일반연구보고서 2023-10

빅데이터 기반 건축물 화재 및 홍수 리스크 분석 모델 개발 연구

A Study of Developing a Big-Data-Based Building Fire and Flood Risk Analysis Model

조영진 Cho, Youngjin
허한결 Heo, Hankyul
송유미 Song, Yumi
현태환 Hyeon, Taehwan

일반연구보고서 2023-10

빅데이터 기반 건축물 화재 및 홍수 리스크 분석 모델 개발 연구

A Study of Developing a Big-Data-Based Building Fire and Flood Risk Analysis Model

지은이 조영진, 허한결, 송유미, 현태환
펴낸곳 건축공간연구원
출판등록 제2015-41호 (등록일 '08. 02. 18.)
인쇄 2023년 12월 29일, 발행: 2023년 12월 31일
주소 세종특별자치시 가름로 143, 8층
전화 044-417-9600
팩스 044-417-9608

http://www.auri.re.kr
가격: 10,000원, ISBN: 979-11-5659-457-4

이 연구보고서의 내용은 건축공간연구원의 자체 연구물로서
정부의 정책이나 견해와 다를 수 있습니다.

연구책임

조영진 연구위원

연구진

허한결 부연구위원
송유미 연구원
현태환 연구원

외부연구진

김호걸 청주대학교 조교수
박현준 청주대학교 교수
백경열 청주대학교 연구원
김유진 한국과학기술원 연구원

연구평가위원

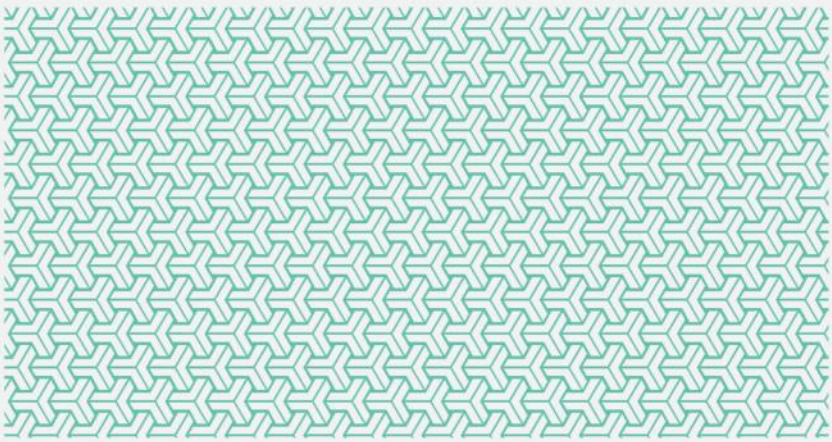
염철호 선임연구위원
김상호 선임연구위원
이여경 연구위원
김영철 한국과학기술원 교수
민범식 한아도시연구소 부회장
이승일 서울시립대학교 교수

자문위원

김길호 한국건설기술연구원 수석연구원
김민석 부경대학교 교수
김태린 상지대학교 조교수
문근종 계명대학교 부교수
박진한 한국환경연구원 부연구위원
성선용 한국전통문화대학교 교수
손동화 충북대학교 조교수
윤석환 서울대학교 박사
이정원 충남대학교 교수
임리사 한국과학기술원 조교수
최지현 서울특별시의회 입법조사관
최현철 희림종합건축사사무소 수석

연구요약

Summary



1. 서론

건축물의 화재와 홍수 피해가 국민 안전을 위협하고 있으며, 기후변화로 인해 이 같은 피해는 앞으로 더욱 증가할 것으로 전망된다. 건축물 화재와 홍수는 건축물에 피해를 입히는 재난·재해 중 가장 피해액이 크며 국민 안전을 위협하는 주요 원인이다. 특히, 2021년 이천 쿠팡 물류센터 화재와 2022년 여름 홍수로 인한 강남역 인근지역 침수와 반지하 거주자들의 사망사고는 국민의 건축물 화재 및 홍수 피해에 대한 불안감을 가중시키고 있다.

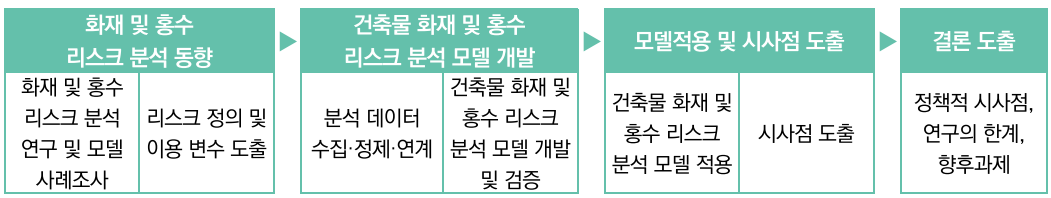
개별 건축물에 대한 화재 및 홍수 리스크를 사전에 인지하는 것은 건축물 안전 확보를 위해 매우 중요하다. 건축물 특성과 입지가 화재 및 홍수 발생 시 피해 정도에 영향을 주며, 이에 대한 리스크에 따라 사전에 건축물의 화재 및 홍수를 대비할 수 있다. 예를 들어 연면적이 큰 건축물은 상대적으로 화재가 발생할 가능성이 높고, 저지대에 위치한 건축물은 홍수피해를 입을 가능성이 높다. 연구자들은 건축물의 재난·재해 피해 발생 특성을 이용하여 건축물 화재 및 홍수 리스크 분석을 수행해 왔다. 이 같은 연구는 최근 관련 기술 발달로 전환점을 맞이하였다.

머신러닝, 딥러닝, IoT, ICT 등 제4차 산업혁명 기술 발달로 리스크 분석 정확도가 증가하고 있다. 의사결정나무, 랜덤 포레스트, 서포트 벡터 머신 등 머신러닝 알고리즘들이 리스크 분석 기술 향상을 가져왔으며, RNN (Recurrent Neural Network)과 LSTM (Long Short Term Memory), CNN (Convolutional neural network) 등 딥러닝 기반 기술 발전이 분석 정확도를 비약적으로 증가시켰다. IoT, ICT 기술 향상은 분석에 사용할 수 있는 데이터 확보에 기여하였으며, 빅데이터를 활용한 문제해결이 가능하도록 이끌고 있다.

건축물의 화재 및 홍수 피해에 대응하기 위해 화재와 홍수의 리스크가 높은 건축물의 식별이 필요하다. 또한, 리스크 상승에 기여하는 요인을 판단하여 이를 중심으로 건축물의 안전 관리를 수행할 필요가 있다. 이미 건축물의 안전을 확보하기 위하여 「건축물관리법」, 「소방시설 설치 및 관리에 관한 법률」, 「시설물의 안전 및 유지관리에 관한 특별법」 등에 따른 건축물 점검이 시행 중이며, 더 합리적인 건축물 정기점검 기반 마련에 본 연구의 결과가 기여할 수 있다.

이에 본 연구는 1) 건축물 단위의 홍수 및 화재 리스크 분석 모델 개발, 2) 서로 다른 데이터 연계 방법 실증 및 향후 데이터 연계를 위한 기초 확립, 3) 홍수 및 화재 대응이 필요한 건축물 도출 및 모델의 정확도 검증, 4) 건축물 화재 및 홍수 안전 제고를 위한 건축물 관리방안 제안을 목적으로 한다.

본 연구는 4단계로 진행하였다. 1) 화재 및 홍수 리스크 연구와 분석 모델을 조사하여 본 연구에서 다룰 건축물 화재 및 홍수 리스크를 정의하고 분석에 사용할 변수를 도출하였다. 2) 건축물 화재 및 홍수 리스크 분석 모델을 각각 개발하고 검증하였다. 모델에는 머신러닝(딥러닝) 알고리즘을 적용하였고, 이를 위해 분석 데이터를 수집·정제·연계하는 과정을 수행하였다. 3) 관악구 지역을 대상으로 건축물 화재 리스크 분석모델과 홍수 리스크 분석 모델을 각각 적용하고 이를 통합하여 건축물 화재 및 홍수 리스크 분석을 수행하였다. 4) 전체 연구를 바탕으로 정책적 시사점을 도출하고, 연구의 한계를 정리하는 한편 향후 연구과제를 제안하였다.



연구 흐름도

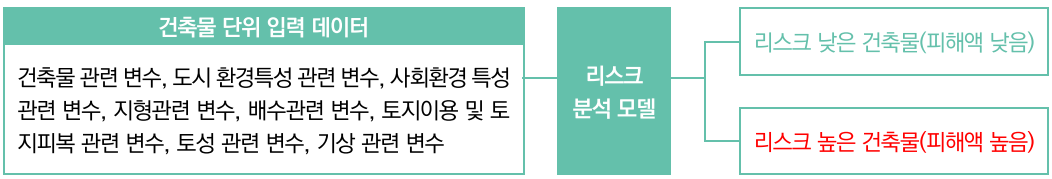
2. 화재 및 홍수 리스크 분석 관련 동향

건축물 화재 및 홍수 리스크 분석을 위해 화재 및 홍수 리스크 분석에 관한 선행연구와 분석 모델을 조사 및 분석하였다. 선행연구에서는 발생확률과 손상함수를 이용하여 리스크를 분석하는 것으로 나타났으며, 최종 리스크는 피해금액을 단위로 산정하는 경우가 가장 보편적이었다. 이는 리스크 분석이 실재 피해 규모를 평가하기 위해 피해액 단위의 결과를 도출하고 있는 것으로 판단된다.

리스크 산정 개념

화재 예측 모델의 경우 주로 통계적 방법론에 기반하여 연구가 진행되어 왔으며, 발생 강도와 피해액을 분석하지만 화재발생 자체를 예측하기는 어려운 것으로 판단된다. 이는 화재의 특성상 자연 발화 가능성이 낮고 대부분 인간의 실수나 방화에 의해 발생하는 경우가 많아 환경적 특성이나 입지가 화재 발생에 미치는 영향이 낮기 때문이다. 리스크 분석 연구는 통계적 화재 발생 모델을 기반으로 손상함수와 연계하여 최종적으로 피해액을 도출하는 방식으로 리스크를 산출한다. 건축물의 리스크 분석도 이와 유사한 방법을 따르는 것으로 판단된다. 이 같은 분석방식은 지역단위 분석 뿐 아니라 건축물단위 분석에서도 동일하게 적용된다.

홍수와 관련된 연구와 분석모델은 수리·수문학적 모델을 기반으로 홍수 발생 가능성과 발생 강도 예측을 주로 수행하고 있다. SWMM과 SWAT같은 모델들이 대표적이며, 해당 모델은 이미 지역별 홍수 대책 수립에 널리 적용되고 있다. 홍수 분석 모델은 강수량의 변화와 지형적 특성을 고려하여 다양한 상황에 따른 홍수 발생을 시뮬레이션 할 수 있으며, 이를 통해 발생 확률과 강도를 예측할 수 있다. 또한, 발생강도에 따른 피해액을 계산하는 손상함수와 결합하여 홍수 발생에 따른 리스크 분석을 수행하는 모델이 개발되었다. Hazus-MH, HEC-FIA 등이 대표적인 피해액 기반의 홍수 리스크 분석 모델로 사용되고 있다.



건축물 화재 및 홍수 리스크 분석모델의 틀

선행연구 및 모델 분석을 통해 본 연구에서 개발하고자 하는 건축물 화재 및 홍수 리스크 분석모델의 틀을 정립하였다. 건축물의 화재 및 홍수 리스크는 모두 피해금액 단위로 도출되도록 설정하였으며, 리스크 분석 기준은 예상되는 피해금액이 낮은 건축물과 높은 건축물을 기준으로 하였다. 발생확률과 손상함수는 통계적 모델에 기반을 두고 있는 머신러닝(딥러닝) 알고리즘을 사용하였다. 또한, 선행연구 및 모델 분석 결과를 바탕으로 건축물 화재와 홍수 리스크 분석을 위한 변수를 선정하였다.

재난·재해 분석 변수 정리

구분	입력변수
화재	도시 환경특성 관련 변수, 사회환경 특성 관련 변수, 기상 특성 관련 변수, 건축물 관련 변수
홍수	지형관련 변수, 배수 관련 변수, 토지이용 및 토지피복 관련 변수, 토성 관련 변수, 기상 관련 변수, 건축물 관련 변수

3. 건축물 화재 및 홍수 리스크 분석 모델 개발

건축물 화재 및 홍수 리스크 분석 모델 개발을 1) 데이터 구축 단계와 2) 모델 개발 및 검증 단계로 구분하여 수행하였다. 데이터 구축 과정은 ① 건축물 화재와 홍수 피해 발생 및 피해비용 데이터를 각각 건축물대장 데이터와 연계하는 과정과, ② 추가적인 독립변수를 생산 및 연계하는 과정으로 구성된다. 모델 개발 및

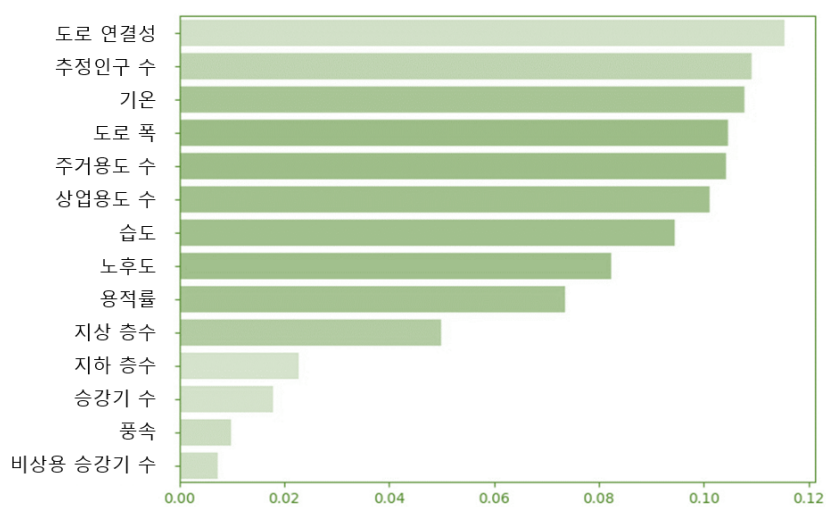
검증 과정은 ① 구축한 분석 데이터셋을 학습시킬 알고리즘을 선정하고, 선정된 알고리즘을 바탕으로 모델을 학습시킨 후, ② 검증 및 테스트 과정을 수행하는 피드백 과정을 거쳐 모델을 개발하는 과정이다. 건축물 화재 및 홍수 리스크 분석 모델은 각각 독립적으로 개발하였다.

