaster and Safety Management (announcement of safety index).)
- [4] 서민송 외 3명, 머신러닝을 이용한 경기도 화재위험요인 예측분석, 한국측량학회지, 제39권 제6호, 2021.06, pp.351-361. // (M.S. Seo et al., Predictive analysis of fire risk factors in Gyeonggi-do using machine learning, Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 39(6), 2021.06, pp.351-361.)
 - [5] 김태준, 김한준, DNN 기반 감독형 매니폴드 학습과 계층 군집 분석을 통한 화재 예측 기법, 정보과학회 컴퓨팅의 실제 논문지, 제24권 제8호, 2018.08, pp.422-426. // (T.J. Kim, H.J. Kim, Fire prediction technique using DNN-based supervised manifold learning and hierarchical clustering analysis, KIISE Transactions on Computing Practices, 24(8), 2018.08, pp.422-426.)
 - [6] 고경석 외 3명, 기계학습을 통한 전기화재 예측모델 연구, 한국정보전자통신기술학회 논문지, 제11권 제6호, 2018.06, pp.703-710. // (K.S. Ko et al., Electrical fire prediction model study using machine learning, The Journal of Korea Institute of Information, Electronics, and Communication Technology, 11(6), 2018.06, pp.703-710.)
 - [7] 고경석 외 5명, 인공지능 기반 건물화재 예측모델 연구, 한국통신학회논문지, 제45권 제7호, 2020.07, pp.1210-1218. // (K.S. Ko et al., Building fire prediction model study using AI, The Journal of Korean Institute of Communications and Information Sciences, 45(7), 2020.07, pp.1210-1218.)
 - [8] 윤희석 외 2명, 단일 및 다층 건축물에서 화재특성에 대한 CFAST의 예측성능 평가, 대한기계학회 논문집 B권, 제46권 제8호, 2022.08, pp.441-451. // (H.S. Yun et al., Evaluation of predictive performance of CFAST for fire characteristics in single and multi-storey buildings, Transactions of the KSME, B, 46(8), 2022.08, pp.441-451.)
 - [9] 김도형, 조병완, 빅데이터 분석을 활용한 초기 정보 기반 화재현장 위험도 예측 모델 개발 연구, 한국재난정보학회 논문집, 제17권 제2호, 2021.02, pp.245-253. // (D.H. Kim, B.W. Jo, A study on the development of a fire site risk prediction model based on initial information using big data analysis, Journal of the Korean Society of Disaster Information, 17(2), 2021.02, pp.245-253.)
 - [10] 안상훈 외 4명, ARIMA 모델에 기반한 화재발생 빈도 예측모델의 설계, 한국가스학회지, 제19권 제2호, 2015.02, pp.20-28. // (S.H. Ahn et al., Forecasting model design of fire occurrences with ARIMA models, Journal of the Korean Institute of Gas, 19(2), 2015.02, pp.20-28.)
 - [11] 김창완, 신동일, 초대형화재사고 예측을 위한 화재사고 분류의 개선 및 발생의 주기성 분석, 한국가스학회지, 제24권 제1호, 2020.01, pp.56-65. // (C.W. Kim, D.I. Shin, Improved classification of fire accidents and analysis of periodicity for prediction of critical fire accidents, Journal of the Korean Institute of Gas, 24(1), 2020.01, pp.56-65.)
 - [12] 이종호, 공공데이터를 활용한 머신러닝 기반 공장 화재 피해 크기 예측 연구, 연세대학교 일반대학원, 2024. // (J.H. Lee, Predicting the size of factory fire damage using public data and machine learning, Yonsei University Graduate School, 2024.)
 - [13] J.H. Lee, J.W. Shin, J.W. Lee, C.R. Park, D.W. Sohn, Development of a data-based machine learning model for classifying and predicting property damage caused by fire, Applied Sciences, 13(21), 11866, 2023.

건물 일반 정보 서비스"변수 특성

- 7) (J.H. Lee et al., 2023)에 따르면, 건축물 정보 8개 변수만을 가지고 화재피해를 예측하는 RF 모델의 종합 F1-Score가 0.400에 그치는 반면, 건축물정보 8개 변수에 화재 시나리오 7개 변수를 포함하여 개발한 RF모델은 종합 F1-Score가 0.740으로 예측성능이 크게 향상됨을 알 수 있다.

- 1) (보도자료) '소방청, 공장밀집 산업단지 화재안전 강화한다'. https://www.nfa.go.kr/nfa/news/pressrelease/press/?boardId=bbs_0000000000000010&mode=view&cntId=1115
- 2) 대한민국의 화재조사 및 보고규정에서는 인명피해가 사망 5명 이상 또는 사상자 10명 이상 발생한 화재이거나, 재산피해가 50억 원 이상 추정되는 화재를 대형화재로 정의하고 있다.
- 3) (뉴스) '대구 산단 공장 화재로 한때 동원령까지...'에 따르면 밀집된 형태의 공장구조로 인해 인근 공장 10여개 동으로 화재가 확산됨. <https://m.yonhapnewstv.co.kr/news/MYH20230616000800641>
- 4) 공공데이터포털. <https://www.data.go.kr/>
- 5) 일반적으로 공장의 업종 데이터는 건축물 또는 위치 기반이 아닌, 상호명과 함께 제공이 되고 있어 예측을 위한 데이터 셋을 구축하는 것이 용이하지 않다. 다만 국가산업단지의 경우 단지 내 블록에 업종 구역을 지정하고 있어, 블록 내 건축물에 업종을 병합하는 방식으로 데이터 셋 구축이 가능하다.
- 6) https://www.vworld.kr/dtna/dtna_fileDataView_s001.do / 파일명"GIS