데이터 구축		▶	모델 개발 및 검증	
화재 및 홍수 발생 데이터와 건축물 대장 연계	분석을 위한 입력변수 생산 및 건축물 대장과 연계		머신러닝(딥러닝) 알고리즘 선정 및 구축된 데이터 기반 모델 학습	검증 및 테스트를 수행하며 모델 피드백 및 최종 리스크 분석 모델 선정

모델 개발 과정 요약

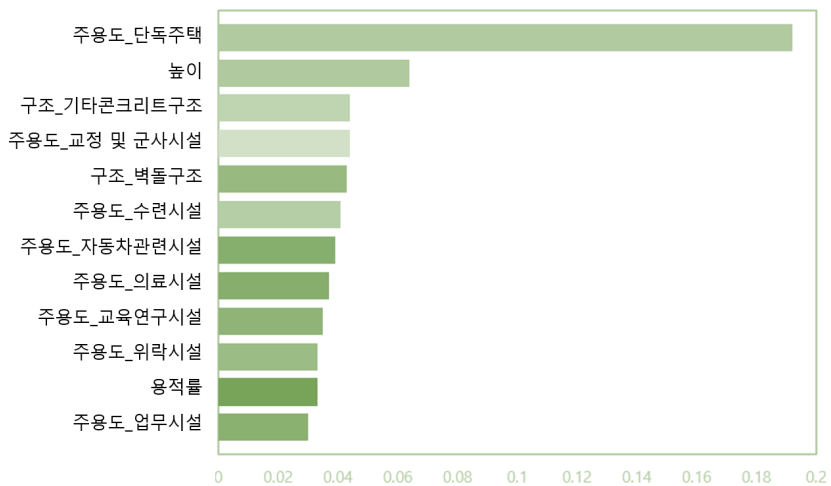
건축물 화재 리스크 분석 모델 개발에는 2017년부터 2021년까지의 총 28,266건의 서울시 건축물 화재피해 데이터를 사용하였다. 피해금액이 20만원 이상인 건축물의 경우 화재 리스크가 높은 건축물로, 20만원 미만인 경우 화재 리스크가 낮은 건축물로 분류하여 분석을 진행하였다. 랜덤포레스트, 로지스틱 회귀 분석, LightGBM, XGBoost 등 다양한 머신러닝 알고리즘을 적용해 보았으며, 가장 높은 정확도를 나타낸 램덤포레스트를 선정하여 모델을 도출하였다.

건축물 화재 리스크 분석의 주 영향 요인으로는 도로 연결성, 추정인구 수, 기온, 도로 폭, 주거용도 수, 상업용도 수 등으로 도출되었다. 또한, 건축물 화재 리스크 분석 모델의 최종 정확도는 78% 수준으로 나타났고, F1스코어는 0.75, ROC곡선의 아래 부분의 면적을 나타내는 AUC값은 0.89로 나타나 준수한 수준의 리스크 분석 모델을 개발하였다.



화재 리스크 분석 모델 변수 중요도

건축물 홍수 리스크 분석 모델 개발에는 2016년부터 2022년까지 총 27,438건의 서울시 건축물 홍수피해 데이터를 활용하였다. 300만원 이하와 초과의 총 2개 클래스를 홍수 리스크가 낮은 건축물과 높은 건축물로 구분하여 분석을 수행하였다. 건축물 홍수 리스크 분석 모델은 TabNet classifier를 사용하였다. 홍수 리스크 분석에 영향력이 높은 것으로 도출된 변수는 건축물의 용도와 구조, 그리고 건축물의 높이와 용적률 등으로 나타났다. 최종 건축물 홍수 리스크 분석 모델의 정확도는 88%로 매우 높은 수준의 정확도를 갖는 것으로 도출되었다.



홍수 리스크 분석 모델 변수 중요도

4. 건축물 화재 및 홍수 리스크 분석 모델 적용

서울시 관악구를 대상으로 건축물 화재 및 홍수 리스크 분석 모델을 적용하였다. 모델 개발에는 화재 및 홍수 피해가 발생한 건물을 이용하였으며, 모델 적용에는 서울시 관악구에 있는 모든 건축물에 대해 화재 및 홍수 리스크 분석을 수행하였다. 서울시 관악구에는 총 32,079동의 건축물이 존재하며, 분석을 위해 입력변수를 구축하고 모든 건축물 포인트에 데이터를 연계하였다.

[화재 리스크 지도] 빨간색: 화재 리스크 높음　　[홍수 리스크 지도] 파란색: 홍수 리스크 높음

건축물 홍수 및 화재 리스크 지도(개별 분석 결과)

구축된 데이터를 모델에 입력하여 건축물 화재 리스크 지도와 홍수 리스크 지도를 도출할 수 있다. 분석 결과 화재 피해가 클 것으로 예측되는 건축물은 총 8,405동(약 31%), 홍수 피해가 클 것으로 예측되는 건축물은 446동(약 1.5%)로 나타났다.

노란색: 화재 및 홍수 리스크 모두 낮음, 빨간색: 화재 리스크 높음
파란색: 홍수 리스크 높음, 별표: 화재 및 홍수 리스크 모두 높음

건축물 홍수 및 화재 리스크 지도(통합 분석 결과)

건축물 화재 및 홍수 리스크를 통합적으로 확인하기 위해 개별 예측 결과물을 중첩분석 하였다. 화재 및 홍수 리스크가 높고 낮음에 따라 총 4가지 유형의 건축물을 도출할 수 있다. 관악구 내 건축물 32,079동 중 화재 또는 홍수 리스크가 모두 낮은 건축물은 23,276동, 화재 리스크만 높은 건축물은 8,357동, 홍수 리스크만 높은 건축물은 398동, 화재 및 홍수 리스크가 모두 높은 건축물은 총 48동으로 도출되었다.

건축물 화재 및 홍수 리스크 유형

	화재 리스크 낮음	화재 리스크 높음
홍수 리스크 낮음	23,276동 (72.56%)	8,357동 (26.05%)
홍수 리스크 높음	398동 (1.24%)	48동 (0.15%)

5. 결론

본 연구는 크게 네 가지 정책적 시사점을 갖는다. 첫째, 건축물의 재난·재해 대응을 위해 제4차 산업혁명 기술의 활용 필요성을 제시하였다. 제4차 산업혁명 기술을 대표한다고 볼 수 있는 머신러닝, 딥러닝, 빅데이터 등 데이터 활용을 통해 건축물 화재 및 홍수 리스크 분석이 가능하며, 기존 방식보다 높은 정확도를 담보한다. 향후 이 같은 기술을 지속적으로 도입하고, 활용성을 제고함으로써 재난·재해로부터 건축물의 안전을 개선할 수 있을 것으로 판단된다.

둘째, 건축물 화재 및 홍수 리스크를 저감할 수 있는 방안을 제안하였다. 리스크 분석 모델의 개발 결과를 토대로 리스크 증가에 영향을 미치는 요인을 도출하였다. 건축물 화재와 관련하여 도로 연결성, 추정인구 수, 기온, 도로 폭, 주거용도 수 등의 변수가 주요 요인이었으며, 건축물 홍수와 관련해서는 건축물의 용도, 구조, 높이, 용적률 등이 도출되었다. 향후 이 같은 요인들에 집중하여 관리 정책을 개선할 경우 화재 및 홍수에 따른 피해를 저감시킬 수 있을 것으로 기대된다.

셋째, 건축물 단위 재난·재해 통합 대응 필요성을 제안하였다. 기존 화재 및 홍수 관련 모델과 연구들은 개별적인 재난과 재해에 대해 리스크 분석을 수행하였다. 그러나, 본 연구에서는 화재와 홍수 두 가지 모두에 대해 통합적인 대응이 가능하도록 모델을 개발하였다. 화재와 홍수는 발생 원인이나 대응 방식이 전혀 다를 수 있으나, 건축물 차원에서의 대응 우선순위를 도출한다는 측면에서는 이를 복합적으로 보는 것이 더 효율적일 수 있다. 성공적인 건축물 재난·재해 예방을 위해 향후 이 같은 건축물 통합예방안전 시스템 구축이 필요하다.

넷째, 재난·재해 취약 건축물 우선 관리를 위한 제도개선 필요성을 확인하였다. 화재와 홍수에 보다 취약한 건축물은 관계법에 따른 의무 점검 대상건축물 보다는 규모가 작은 소규모건축물이 많았다. 즉, 화재와 홍수에 취약한 건축물은 법적 점검이 이루어지지 못하는 실정으로, 이에 대응하기 위한 제도 개선이 필요하다. 현재 「건축물관리법」 제15조(소규모 노후 건축물등 점검의 실시)에서 소규모 노후 건축물에 대한 점검은 규정하고 있으나, 이에 대한 강화 방안이 필요하다.

뿐만 아니라, 본 연구와 관련하여 향후 공공데이터 연계의 강화 및 새로운 알고리즘 적용을 통해 리스크 분석의 정확도를 높이고, 화재 및 홍수 이외의 재난·재해 요소를 포함하는 등 건축물 안전 향상을 위한 관련 연구가 지속될 필요가 있다.

주제어

빅데이터, 건축물 화재 위험, 건축물 홍수 위험, 머신러닝, 딥러닝

차 례 CONTENTS

제1장 서론 1

1. 연구의 배경 및 목적 2

1) 연구의 배경 2

2) 연구의 목적 및 기대효과 6

2. 연구의 범위 및 방법 8

1) 연구의 흐름 8

2) 연구의 범위 9

3) 연구의 방법 10

제2장 화재 및 홍수 리스크 분석 관련 동향 13

1. 화재 피해 및 리스크 분석 14

2. 홍수 피해 및 리스크 분석 30

1) 홍수 발생 예측 모형 및 선행연구 30

2) 리스크 분석 모형 36

3. 소결 42

제3장 건축물 화재 및 홍수 리스크 분석 모델 개발 47

1. 모델 개발 절차 및 방향 48

1) 모델 개발 절차 48

2) 데이터 연계 방향 49

3) 대상지 선정 및 특성 50

2. 건축물 및 재해 데이터 연계 54

1) 건축물 화재 발생 데이터 공간정보화 54

2) 건축물 홍수피해 데이터 공간정보화 60

차 례 CONTENTS

3. 건축물 화재 리스크 분석 모델 개발 65
1) 입력 데이터 선정 65
2) 데이터 가공 67
3) 분석 모델 선정 및 학습 72
4) 모델 검증 및 영향 요인 도출 74
4. 건축물 홍수 리스크 분석 모델 개발 79
1) 입력 데이터 선정 79
2) 데이터 가공 83
3) 분석 모델 선정 및 학습 87
4) 모델 검증 및 영향 요인 도출 89
5. 소결 92

제4장 건축물 화재 및 홍수 리스크 분석 모델 적용 95

1. 건축물 화재 및 홍수 리스크 분석 96
1) 건축물 화재 리스크 분석 96
2) 건축물 홍수 리스크 분석 99
2. 건축물 화재 및 홍수 리스크 통합 분석 103
3. 소결 및 시사점 106
1) 리스크 분석 모델 106
2) 건축물 화재 및 홍수 리스크 분석 모델 107

차 례 CONTENTS

제5장 결론 109

1. 연구의 요약 110

1) 재난·재해 리스크 분석 동향 110

2) 건축물 화재 리스크 분석 모델 개발 111

3) 건축물 홍수 리스크 분석 모델 개발 111

4) 건축물 화재 및 홍수 리스크 통합 분석 모델 개발 112

2. 정책적 시사점 113

3. 연구의 한계 및 향후 연구 과제 116

참고문헌 117

Summary 123

표차례 LIST OF TABLES

[표 1-1] 건축물 단위 데이터 구축 예시(홍수) 11
[표 2-1] FWI 구성 Code 및 Index 16
[표 2-2] FBP 구성요소 18
[표 2-3] CWFIS에서 제공하는 산불 예측 및 위험도 관련 지도 18
[표 2-4] SWAT 모델 입력자료 34
[표 2-5] 재난·재해 분석 모델 및 연구 정리 42
[표 2-6] 재난·재해 분석 변수 정리 44
[표 3-1] 2017-2021년 서울시 발화요인별 화재발생 건수 51
[표 3-2] 2017-2021년 용도별 서울시 화재피해 건수 51
[표 3-3] 2016-2022년 피해요인별 서울시 홍수피해 건수 52
[표 3-4] 2016-2022년 피해금액대별 서울시 홍수피해 건수 53
[표 3-5] 2017-2021년 서울시 발화요인별 화재발생 건수 54
[표 3-6] 2017-2021년 서울시 장소별 화재발생 건수 55
[표 3-7] 화재데이터 지오코딩에 따른 오류구분 및 예시 57
[표 3-8] 화재 건축물 연계 데이터 최종 매칭 결과 59
[표 3-9] 화재 데이터 및 건축물 대장 매칭 결과 칼럼Code 및 설명 59
[표 3-10] 화재발생 건축물 데이터셋의 독립변수 67
[표 3-11] 딥러닝 기반 건축물 홍수 리스크 분석을 위한 종속변수와 독립변수 리스트 - 79
[표 3-12] 기계학습 기반 홍수 민감지역 분석을 위한 종속변수와 독립변수 80
[표 3-13] 데이터 전처리방법 87
[표 3-14] 피해액 기준 단위별 클래스 학습 88
[표 3-15] 홍수리스크 분석 모델 변수별 중요도 91
[표 4-1] 분석 모델 적용 결과: 화재 98
[표 4-2] 분석 모델 적용 결과: 홍수 101
[표 4-3] 분석 모델 적용 결과: 홍수(건축물 용도 및 구조) 102
[표 4-4] 건축물 리스크 유형 및 동 수 103
[표 4-5] 건축물 리스크 유형 107

그림차례 LIST OF FIGURES

[그림 1-1] 건축물 홍수 및 화재 발생 현황 — 2
[그림 1-2] Hazus-MH의 홍수분석 결과와 홍수 피해 건축물 지도 — 3
[그림 1-3] 건축물 화재 발생가능성 예측 결과 — 4
[그림 1-4] CNN, RNN을 이용한 홍수 발생가능성 분석 — 5
[그림 1-5] 연구의 흐름 — 8
[그림 2-1] CFFDRS 구조도 — 15
[그림 2-2] FWI 구조도 — 16
[그림 2-3] FBP 구조도 — 17
[그림 2-4] NeuroFire 구조도 — 21
[그림 2-5] CityGuard 시스템 구성도 — 22
[그림 2-6] CityGuard 인터페이스 — 22
[그림 2-7] 화재예측 모델 및 결과 — 24
[그림 2-8] 벤쿠버 시 및 뉴웨스트민스터 시의 건축물 화재위험도 — 25
[그림 2-9] 벤쿠버 시 및 뉴웨스트민스터 시의 Lift 분석 결과 — 25
[그림 2-10] 모델 활용을 통한 점검 프로그램의 효율성 향상 비율 — 26
[그림 2-11] TabNet 모델의 신경망 구조 — 27
[그림 2-12] 건축물 화재 발생 예측 의사결정나무 모델 — 28
[그림 2-13] 건축물 화재 예측에서 TabNet 모델의 변수 중요도 — 28
[그림 2-14] 건축물 화재 발생가능성 예측 결과 — 29
[그림 2-15] 강우를 고려한 도시 홍수 시뮬레이션 결과 — 30
[그림 2-16] SWMM 분석 개념도 — 31
[그림 2-17] 강우, 토양침투, 증발 등을 고려하는 SWMM의 강우 유출 분석 개념도 — 31
[그림 2-18] 저영향 개발 도입 시나리오에 따른 강우 유출량 분석 — 32
[그림 2-19] SWAT의 수문 분석 개념도 — 33
[그림 2-20] SWAT의 하위유역 구분 사례(붉은 선) — 33
[그림 2-21] SWAT의 입력 및 출력 데이터 — 34
[그림 2-22] TUFLOW의 홍수 분석 예시 — 35
[그림 2-23] TUFLOW의 홍수 입출력 데이터 예시 — 35
[그림 2-24] FLO-2D의 홍수 분석 예시 — 36
[그림 2-25] Hazus flood model의 분석 구조 — 37
[그림 2-26] Hazus flood model의 손상함수 — 38

그림차례 LIST OF FIGURES

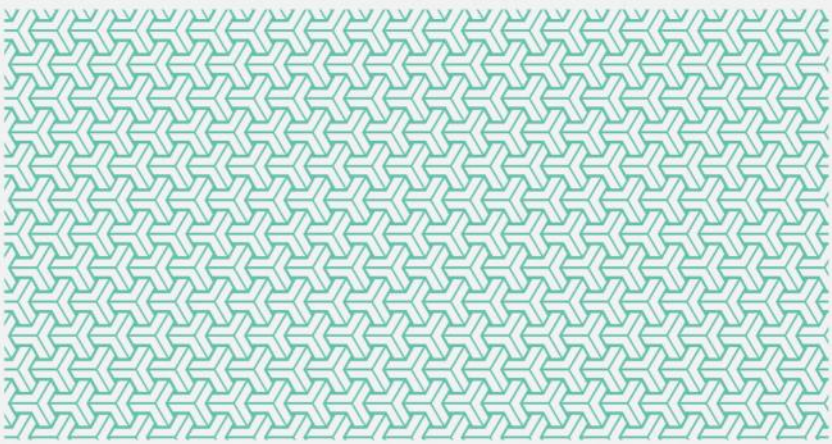
[그림 2-27] HEC-FIA 모델 입력자료 예시 — 39
[그림 2-28] HEC-FIA 모델의 피해비용 예측 모식도 — 40
[그림 2-29] Riskscape 분석 프로세스 및 모델 파이프라인 개념도 — 41
[그림 2-30] 리스크 산정 방식 — 43
[그림 2-31] 리스크 분석 모형의 개념 — 45
[그림 3-1] 건축물 화재 및 홍수 리스크 분석 모델 개발 프로세스 — 48
[그림 3-2] 피해금액대별 화재피해발생 건수 — 52
[그림 3-3] 피해금액대별 홍수피해발생 건수 — 53
[그림 3-4] 대상지 선정기준 — 54
[그림 3-5] 건축물 화재 발생 데이터 구축 프레임워크 — 55
[그림 3-6] 건축물 데이터와 화재 데이터 연계 절차 — 56
[그림 3-7] 화재 데이터의 공간 정보화 — 58
[그림 3-8] 건축물 홍수피해 데이터 구축 프레임워크 — 60
[그림 3-9] biz-gis 지오코딩툴의 지오코딩 결과 화면 예시 — 61
[그림 3-10] biz-gis 지오코딩툴의 지오코딩 결과 — 62
[그림 3-11] 홍수피해 및 건축물대장 지오코딩 결과 — 63
[그림 3-12] 홍수피해 및 건축물대장의 공간기반 연계 결과 — 64
[그림 3-13] 건축물 화재 변수 간의 상관관계 분석 — 65
[그림 3-14] 건축물 화재 리스크 분석 모델 변수별 분석 단위 — 68
[그림 3-15] 버퍼를 활용한 공간분석(400m 기준) — 68
[그림 3-16] 건축물 화재 리스크 분석 입력 변수 구축 결과 — 70
[그림 3-17] 랜덤 포레스트 모델 — 72
[그림 3-18] 건축물 화재발생 피해금액 및 발생빈도 — 74
[그림 3-19] 화재 리스크 분석 모델 혼돈행렬 — 76
[그림 3-20] 화재 리스크 분석 모델 ROC 곡선 — 77
[그림 3-21] 화재 리스크 분석 모델 변수중요도 시각화 — 78
[그림 3-22] Euclidean distance의 거리 계산 방법 — 80
[그림 3-23] 건축물 홍수 민감 지역 — 82
[그림 3-24] 홍수 민감 지도 분석 환경 변수 시각화 — 84
[그림 3-25] 건축물 홍수 피해액 분포 — 88
[그림 3-26] 피해액 기준, 클래스 별 학습 진행 결과: 100만원, 총 11개 클래스 — 89

그림차례 LIST OF FIGURES

[그림 3-27] 피해액 기준, 클래스 별 학습 진행 결과: 200만원, 총 6개 클래스 —— 89
[그림 3-28] 피해액 기준, 클래스 별 학습 진행 결과: 300만원, 총 4개 클래스 —— 90
[그림 3-29] 피해액 기준, 클래스 별 학습 진행 결과: 500만원, 총 3개 클래스 —— 90
[그림 3-30] 피해액 기준, 클래스 별 학습 진행 결과: 300만원 이하, 초과 2개 클래스 — 90
[그림 3-31] 홍수리스크 분석 모델 변수별 중요도 —— 91
[그림 4-1] 관악구 건축물 포인트 지도 —— 96
[그림 4-2] 건축물 주변 공간정보 전처리 과정 —— 97
[그림 4-3] 서울시 건축물 홍수 리스크 지도 —— 100
[그림 4-4] 관악구 건축물 홍수 리스크 지도 —— 101
[그림 4-5] 서울시 관악구 내 건축물 분포 현황 —— 103
[그림 4-6] 건축물 화재 리스크 지도 —— 104
[그림 4-7] 건축물 홍수 리스크 지도 —— 104
[그림 4-8] 건축물 홍수 및 화재 리스크 지도 —— 105

제1장 서론

Chapter. 1



1. 연구의 배경 및 목적
2. 연구의 범위 및 방법

1. 연구의 배경 및 목적

1) 연구의 배경

□ 건축물 화재 및 홍수 피해 증가

지난 수년 동안 우리는 건축물의 화재와 홍수에 의한 피해를 체감하고 있다. 이 같은 문제는 과거 몇 년 동안 발생한 대형 화재와 홍수를 통해 잘 드러난다. 특히 2021년 이천 쿠팡 물류센터 화재와 2022년 여름 홍수로 인한 강남역 침수와 반지하 침수에 의한 사망사고 발생은 국민의 건축물의 화재와 홍수로 인한 안전 불안감을 더욱 증폭시켰다. 또한 2012년부터 2021년까지의 기간 동안, 태풍과 호우로 인한 건축물 피해액이 1,600억원을 상회하는 것으로 나타나 문제의 심각성이 부각되고 있다(행정안전부 2022, p.19). 이러한 문제로 인해 건축물 안전 강화 필요성이 제기되고 있으며, 향후 문제 발생을 대비하여 적절한 대응 방안 마련이 화두에 올랐다.

건축물 홍수 피해

건축물 화재 피해

[그림 1-1] 건축물 홍수 및 화재 발생 현황

출처: 정세진. 머니투데이. https://news.mt.co.kr/mtview.php?no=2022080917011789586. 2023.03.06. 접속;
서정익. 전국매일신문. https://www.jeonmae.co.kr/news/articleView.html?idxno=896782. 2023.03.06. 접속

기후 변화가 야기할 재난 및 재해의 빈도와 강도 증가를 고려할 때, 건축물의 안전 문제는 더 시급히 다루어질 필요가 있다. 최근 IPCC (Intergovernmental Panel on Climate Change)는 1.5℃ 기후 목표 달성이 어려울 것으로 발표하였고, 기후변화의 영향과 피해가 예상했던 것 보다 강할 것으로 예측하고 있다[1]. 결국 건축 및 도시 부문에서 기후 변화로 인한 피해는 그동안 경험해온 것 보다 증가하게 될 것이고, 이에 대한 적절한 대응 방안 마련이 더욱 시급해졌다. IPCC 또한 다가올 기후변화 영향에 대한 적응방안 마련 필요성을 강조하고 있다[2].

1) IPCC. 2022b. Climate Change 2022: Mitigation of Climate Change

□ 건축물 특성을 반영한 화재 및 홍수 리스크 분석 노력

건축물에서 발생하는 재해 피해는 건축물의 노후도와 같은 건축물 특성의 영향을 받는다. 화재발생시 건축물 노후도는 건축 설비의 노후화 및 난연·불연재 미사용과 연관되어 화재 피해에 대한 취약성을 증가시킨다. 건축물의 특성은 홍수 발생에도 영향을 준다. 노후건축물의 누수, 지하층이 있는 건축물의 침수, 건축물 용도에 따른 피해규모 변화 등 건축물 특성과 홍수 취약성은 높은 상관관계가 있다. 즉, 재난과 재해에 취약한 건축물 요소들을 식별하고 분석함으로써, 향후 발생할 피해를 최소화할 수 있을 것이다.

다가오는 재해위험 속에서 건축물 안전을 확보하기 위해서는 건축물의 특성을 고려한 정량적인 리스크 분석이 필요하다. 노후도, 규모, 구조, 용도와 같은 건축물의 특성을 고려한 안전성 진단이 될 수 있도록 건축물 재해 리스크 분석 모델 개발이 필요하다. 건축물의 재해 리스크 분석 모델은 건축물 특성을 고려한 재난·재해 발생 가능성과 건축물 피해규모를 고려하여 향후 피해를 최소화하기 위한 근거자료를 제시해주어야 한다.

[그림 1-2] Hazus-MH의 홍수분석 결과와 홍수 피해 건축물 지도

출처: FEMA & HAZUS, 2018. Hazus Flood Model: FEMA Standard Operating Procedure for Hazus Flood Level 2 Analysis. p.78

2) IPCC. 2022a. AR6 Climate Change 2022: Impacts, Adaptation and Vulnerability

이미 국내·외 다양한 모델들이 건축물 특성을 반영하여 각종 재난 및 재해 리스크 분석과 예측을 시도하고 있다. 카네기 멜론 대학 프로젝트인 Metro21: Smart Cities Institute의 건축물 화재 위험도 예측모델은 건축물 규모, 유형, 가격 등을 변수로 사용한다[3]. Hazus-MH와 HEC-FIA와 같은 미국의 홍수 리스크 분석 모델은 건축물의 규모, 구조, 유형 등 특성을 활용하여 홍수 피해액을 산정한다. 이처럼 건축물의 특성은 건축물의 재난 및 재해 리스크 분석을 위한 기초적인 자료로 자리 잡고 있다.

발생: 빨간색, 미발생: 파란색

[그림 1-3] 건축물 화재 발생가능성 예측 결과

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. 건축공간연구원. p.103

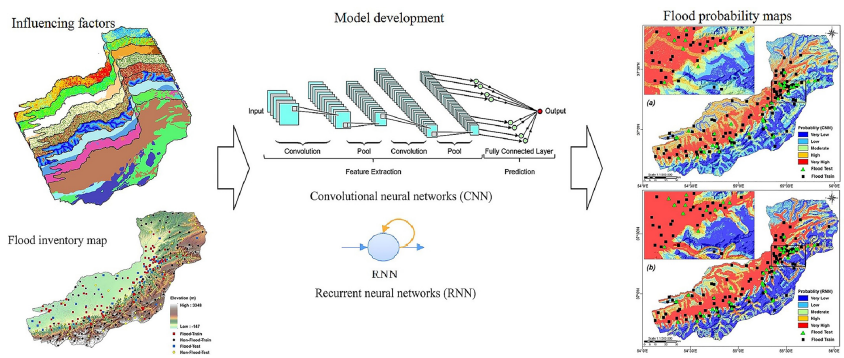
□ 제4차 산업혁명 기술에 따른 리스크 분석 정확도 증가

건축물의 홍수 및 화재 리스크를 분석하기 위한 데이터와 분석 방법론은 제4차 산업혁명과 발맞추어 진척되고 있다. 홍수와 관련해서는 수리·수문기반 모형과 같이 강수량과 물리적 시스템을 이용한 분석모델, 기존의 통계기반 모델을 발전시킨 머신러닝 및 딥러닝 모델이 활용되고 있다. 건축물의 화재 예측은 특히 화재 발생 데이터와 건축물 관련 정보들을 수집하여 통계기반의 분석이 발전하였다.

컴퓨터의 저장 및 연산능력의 향상이 리스크 분석을 위한 머신러닝과 딥러닝을

3) Metro21: Smart Cities Institute. https://www.cmu.edu/metro21/projects/smart-infrastructure/fire-risk-analysis.html. 2023.05.22. 접속;
Madaio, Michael A.. 2018. Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections. Metro21. 을 참고하여 재작성

가능하게 만들었고, 모델의 정확도 개선에 많은 기여를 하였다. 재난 및 재해 리스크 분석 분야에서 의사결정나무, 랜덤 포레스트, 서포트 벡터 머신과 같은 머신러닝 알고리즘들을 기반으로 모델 정확도를 점진적으로 증가시켜왔다. 뿐만 아니라 RNN (Recurrent Neural Network)과 LSTM (Long Short Term Memory), CNN (Convolutional neural network) 등의 딥러닝 기반 기술의 발전이 모델의 정확도를 비약적으로 증가시켰다.



[그림 1-4] CNN, RNN을 이용한 홍수 발생가능성 분석

출처: Panahi *et al.*. 2021. https://www.sciencedirect.com/science/article/pii/S1674987120302085. 2023.05.22. 접속

리스크 분석 정확도의 증가에는 가용 데이터의 품질 향상도 기여한 바가 크다. IoT (Internet of Things), ICT (Information and Communications Technology) 기반 데이터 수집능력 발전으로 가용 데이터의 범위 및 용량이 늘어났다. IoT와 ICT 기술은 거대한 양의 데이터를 실시간으로 수집, 저장, 전송해주며, 클라우드 서버와 같은 시스템을 통해 공유되기도 한다. 이 같은 기술의 발달로 인해 머신러닝 및 딥러닝 분석을 위해 사용되는 데이터 품질이 고도화되었고, 리스크 분석의 정확도 또한 높아지고 있다.

□ 「건축물관리법」 등 관계 법령에 따른 건축물 정기점검 합리화 기반 마련 필요

국내에서도 화재와 홍수의 리스크가 높은 건축물을 식별하고, 건축물 리스크 상승에 기여하는 요인을 판단하여 이를 중심으로 건축물의 안전 관리를 합리화할 필요가 있다. 정부와 지방자치단체는 건축물의 안전을 확보하기 위하여 「건축물관리법」, 「소방시설 설치 및 관리에 관한 법률」, 「시설물의 안전 및 유지관리에 관한 특별법」 등에 따른 건축물 점검을 법적 의무로 시행하고 있다. 이러한 점검은 다중이용건축물, 일정 규모 이상 건축물 등에 대하여 의무적으로 시행하고 있어 점검의 실효성에 대하여 지속적인 문제가 제기되어 왔다.

따라서 화재, 수해 등 재난에 대하여 위험한 건축물을 판별하고 이를 점검하는 방식에 대하여 고민할 필요가 있다. 과거에는 데이터 구득과 기술적 문제로 개별 건축물의 위험을 측정하는 것이 어려웠지만, 이제 건축물의 속성정보와 더불어 다양한 국가 공공데이터와, 머신러닝 및 딥러닝 등 최신 기술의 활용이 가능한 만큼 이를 적극적으로 활용할 필요가 있다.

2) 연구의 목적 및 기대효과

□ 연구의 목적

본 연구의 주요 목적은 네 가지로 설정하였다. 첫 번째 목적은 건축물 단위의 홍수 및 화재 리스크 분석 모델을 개발하는 것이다. 두 번째 목적은 서로 다른 데이터 연계 방법을 실증하고, 향후 데이터 연계를 위한 기초를 확립하는 것이다. 세 번째 목적은 홍수 및 화재 대응이 필요한 건축물을 도출하고 모델의 정확도를 검증하는 것이다. 마지막으로 건축물 화재 및 홍수 안전 제고를 위한 건축물 관리방안을 제안하고자 한다.

□ 기대효과

본 연구를 통해 건축물 관리 정책과 유관 제도의 개선과 신규 사업 발굴에 기여할 수 있다. 화재 및 홍수 발생 건축물들의 유형과 특성을 도출함으로써 「건축물관리법」에 따른 점검(정기, 긴급, 소규모) 항목을 개선할 수 있으며, 세부적으로는 기존 건축물관리 정기 점검 대상인 '5,000㎡이상 다중이용건축물', '1,000㎡이상 준다중이용건축물 중 지자체 조례로 정한 건축물', '「공동주택관리법」에 따른 의무관리대상이 아닌 건축물' 등의 기준도 개선할 수 있을 것으로 기대된다. 뿐만 아니라 리스크가 높은 건축물을 도출함으로써 정부, 지자체의 화재 및 풍수해 예방 정책 수립 시 기초자료로 활용할 수 있다.

건축물관리법 시행령

제8조(정기점검 대상 건축물 등) ① 법 제13조제1항에서 "다중이용 건축물 등 대통령령으로 정하는 건축물"이란 다음 각 호의 건축물을 말한다. 다만, 「학교안전사고 예방 및 보상에 관한 법률」 제2조제1호에 따른 학교, 「공동주택관리법」 제2조제1항제2호에 따른 의무관리대상 공동주택, 「유통산업발전법」 제2조제3호 · 제4호에 따른 대규모점포 · 준대규모점포 및 정기점검을 실시해야 하는 날부터 3년 이내에 「공동주택관리법」 제34조제2호에 따라 소규모 공동주택 안전관리를 실시한 공동주택은 제외한다.

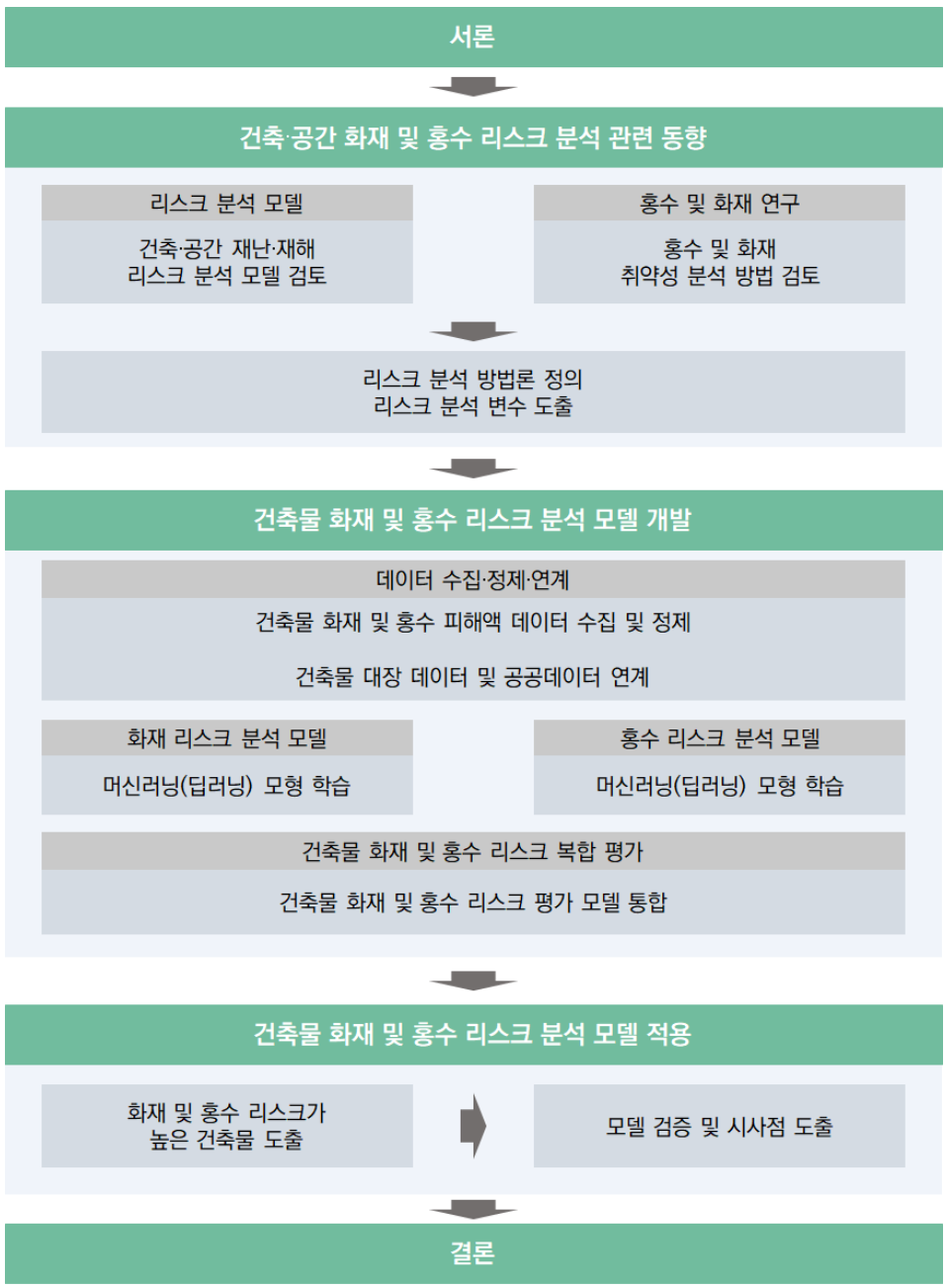
1. 「다중이용업소의 안전관리에 관한 특별법」에 따른 다중이용업소가 있는 건축물로서 특별자치시 · 특별자치도 · 시 · 군 · 구(자치구를 말한다)의 조례(이하 "시 · 군 · 구 조례"라 한다)로 정하는 건축물
2. 「집합건물의 소유 및 관리에 관한 법률」을 적용받는 건축물로서 연면적 3천제곱미터 이상인 건축물
3. 「건축법 시행령」 제2조제17호에 따른 다중이용 건축물
4. 「건축법 시행령」 제2조제17호의2에 따른 준다중이용 건축물로서 같은 조 제18호에 따른 특수구조 건축물에 해당하는 건축물

(생략)

출처 : 건축물관리법 시행령. 대통령령 제33632호, 2023. 7. 11.. 일부개정

2. 연구의 범위 및 방법

1) 연구의 흐름



[그림 1-5] 연구의 흐름
출처: 연구진 작성

2) 연구의 범위

□ 공간적 범위

본 연구의 공간적 범위는 다음의 세 가지 조건을 검토하였다.

첫째, 건축물 단위의 화재 및 홍수 피해 발생 데이터가 잘 구축되어 있는 지역을 연구의 대상지로 하였다. 이는 머신러닝과 딥러닝에 필요한 입력 데이터 확보 및 향후 분석 모델의 정확도 제고와 모형의 검증을 위해 필수적이다.

둘째, 분석을 위한 데이터 모집단이 충분한 지역을 선택하였다. 분석의 정확도 제고를 위해 건축물 피해데이터와 건축물 특성정보 이외의 다양한 데이터가 활용될 필요가 있다.

셋째, 데이터 접근성을 고려하였다. 데이터 구득이 가능하며, 모델 개발 및 검증이 용이한 지역을 선택하였다. 연구 수행에 있어 데이터 접근성은 중요한 요소로 작용하며, 이를 통해 모델 개발과 검증 과정의 신뢰도를 높일 수 있다.

이 같은 조건을 충족하는 지역으로 서울시를 공간적 범위로 설정하였다. 서울시는 건축물 화재 및 홍수 피해 데이터의 구축 및 관리가 잘 되어있으며, 이외에도 다양한 공공데이터를 구축하고 있고, 이 같은 데이터에 대한 접근이 용이하다.

□ 내용적 범위

본 연구의 내용적 범위는 크게 세 가지로 구분할 수 있다.

첫째는 홍수 및 화재 발생 데이터와 건축행정 데이터 및 공공데이터의 연계이다. 이는 연구의 기본적인 출발점으로, 두 가지 이상의 이기종 데이터를 연결하고 이를 분석 가능한 형태로 만드는 과정이다. 데이터 수집, 정제 및 전처리 작업으로 시작되며, 이 과정에서 매칭키 도출과 연계방안을 모색 등을 주로 연구할 것이다. 뿐만 아니라, 향후 데이터 연계 시 발생할 수 있는 문제점을 파악하고, 이를 해결하는 방안을 모색하고자한다. 이 과정을 통해 구축된 데이터를 활용하여 효과적인 건축물 화재 및 홍수 리스크 분석을 수행하고자 한다.

두 번째는 건축물의 홍수 및 화재 리스크를 분석하는 모델을 개발하고 검증하는 과정이다. 첫 단계에서 구축한 데이터가 두 번째 단계에서 머신러닝 및 딥러닝 모델 학습과 검증에 사용된다. 이 과정을 통해 건축물 홍수 및 화재 리스크 분석 모델을 개발하며, 모델의 정확성과 신뢰성을 검증하고, 모델의 개선 및 개선 방안도 함께 제시하고자 한다.

마지막으로, 개발된 모델의 활용 방안을 검토하고자 한다. 모델 적용을 통한 홍수 및 화재 대응이 필요한 건축물의 도출, 그리고 해당 결과를 활용한 건축물의 대응 전략 제시를 목표로 한다. 이는 본 연구가 단순히 사례연구에 그치지 않고, 향후 정책에 활용 가능한 형태로 전환될 수 있도록 확인하는 과정이다.

3) 연구의 방법

모델 개발은 홍수와 화재 발생 가능성과 이에 따른 잠재적 손상을 고려한 건축물의 리스크를 도출할 수 있도록 한다. 개발 모델은 제4차 산업혁명 기술 발전을 고려하여 관련 빅데이터, 머신러닝, 딥러닝 기술을 활용하며 이를 통해 기존 분석 모델 대비 높은 정확도를 확보 하고, 홍수 및 화재 피해발생 데이터, 건물 특성 정보, 그리고 다양한 국가 공공데이터를 종합적으로 활용하여 홍수와 화재 리스크를 분석한다.

위험 건축물 도출 및 모델 검증은 향후 본 연구에서 개발한 모델이 국가 정책에 활용될 수 있도록 신뢰도를 제고하는 과정이다. 개발된 분석 모델을 이용하여 홍수와 화재 위험도가 높은 건축물을 도출하고, 이 같은 결과를 이용하여 데이터기반 정책 마련이 가능한지 확인하고자 한다. 이 과정을 통해 모델의 잠재적인 한계와 문제점을 발견하고, 이를 개선함으로써 보다 효과적인 리스크 분석 모델을 만들 수 있을 것으로 기대한다.

데이터 연계 방법 실증을 위해 화재피해 데이터와 홍수피해 데이터는 각각 건축물 대장과 서로 다른 방식으로 연계 한다. 본 연구에서 두 가지 서로 다른 데이터 연계 방법을 적용해봄으로써 향후 구축될 다양한 이기종 데이터의 연계방안 마련의 지침이 될 수 있다.

□ 문헌조사

- 건축물 홍수 및 화재 관련 선행연구 조사
- 리스크 분석 관련 선행연구 및 기술 동향 조사
  - HAZUS-MH, FLEMO, VESTAP 등 리스크 분석모델 조사
  - 다중회귀분석 모형, 인공신경망 모형 등 기계학습모형 조사

□ 건축물 단위 데이터 구축

- 건축물 형상정보와 속성정보 연계를 통한 건축물 공간정보 구축
  - 건축물 형상정보 및 속성정보 연계방안 도출

- 품질 저하를 최소화하는 건축물 단위 데이터 연계

- 홍수 및 화재 피해 분석을 위한 속성정보 생산 및 건축물 단위 연계
  - 건축물 홍수 및 화재 피해에 영향을 미치는 변수 도출
  - 관련 데이터 수집 및 가공을 통한 데이터베이스 구축
  - 피해정보 및 건축물 데이터 등 이기종 데이터 연계
- 건축물단위 홍수 및 화재 피해액 정보 연계

[표 1-1] 건축물 단위 데이터 구축 예시(홍수)

홍수 피해액(천원)	구조	연면적 (㎡)	…	사용승인일
13,391	경량철골구조	7,503	…	1992
971	블록구조	1,530	…	1993
3,669	철근콘크리트구조	2,030	…	2004
374	철근콘크리트구조	504	…	2021
…	…	…	…	…
5,478	일반목구조	280	…	1984

출처: 연구진 작성

□ 건축물 홍수 및 화재 리스크 분석 모델 개발

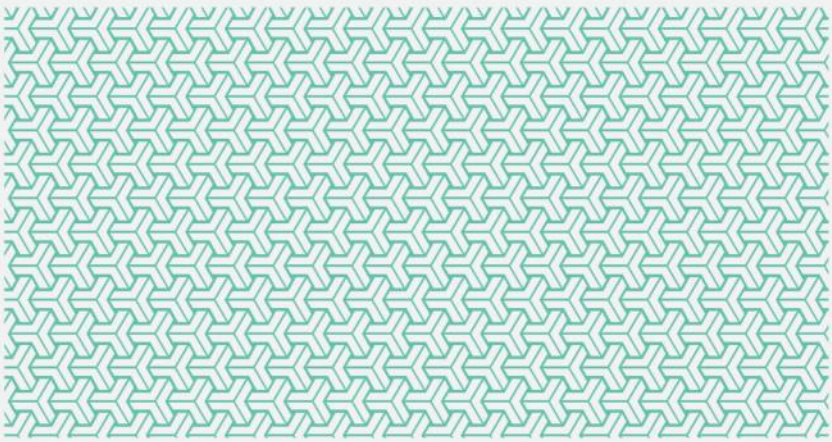
- 홍수 및 화재 리스크 분석 모델 연구
  - 홍수 및 화재 발생 확률 및 손상함수 결정을 위한 발생 메커니즘 조사
  - 홍수 및 화재 발생과 건축물 피해 관계 분석
- 홍수 및 화재 리스크 분석 모델 개발
  - 독립변수와 종속변수 선정
  - 기계학습 및 딥러닝 등 리스크 분석 모형 선정
  - 홍수 및 화재 리스크 분석 모델 개발

□ 전문가 활용

- 홍수 및 화재 전문가 대상 리스크 분석을 위한 주요변수 도출 및 검증
- 분석모형 선정 및 검증 방법 검토
- 데이터 및 개발 모델 활용 방안 검토

Chapter.2

제2장 화재 및 홍수 리스크 분석 관련 동향



1. 화재 피해 및 리스크 분석
2. 홍수 피해 및 리스크 분석
3. 소결

1. 화재 피해 및 리스크 분석

① Canadian Wildland Fire Information System[4)]

Canadian Wildland Fire Information System (CWFIS)는 캐나다의 화재 관리 및 모니터링 시스템이다. CWFIS는 캐나다 Natural Resources Canada가 운영하며, 캐나다 내 모든 산불에 대한 정보를 수집하고 보급하는 역할을 한다. CWFIS는 다양한 데이터를 사용하여 산불 위험, 화재 확산 패턴, 날씨 조건 등을 분석하여 산불 관리자와 대중에게 중요한 정보를 제공하고 있다. 이 정보는 산불 예방, 조기 경보, 화재 진압, 자원 할당 등을 지원하고 있다.

CWFIS 핵심 기능

- 산불 위험 예보 : CWFIS는 산불 위험 등급을 예측하고 예보한다. 이를 통해 지역의 산불 위험 수준을 파악하고 조치를 취할 수 있다.
- 산불 모니터링 : CWFIS는 실시간으로 산불에 대한 정보를 모니터링한다. 이를 통해 산불의 크기, 위치, 확산 경로 등을 추적하고 조치를 취할 수 있도록 하고 있다.
- 날씨 데이터 분석 : CWFIS는 날씨 데이터를 활용하여 산불 위험을 예측한다. 바람의 강도와 방향, 기온, 습도 등의 정보를 분석하여 산불의 확산 가능성을 예측한다.
- GIS 매핑 도구 : CWFIS는 지리 정보 시스템(GIS) 매핑 도구를 제공하여 산불 관련 데이터를 시각화하여 제공하고 있다. 지도 위에 산불 위치, 경계, 피해 지역 등을 표시하여 산불 상황을 시각적으로 이해할 수 있다.

출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background. 2023.06.15. 접속

CWFIS는 산불예측을 위해 캐나다 산불 위험 등급시스템을 중심으로 하위 시스템을 연계하여 운영하고 있으며, 자세한 설명은 아래와 같다.

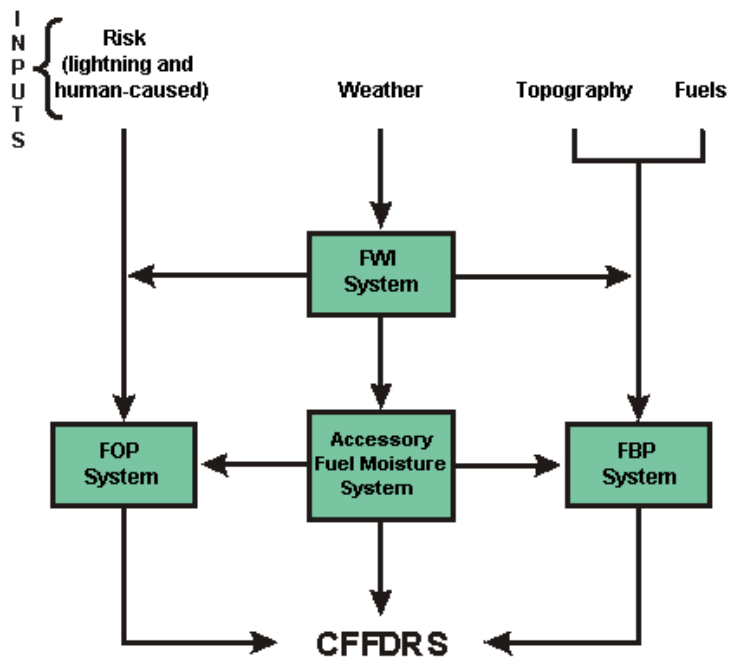
☐ Canadian Forest Fire Danger Rating System (CFFDRS)[5)]

캐나다 산불 위험 등급 시스템(CFFDRS)은 캐나다의 산불 위험을 평가하는 국가 시스템이다. 산불 위험은 화재 환경에서 발화의 용이성, 통제의 어려움 등 다양한 요인을 표현하기 위해 사용되는 일반적인 용어다. 화재 위험 등급 시스템은 다양한 화재 관리 활동에서 지침으로 사용되는 화재 잠재력의 질적 또는 숫자 지수를 생성한다. CFFDRS는 1968년부터 개발 중이며 현재 캐나다 산불 기상 지수(FWI) 시스템과 캐나다 산불 행동 예측(FBP) 시스템의 두 가지 하위 시스템이 캐나다와 국제적으로 광범위하게 사용되고 있다.

4) Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/home. 2023.06.15. 접속

5) Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fdr. 2023.06.15. 접속

CFFDRS 구성요소인 위험, 날씨, 연료 및 지형은 화재 날씨, 화재 발생 및 화재 행동을 예측하는 데 필요한 입력 정보다. 연료 수분 모델(Accessory Fuel Moisture System)은 현재 다양한 캐나다 숲 유형을 위해 개발 중에 있다.



[그림 2-1] CFFDRS 구조도

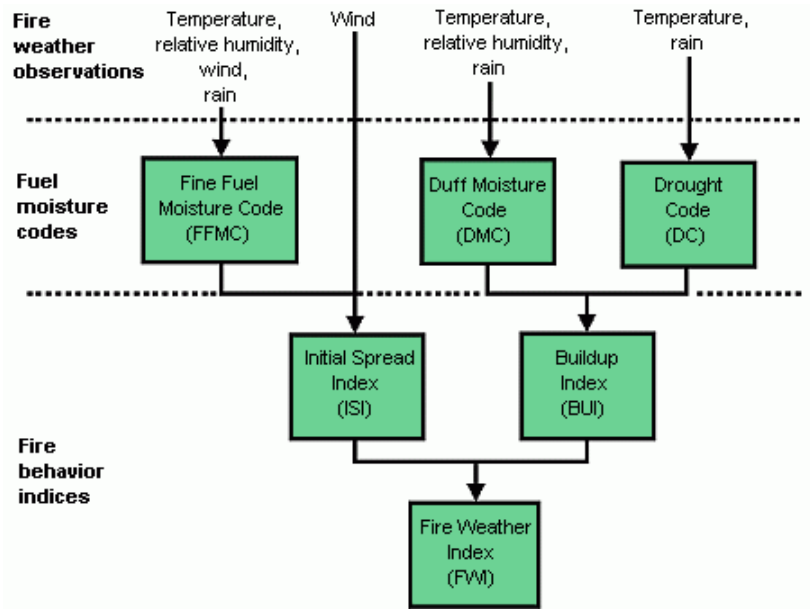
출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fdr. 2023.06.15. 접속

□ Canidian Forest Fire Weather Index (FWI) System[6)]

캐나다 산불 기상 지수(FWI) 시스템은 연료 수분 및 기상 조건이 화재에 미치는 영향을 설명하는 6가지 구성 요소로 구성된다. 처음 세 가지 구성 요소는 숲 바닥과 기타 죽은 유기물의 수분 함량에 대한 수치 등급인 연료 수분 코드이며, 수분 함량이 감소함에 따라 값이 증가한다. 연료 수분 코드는 쓰레기 및 기타 미세 연료, 중간 정도 깊이의 느슨하게 압축된 유기 층, 그리고 깊고 압축된 유기 층으로 구성된다. 나머지 세 가지 구성 요소는 화재 확산 속도, 연소에 사용 가능한 연료, 그리고 전방 화재 강도를 나타내는 화재 행동 지수이다. 이 세 가지 값은 화재 위험이 증가함에 따라 증가한다.

6) Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi. 2023.06.15. 접속

FWI 계산은 온도, 상대 습도, 풍속 및 24시간 강수량의 연속적인 일일 관측치를 기반으로 한다. 6개의 표준 성분은 자연발생 화재의 상대적 잠재력에 대한 수치 등급을 제공한다.



[그림 2-2] FWI 구조도

출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi. 2023.06.15. 접속

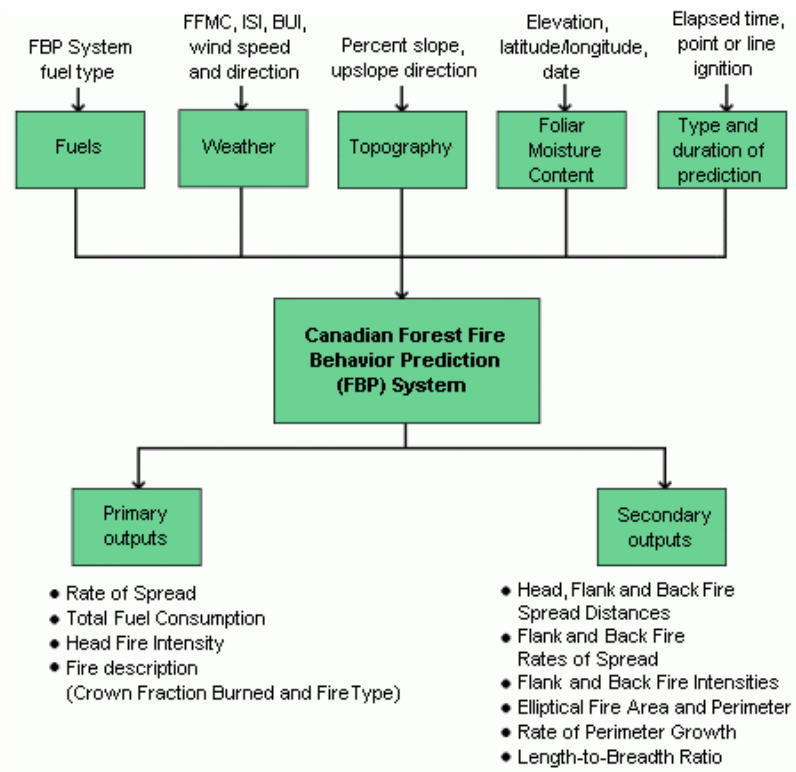
[표 2-1] FWI 구성 Code 및 Index

구분	설명
Fine Fuel Moisture Code (미세연료 수분 코드)	- 쓰레기 및 기타 경화 미세 연료의 수분 함량에 대한 수치 등급으로 상대적인 점화 용이성과 미세 연료의 가연성을 나타내는 지표
Duff Moisture Code (더프 수분 코드)	- 적당한 깊이의 느슨하게 압축된 유기 층의 평균 수분 함량에 대한 수치 등급으로, 중간 크기의 더프 층과 중간 크기의 목질 재료에서 연료 소비량을 나타냄
Drought Code (가뭄 코드)	- 깊고 압축된 유기층의 평균 수분 함량에 대한 수치 등급으로, 삼림 연료에 대한 계절적 가뭄 영향과 깊은 더프 층과 큰 통나무의 연기 양을 나타내는 유용한 지표
Initial Spread Index (초기 확산 지수)	- 예상되는 화재 확산 속도에 대한 수치 등급 - 풍속과 FFMC를 기반으로 하며, 다른 FWI 시스템 구성 요소와 마찬가지로 ISI는 연료 유형을 고려하지 않음 - 실제 확산 속도는 동일한 ISI에서 연료 유형에 따라 다름
Buildup Index (발달 지수)	- 연소에 사용할 수 있는 총 연료량에 대한 수치 등급 - DMC와 DC를 기반으로 하며 BUI는 일반적으로 DMC 값의 2배 미만이며, DMC 층의 수분은 사용 가능한 연료의 깊은 곳에서 물질이 연소되는 것을 방지하는 데 도움이 될 것으로 예상

출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi. 2023.06.15. 접속

☐ Canadian Forest Fire Behavior Prediction (FBP) System[7)]

캐나다 산불 행동 예측(FBP) 시스템은 잠재적인 산불 확산 속도, 연료 소비 및 화재 강도에 대한 정량적 추정치와 화재 설명을 제공한다. 타원형 화재 성장 모델의 도움을 받아 화재 면적, 화재발생 주변지역, 주변지역으로의 확산률 및 화재의 진행 방향, 측면 및 후면에서의 화재 거동에 대한 추정치를 제공한다.



[그림 2-3] FBP 구조도

출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fbp. 2023.06.15. 접속

7) Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fbp. 2023.06.15. 접속

[표 2-2] FBP 구성요소

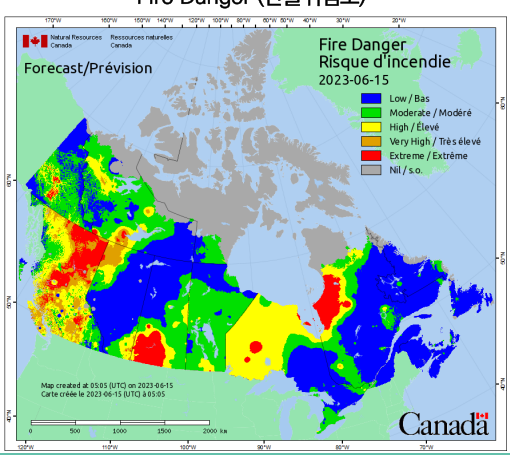
구분	설명
Rate of Spread (확산 속도)	- 화재의 전면(불이 가장 빠르게 이동하는 곳)에서 예상되는 화재 속도이며 발로와 점멸을 모두 고려 - 분당 미터 단위로 측정되며 연료 유형, 초기 확산 지수, 빌드업 지수 및 낙엽수의 현상학적 상태(잎이 없는 상태 또는 녹색), 침엽수의 높이, 풀의 경화율과 같은 몇 가지 연료별 매개 변수를 기반으로 함
Total Fuel Consumption (총 연료 소비량)	- 숲 바닥과 나무 꼭대기에서 화재로 인해 소비되는 연료의 예상 무게로, 지면 표면의 제곱미터 당 킬로그램으로 측정되며, 나뭇잎 수분 함량, 표면 연료 소비 및 확산 속도를 기준으로 함
Head Fire Intensity (화재진행 강도)	- 화재의 전면에서 발생하는 화재의 예측된 강도 또는 에너지 출력 - 소방 관리자들이 화재 통제의 어려움을 추정하고 적절한 진압 방법을 선택하는 표준 값 중 하나로 화재 현장의 미터당 킬로와트로 측정되며 확산 속도와 총 연료 소비량을 기준으로 함
Crown Fraction Burned (상위 부분 연소)	- 화재로 인해 소모되는 나무 상위부분의 예측 부분으로 빌드업 지수, 나뭇잎 수분 함량, 표면 연료 소비 및 확산 속도를 기반으로 함
Fire Type (화재 유형)	- 연소된 크라운 분율(CFB)을 기반으로 하며, CFB가 0.1(10%)보다 작으면 표면 화재이고, CFB가 0.9(90%) 이상이면 화재는 연속 상위 화재, CFB가 0.1과 0.9 사이이면 화재는 간헐적인 상위 화재임

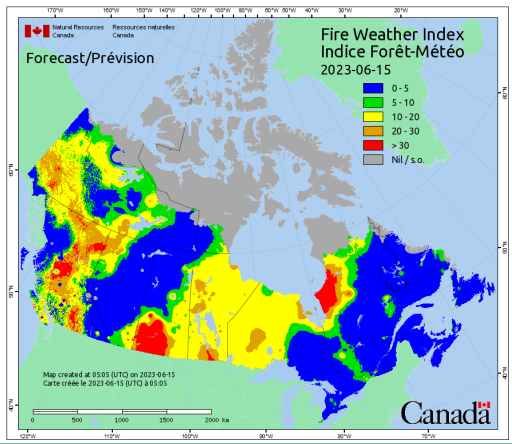
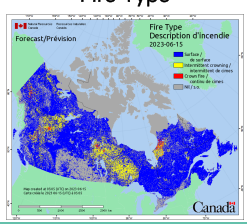
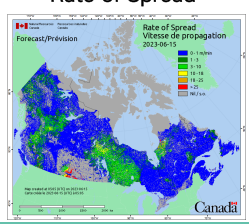
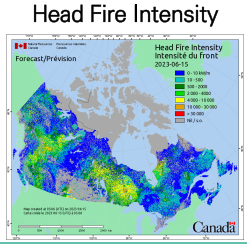
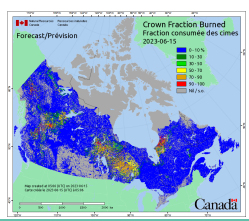
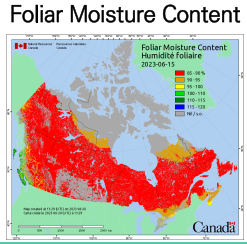
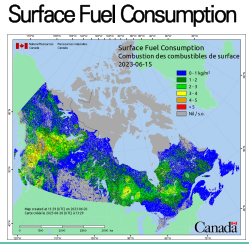
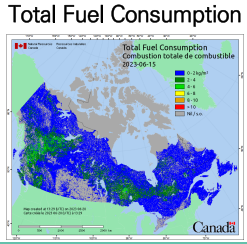
출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/background/summary/fbp. 2023.06.15. 접속

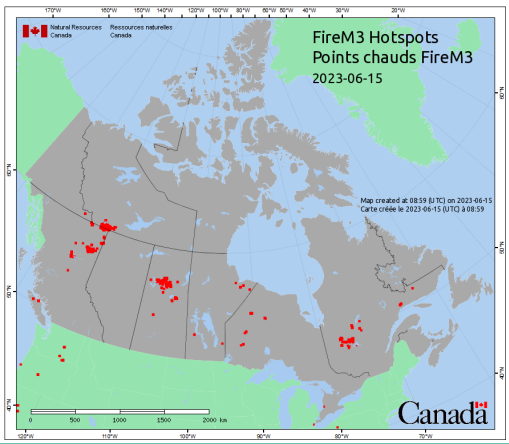
□ CWFIS에서 제공하는 지도

CWFIS는 위에서 설명한 다양한 정보 및 지표를 바탕으로 국가적 차원에서 산불 위험도, 산불예측, 산불 발생 시 산불진행 예측 등에 관한 지도를 실시간으로 홈페이지를 통해 발표하고 있다.

[표 2-3] CWFIS에서 제공하는 산불 예측 및 위험도 관련 지도

지도	설명
Fire Danger (산불위험도) 	나무에 불을 붙이는 것이 얼마나 쉬운지, 불이 얼마나 통제하기 어려울 수 있는지, 그리고 불이 얼마나 많은 피해를 줄 수 있는지에 대한 상대적 지수

지도	설명
Fire Weather Index (산불 기상 지수도) 	화재 강도의 수치를 나타내는 지도로 캐나다의 삼림 지역 전체에서 화재 위험의 일반적인 지표로 사용
Fire Behavior Maps (산불 행동 지도) Fire Type  Rate of Spread  Head Fire Intensity  Crown Fraction Burned  Foliar Moisture Content  Surface Fuel Consumption  Total Fuel Consumption 	산불 행동 지도는 Fire Type, Rate of Spread, Head Fire Intensity, Crown Fraction Burned, Foliar Moisture Content, Surface Fuel Consumption, Total Fuel Consumption 총 7가지 지표에 대한 지도를 제공하고 있음

지도	설명
Fire M3 Hotspots 	적외선 강도가 높은 위성 이미지 픽셀로, 열원을 나타냄 산업을 통해 발생한 열원 핫스팟은 제거되고, 표기된 핫스팟은 숲, 잔디, 농경지 또는 벌목 잔해에 있을 수 있는 초목 화재를 나타냄 화재가 너무 작거나 구름이 위성의 지상 시야를 가리기 때문에 위성 이미지에서 모든 화재를 식별할 수 있는 것은 아님

출처: Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/maps/fw;
https://cwfis.cfs.nrcan.gc.ca/maps/fw?type=fw;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=ft&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=ros&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=hfi&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=cfb&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=fmc&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=sfc&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fb?type=tfc&year=2023&month=6&day=15;
https://cwfis.cfs.nrcan.gc.ca/maps/fm3?type=tri. 2023.06.15. 검색

② 도시 및 건축물 단위 화재 예측 선행연구

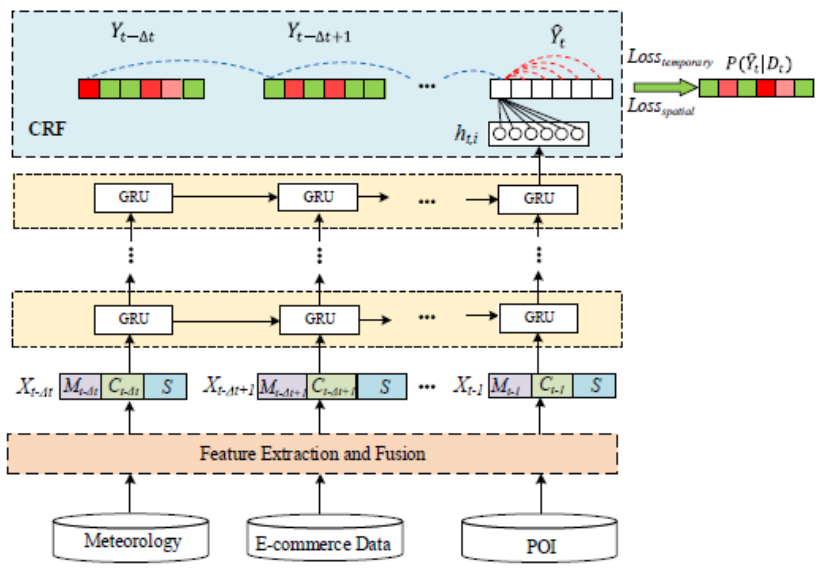
□ 도시 단위 화재 예측[8)]

Wang *et al.* (2019)는 도시단위에서의 화재예측을 위해 일명 NeuroFire라는 머신러닝 모델을 개발하여 화재를 예측하였다. 화재발생 예측을 위한 변수 범위는 크게 두 가지로 시간적 효과(Temporal Effect)와 공간적 요인(Spatial Factors)으로 구분하였다. 시간적 효과에서는 내재요인(과거 화재 발생)과 외부적요인(온도, 습도, 전자상거래 주문 건 수)과의 상관관계를 분석하였으며, 즉각적으로 화재위험 확률을 높이는 요인과 누적으로 화재위험 확률을 높이는 요인을 구분하여 분석하였다.

또한, 공간적 요인의 경우 지역적 공간 속성(POI (Point of Interest), 인간활동, 지역인구 등)과 국지적 특성을 통합하여 분석을 수행하였다.

8) Wang *et al.*. 2019. CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach.

연구수행 방법은 우선 GRU-CRF (Gate Recurrent Unit-Conditional Random Field)를 활용하며, GRU를 통해 도시 데이터의 시간적 표현을 학습하고, 이를 CRF를 통해 화재 위험 순서에 통합하는 NeuroFire라는 도시 광역 화재 예측 모델을 제안한다. 다음으로 그리드 간의 공간적 종속성을 모델링하기 위해 공간-시간 손실 함수를 설계한 후 실제 도시를 선정하여 모델의 당위성을 입증하였다.

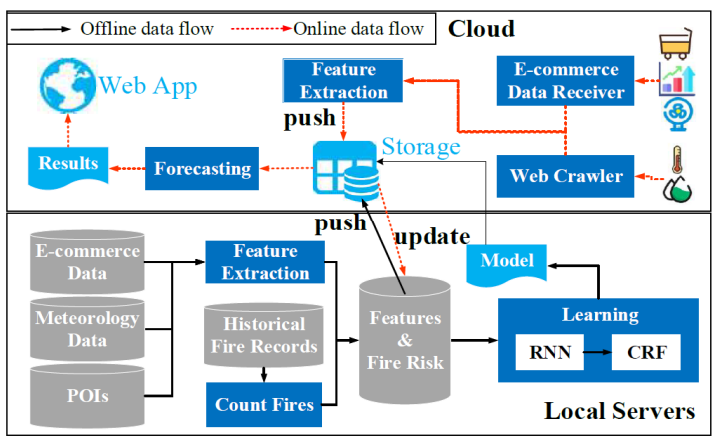


[그림 2-4] NeuroFire 구조도
출처: Wang *et al.*. 2019. p.8.

데이터 분석을 통해, 화재 위험은 시간적 관점에서 내부 시간적 영향과 외부 시간적 영향을 받는다는 것을 알 수 있다. 내부 시간적 영향은 과거 화재 위험과의 상관 관계를 나타내며, 즉각적인 영향과 지연된 영향에 의해 화재 위험에 영향을 미친다. 현재 화재 위험을 예측하기 위해, GRU-CRF를 사용하여 이러한 두 가지 요소를 결합한다. 구체적으로, 먼저 GRU를 사용하여 내부 시간적 영향의 시간적 표현을 학습하고, CRF를 사용하여 내부 시간적 영향과 시간적 표현을 결합한다.

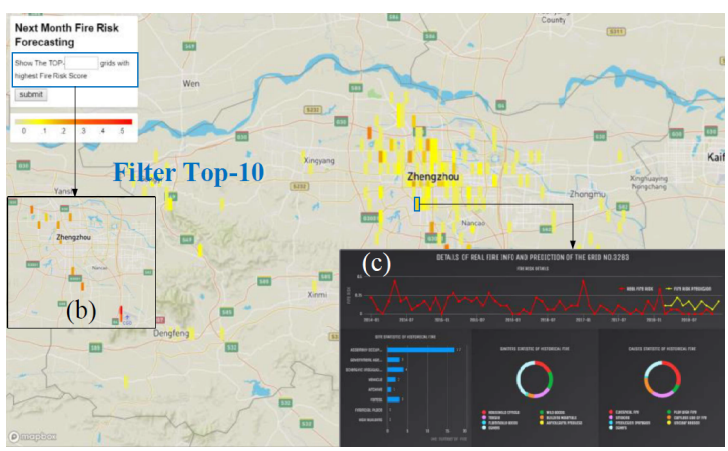
시간적인 화재 예측은 각 구역의 화재 위험(확률)을 계산하지만, 구역별 화재 위험은 독립적이기 때문에 공간적 요소에 대한 분석이 필요하다. 공간적 종속성을 학습하기 위해, 구역 간의 화재 위험 차이를 고려해야 하며, 이 문제는 추천 시스템의 순위 문제와 유사하다. 각 구역에 대해 다른 구역은 두 개의 층위로 분류될 수 있다. 하나의 층위에는 구역보다 더 큰 화재 위험이 있는 구역이 포함되고, 다른 층위에는 나머지 구역들이 포함된다.

연구결과 NeuroFire모델은 다른 비교 모델에 비해 화재발생 예측력이 뛰어났으며, 연구자는 이를 통해 CityGuard라는 시스템을 제안하고 있다. CityGuard는 로컬서버와 클라우드 서버로 구성된다. 로컬서버는 각 구역에 대해 일정 기간의 시간적 및 공간적 특성을 추출하고 구성 모듈 특성 맞춰 형식화된 특성을 획득하여 로컬에 저장하고, 과거 화재 기록을 수집하고, 계산 모듈을 적용하여 각 구역의 화재 위험 데이터를 얻은 다음 로컬에 저장한다. 또한 화재 위험 데이터와 특성을 사용하여 예측 모델을 훈련시키고, 훈련된 모델을 클라우드에 업로드한다. 클라우드 서버는 지속적으로 기상 데이터를 수집하고 월별 전자 상거래 데이터를 수신한다. 클라우드 서버에서는 훈련된 모델과 새로운 특성 데이터를 사용하여 도시의 모든 구역에 대한 화재 위험을 계산하고, 웹 서비스를 통해 예측된 화재 위험을 시각화하는 앱 서비스를 공유한다.



[그림 2-5] CityGuard 시스템 구성도

출처: Wang *et al.*. 2019. p.15



[그림 2-6] CityGuard 인터페이스

출처: Wang *et al.*. 2019. p.15

□ 건축물 단위 화재예측[9)]

McGuinness *et al.* (2019)는 기존 건축물 단위 화재예측 모형인 뉴욕시 데이터 분석국(MODA: Mayor's Office of Data Analytics)의 모델, 조지아 공과대학과 애틀란타 소방구조국이 함께 만든 Firebird 모델, 조지아 공과대학과 피츠버그 소방국이 함께 만든 Metro21 모델에 기반하여 캐나다 로어 메인랜드 지역의 벤쿠버 시와 뉴웨스트민스터 시를 대상으로 건축물 화재예측 모델을 개발하였다.

각 모델에서 활용한 데이터를 기반으로 화재사건데이터, 화재검사데이터, 재산평가데이터, 인구조사데이터(인구통계 및 사회경제적 데이터), 지적 데이터, 311 데이터(공중위생 데이터 포함), 범죄 데이터, 가용가능한 자산 데이터를 구축하였다. 특히 건축물 관련 속성데이터는 지상 및 지하 층 수, 점유 및 사용, 소화시스템 설치 여부가 포함되었다. 또한 공공데이터 중 요양원, 주거 및 임대 주택, 용도지역 및 토지이용, 주소 및 범죄데이터 등을 활용하였다.

분석 데이터 셋 구축은 건축물 수준, 주소 수준, 지적 수준으로 구분하였다. 화재검사 및 사건 데이터는 주소 수준으로 구축하고, 다른 데이터는 지적 ID를 설정하여 결합하였으며, 동일한 ID 내 다수 데이터가 있는 경우 합산하였다. 지적 수준에서 결합되지 못한 데이터는 제한하였으며, 지적 수준보다 더 큰 범위로 제공하는 데이터는 블록 수준으로 매칭하였다.

벤쿠버 시 및 뉴웨스트민서터 시 대상의 화재예측 모델 개발을 위한 빅데이터 분석은 학습 데이터셋으로 2014~2016데이터를 활용하고, 테스트 데이터셋은 2017년 데이터를 활용하였으며, 2013년 데이터는 특성 선택에 사용되었다. 모델의 성능은 우선적으로 다음의 세가지 지표를 사용하여 설명하였다.

건축물화재 예측 모델 지표

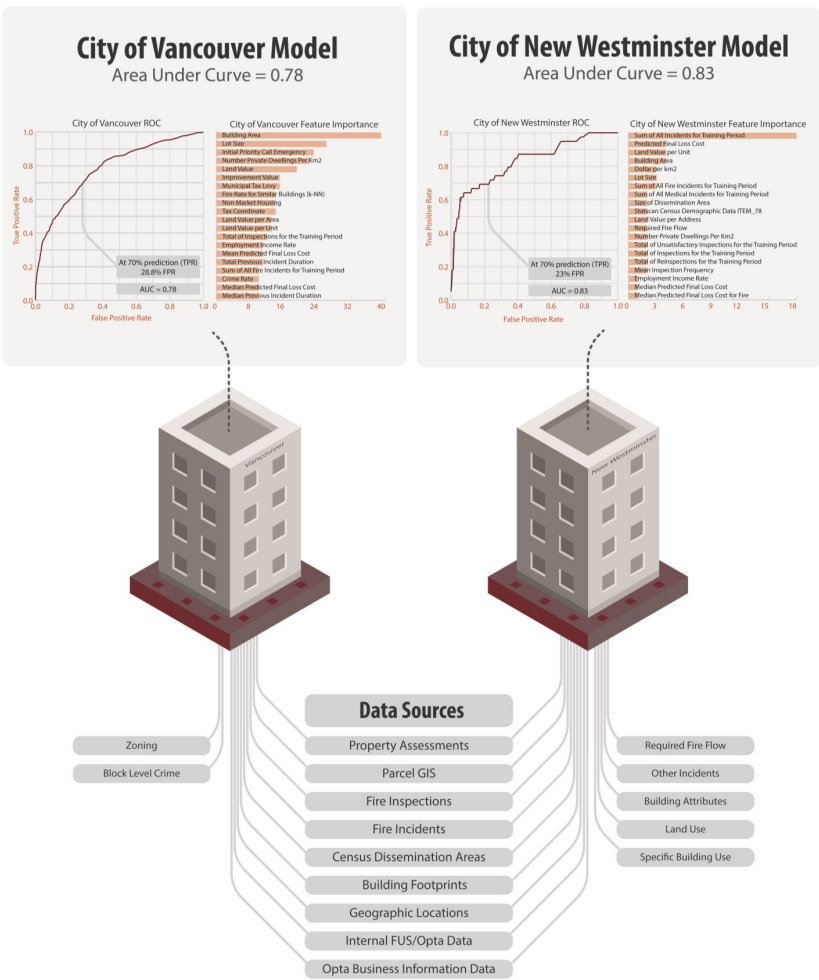
- True Positive Rate (TPR) : 모든 화재 사건 중 올바르게 식별된 화재 사건의 비율
- False Positive Rate (FPR) : 잘못 예측된 화재 사건의 비율
- Area Under Curve (AUC) : ROC (Receiver Operating Characteristic) 곡선 아래의 면적으로 TPR 대 FPR 그래프 모형

출처 : McGuinness *et al.*. 2019. A Building Fire Risk Prediction Validation Project. p.8

데이터 분석을 위한 머신러닝 모델은 XGBoost알고리즘을 활용하였다. 분석 결과 벤쿠버 시의 AUC는 0.78, 뉴웨스트민트 시의 AUC는 0.83을 보였다.

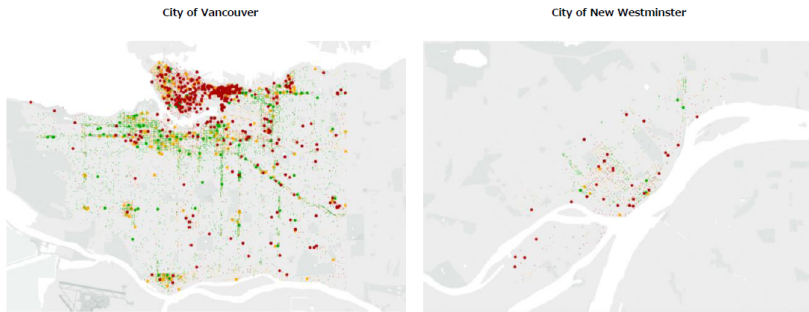
9) McGuinness *et al.*. 2019. A Building Fire Risk Prediction Validation Project. pp. 1-16

이는 벤쿠버 시의 경우 2017년에 대해 70%의 화재를 정확하게 예측(TPR: True positive rate)하면서 잘못 예측된 화재 사건(FPR: False positive rate)은 28.8%이며, 뉴웨스트민스트 시의 경우 70%의 화재를 정확하게 예측(TPR)하고 잘못 예측된 화재 사건(FPR)은 23%였다는 것을 의미한다.



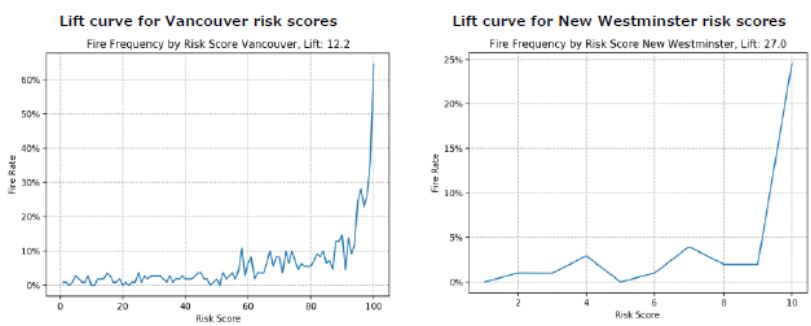
[그림 2-7] 화재예측 모델 및 결과
출처: McGuinness *et al*.. 2019. A Building Fire Risk Prediction Validation Project. p.9

분석 결과를 살펴보면 벤쿠버 시의 경우 화재발생에 영향을 미치는 특성은 토지의 가치, 건물 바닥면적, 기존 화재 사건 및 검사, 고용 소득률, 범죄율 등의 영향이 높았으며, 뉴웨스트민스터 시의 경우 건물 연면적, 예상 손실 비용, 고용 소득률, 인구 통계 및 사회데이터의 영향력이 높게 나타났다.



[그림 2-8] 벤쿠버 시 및 뉴웨스트민스터 시의 건축물 화재위험도
출처: McGuinness *et al.*. 2019. A Building Fire Risk Prediction Validation Project. p.10

구축한 모델의 평가를 위해 Kappa Statistic (Kappa 통계량), Precision (정밀도), Recall (재현율) 분석을 수행하였다. 정밀도는 화재발생 예측 건축물 중 실제 화재 발생 비율을 의미하며, 재현율은 건축물 중 화재발생이 예측된 건축물의 비율을 총 관련 건물 수로 나눈 값이다. Kappa 통계량은 실제 값과 예측 값 사이의 일치성을 측정하며, 우연에 의해 발생하는 일치도를 고려한다. 측정된 값을 통해 각 시별로 정밀도-재현율 곡선을 구축하고, 이를 Lift 분석을 수행하여 예측 모델의 효과를 측정하였다. 분석 결과 벤쿠버 시의 Lift 값은 12.2 이며, 뉴웨스트민스터 시의 Lift 값은 27로 나타났다.



[그림 2-9] 벤쿠버 시 및 뉴웨스트민스터 시의 Lift 분석 결과
출처: McGuinness *et al.*. 2019. A Building Fire Risk Prediction Validation Project. p.11

모델 결과를 활용하여 각 도시의 건물에 화재에 대한 위험점수를 할당하고, 높은 위험점수(90점 이상)를 받은 건물들을 그룹화 하였다. 이를 활용하여 화재 위험이 높은 건축물에 대한 화재 점검 수행 시 효과에 대해 분석하고, 향후 건축물 화재 점검 수행 방안을 제안하였다.

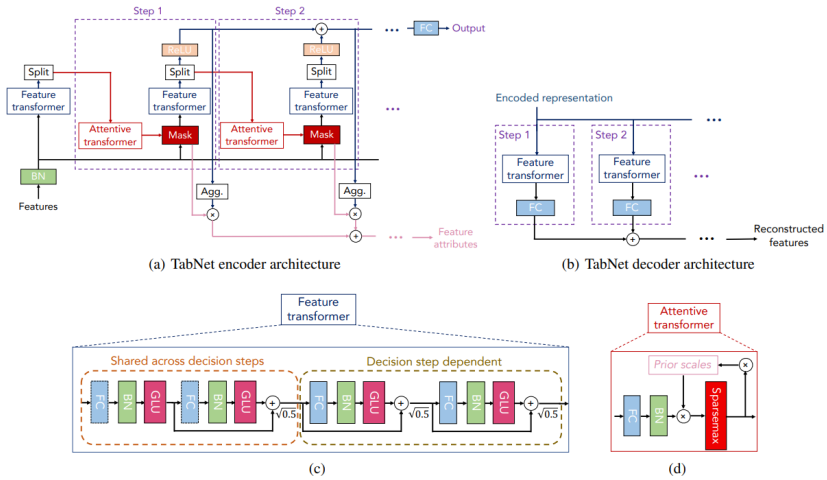
[그림 2-10] 모델 활용을 통한 점검 프로그램의 효율성 향상 비율

출처: McGuinness *et al.*. 2019. A Building Fire Risk Prediction Validation Project. p.13

③ 건축물 화재 발생 예측 모델[10)]

조영진 외(2022)는 '빅데이터 기반 건축물 화재 예측 모델 개발 연구'를 통해 건축물 단위의 화재 예측 모델을 개발하였다. 이를 위해 결과에 대한 해석이 용이한 의사결정나무 모델과, 테이블 형태의 데이터 학습에 높은 성능을 나타내는 딥러닝 모델인 TabNet을 활용하였다. 또한, 5년간의 서울시 건축물 화재발생 데이터를 건축물의 속성정보가 담긴 건축물대장과 연계하여 학습 및 검증 데이터로 사용하였다. 데이터 중 80%는 훈련데이터, 10%는 검증데이터, 나머지 10%는 시험데이터로 활용하여 모델 학습 결과를 검증 및 시험하였다.

10) 조영진, 허한결, 안의순, 류수연, 송유미, 현태환. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. 건축공간연구원

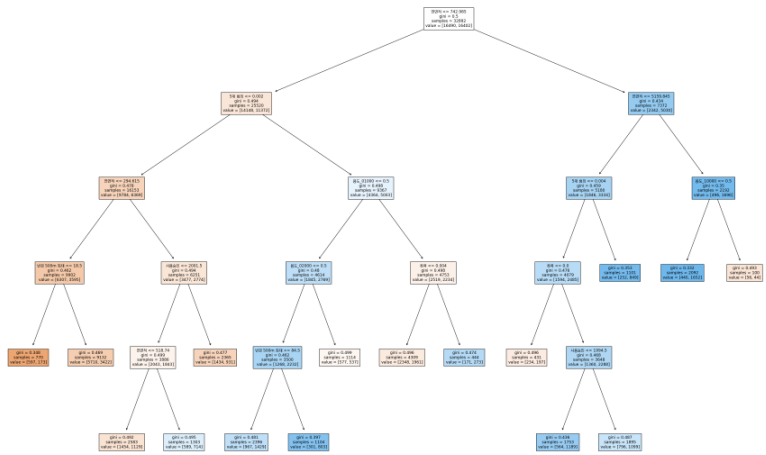


[그림 2-11] TabNet 모델의 신경망 구조

출처: Arik, S. Ö. and Pfister, T. 2020. TabNet: Attentive Interpretable Tabular Learning. p.4. 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. 건축공간연구원, p.76.으로부터 재인용

딥러닝 모델 학습을 위해 분석데이터 전처리 과정을 수행하였다. 분류 범주 불균형을 해결하기 위한 언더샘플링, 학습·검증·시험을 위한 데이터셋 분할, 범주형 변수의 이진화, 결측값 대체 등을 수행하였다.

의사결정나무 모델의 학습결과 16개의 잎과 노드를 가진 모델이 도출되었으며, 최종 정확도는 61%로 나타났다. 의사결정나무 모델의 학습 결과를 토대로 화재에 취약한 건축물 유형 상위 3가지를 도출하였다. 전체 건축물의 화재 발생확률은 3.8%수준으로 나타났으며, 가장 취약한 유형은 "연면적 5,160m^2 이상, 교육연구시설 제외"(화재 발생확률 12%), 두 번째는 "연면적 743m^2 이상 5,160m^2 미만, 5대 범죄 발생률 0.004건/m^2 이상"(화재 발생확률 10.7%), 세 번째는 "연면적 743m^2 미만, 5대 범죄 발생률 0.002건/m^2 이상, 단독주택(대분류) 및 공동주택 제외, 반경 500m 화재 발생 85건 이상"(화재 발생확률 8.1%)로 나타났다.

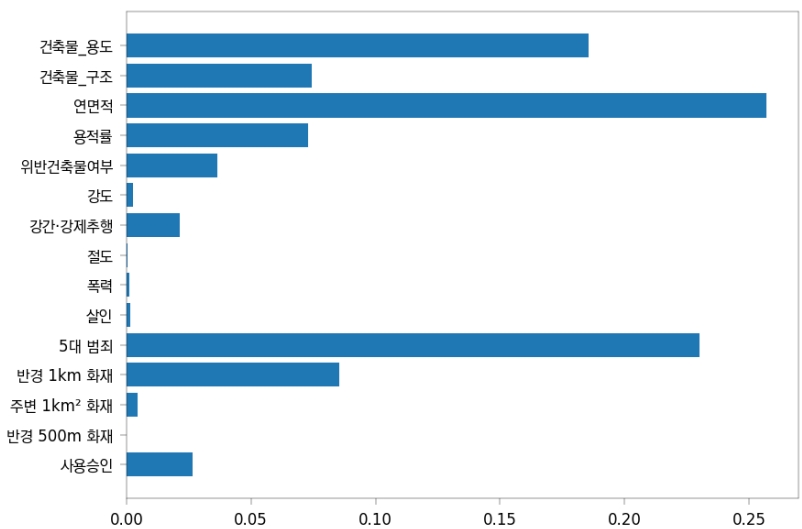


(파랑: 화재 발생률 높음, 주황: 낮음)

[그림 2-12] 건축물 화재 발생 예측 의사결정나무 모델

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.98

TabNet을 활용한 딥러닝 학습의 경우 정확도 65.8%로 나타났다. 딥러닝 모델의 경우 모델특성상 개별 변수가 건축물 화재발생에 미치는 영향에 대해 설명할 수 없다. 다만, 건축물 화재발생 예측에 중요하게 사용된 변수의 중요도를 설명할 수 있다. 이에 따르면 건축물의 연면적과 용도가 가장 중요한 것으로 나타났다. 또한, 독립변수로 활용된 5대범죄 발생여부가 중요도가 높게 도출되었다.



[그림 2-13] 건축물 화재 예측에서 TabNet 모델의 변수 중요도

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.101

마지막으로, 학습된 TabNet 모형을 이용하여 서울시 전체의 화재 취약 건축물을 모델링하고 이를 공간정보화 하였다. 이를 통해 화재발생 가능성이 높은 건축물을 사전에 예측하고, 해당 건축물이 있는 위치를 확인할 수 있으며, 최종적으로 건축물 화재발생을 예방할 수 있도록 기반을 마련하였다.

발생: 빨간색, 미발생: 파란색

[그림 2-14] 건축물 화재 발생가능성 예측 결과

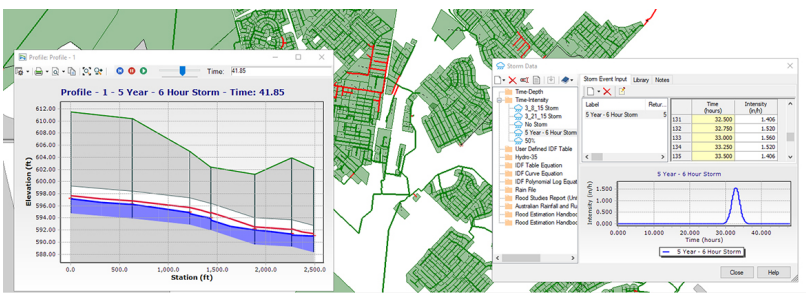
출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.103

2. 홍수 피해 및 리스크 분석

1) 홍수 발생 예측 모형 및 선행연구

① SWMM[11]

SWMM (Storm Water Management Model)은 1971년 미국 환경 보호국(EPA: Environmental Protection Agency)에서 개발된 우수 관리 모델이다. SWMM은 도시의 유출수 및 배수 시스템을 시뮬레이션 하는 소프트웨어 패키지로, 우수 관리 시스템의 분석 및 설계를 위해 주로 사용된다. 또한, 유출수 및 배수 시스템에 대한 시뮬레이션 능력이 우수하여 도시 홍수 분석에 주로 활용되는 모형이기도 하다.



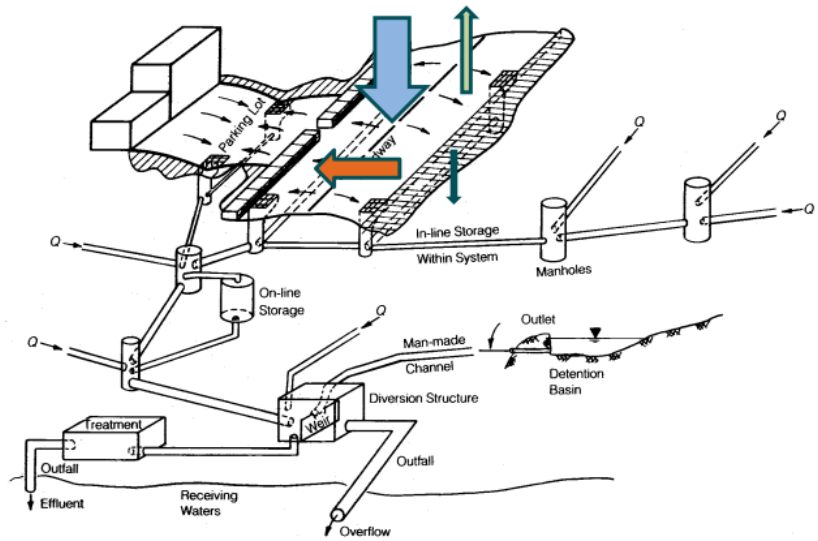
[그림 2-15] 강우를 고려한 도시 홍수 시뮬레이션 결과

출처: BURGESS&NIPLE. https://www.burgessniple.com/insights/2021/8/16/a-deep-dive-into-swmm-modeling-5-common-mistakes/. 2023.06.10. 접속.

SWMM은 배수분구의 강우, 지형, 하수관거 및 맨홀 자료를 사용하여 대상지의 배수능력을 평가하기 때문에 다른 모형에 비해 정확도가 높다. 도시지역의 동일 배수분구를 대상으로 시뮬레이션을 구동하며, 시간의 따른 강우량의 변화, 지표수의 증발, 저류지를 통한 강우 차단, 토양층으로의 강우 침투, 지하수와 배수 시스템 간 상호작용, 비선형 저류식 및 연속식 수로와 관거, 다양한 저영향 개발(LID: Low Impact Development) 요소까지 고려할 수 있다[12]. 이 같은 특성으로 인해 SWMM이 동일 배수분구 내에서는 가장 높은 정확도를 나타내는 유출수 분석 모델로 알려져 있다.

11) EPA. https://www.epa.gov/water-research/storm-water-management-model-swmm. 2023.06.10. 접속

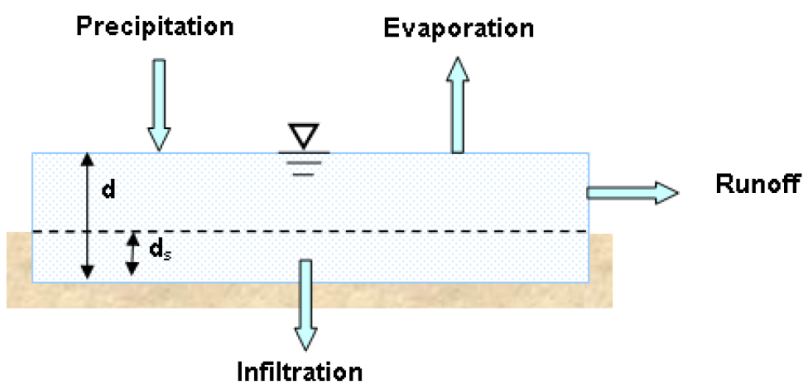
12) EPA. 2022. Storm Water Management Model User's Manual Version 5.2. p.15



[그림 2-16] SWMM 분석 개념도

출처: Simon, Michelle.. 2021. Getting Started with EPA's Storm Water Management Model (SWMM). p.8.

SWMM의 강우 흐름 계산은 총 9가지 물리적 프로세스를 고려한다. 지표 유출수, 지하수, 표면 연못, 침투, 눈 녹음, 수질 라우팅, 지하수, 유량 라우팅, 저영향 개발 등 9가지 요소가 이에 해당한다. 홍수와 밀접한 관련이 있는 지표 유출의 경우 침투, 증발, 지표유출이 고려된다. 이를 산정하기 위한 우수 유입량은 기본적으로 강수량과 해당 지역 상류 지역으로부터의 유입량을 더하여 결정된다[13].



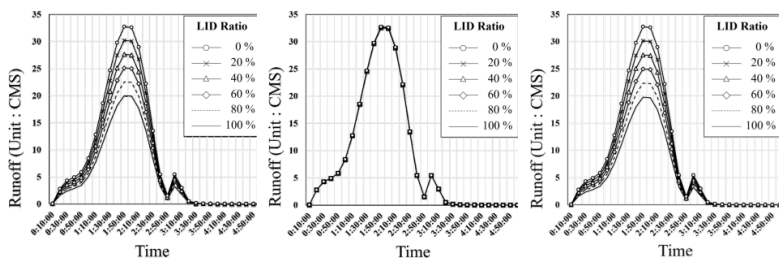
[그림 2-17] 강우, 토양침투, 증발 등을 고려하는 SWMM의 강우 유출 분석 개념도

출처: EPA. 2022. Storm Water Management Model User's Manual Version 5.2. p.85

13) EPA. 2022. Storm Water Management Model User's Manual Version 5.2. p.84

SWMM은 강우에 의한 유출수 분석과 홍수 모의 뿐 아니라, 유출수에 의한 환경 오염물질 확산과 축적 모델링이 가능하다. 비가 오지 않는 시점의 토지 이용에 따른 오염물질 축적, 강우 발생 시 특정 토지 용도로부터의 오염물질 확산, 오염물질 확산에 대한 강우의 기여정도, 도시 정화 활동에 따른 오염물질 감소, 배수 시스템을 통한 수질 관리를 고려하여 환경오염물질의 축적과 확산을 분석 한다[14].

미국 환경 보호국에 따르면 SWMM은 주로 7가지 부문에 적용되고 있다. 도시계획에 있어 홍수 대응을 위한 배수 시스템 설계, 홍수 대응 및 수질 관리를 위한 저류시설 설계, 자연하천의 범람원 도출, 하수도 역류 최소화를 위한 설계 전략 도출, 강우에 따른 하수도 역류 영향 평가, 강우에 의한 비점 오염원 부하량 검토, 강우 시 오염물질 부하를 줄이는 최적관리기법(BMP: best management practices)의 효과 평가에 적용된다[15].



[그림 2-18] 저영향 개발 도입 시나리오에 따른 강우 유출량 분석

출처: 김재경, 강준석. 2022. SWMM 모델을 활용한 상도동 지역의 우수유출관리 및 저영향개발의 유출저감효과 분석 연구. 한국산학기술학회논문지. 23(8). p.344

② SWAT[16]

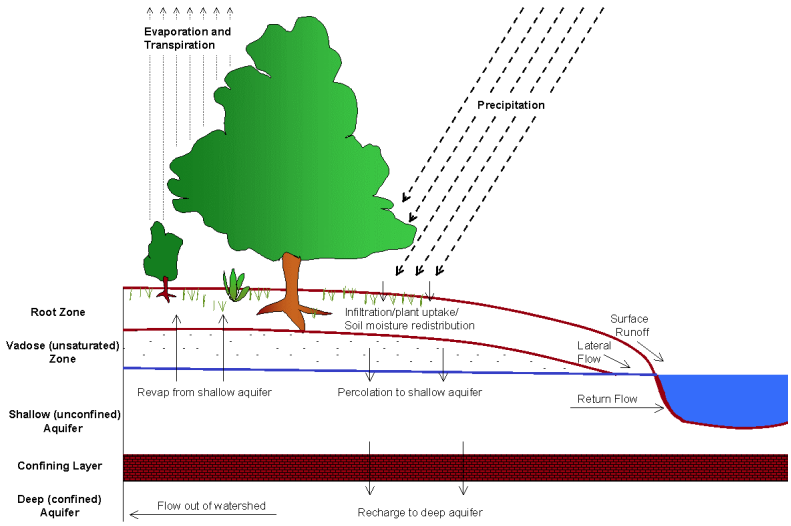
SWAT (Soil and Water Assessment Tool)은 미국 농무부(USDA: United States Department of Agriculture)에서 개발한 토양 및 수자원 평가 도구이다. 유역 내 토지이용 및 관리에 의한 수자원 영향을 시뮬레이션하는 수문학 기반 모델로, 유역 수준에서 넓은 공간 규모에서 사용된다. 강수, 증발산, 침투, 지표 유출, 지하수 흐름, 하천 흐름 등 수문 순환의 다양한 구성요소를 포함하고 있고, 물과 함께 이동하는 퇴적물, 영양분 및 농약 이동 시뮬레이션이 가능하다[17].

14) EPA. 2022. Storm Water Management Model User's Manual Version 5.2. p.16

15) EPA. 2022. Storm Water Management Model User's Manual Version 5.2. p.16

16) NEITSCH, S.L. *et al.*. 2002. SOIL AND WATER ASSESSMENT TOOL THEORETICAL DOCUMENTATION VERSION 2000

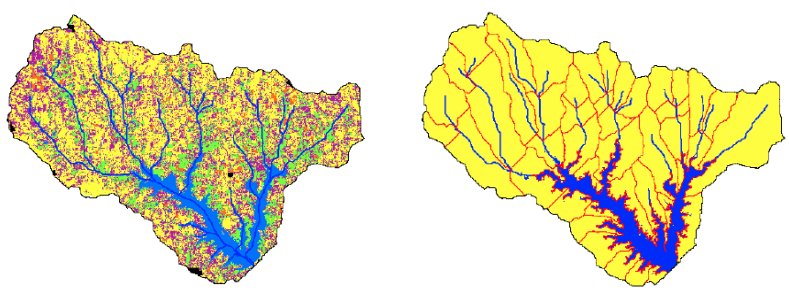
17) NEITSCH, S.L. *et al.*. 2002. SOIL AND WATER ASSESSMENT TOOL THEORETICAL DOCUMENTATION VERSION 2000. p.1



[그림 2-19] SWAT의 수문 분석 개념도

출처: NEITSCH, S.L. *et al.*. 2002. SOIL AND WATER ASSESSMENT TOOL THEORETICAL DOCUMENTATION VERSION 2000. p.8

SWAT은 넓은 지역에 대한 수문 분석을 수행하기 위해 거대 유역을 여러 개의 하위유역으로 분할하여 모델링할 수 있다. 특히 하위유역 간 토지이용이나 토양유형 등 수문에 영향을 미치는 특성이 다른 경우 이 같은 하위유역 분할이 효과적이다. 하위유역으로 구분된 개별 소유역에는 기후, 연못 및 습지, 지하수, 배수로, 수문 특성 등 정보가 입력되어 분석에 사용된다[18]. 이에 따라 SWAT은 다른 모델에 비해 높은 정확도로 넓은 지역의 수문 분석이 가능하다.



[대상지 토지이용 지도] [하위유역 구분 지도]

[그림 2-20] SWAT의 하위유역 구분 사례(붉은 선)

출처: NEITSCH, S.L. *et al.*. 2002. SOIL AND WATER ASSESSMENT TOOL THEORETICAL DOCUMENTATION VERSION 2000. pp.6-7

18) NEITSCH, S.L. *et al.*. 2002. SOIL AND WATER ASSESSMENT TOOL THEORETICAL DOCUMENTATION VERSION 2000. pp.6-7

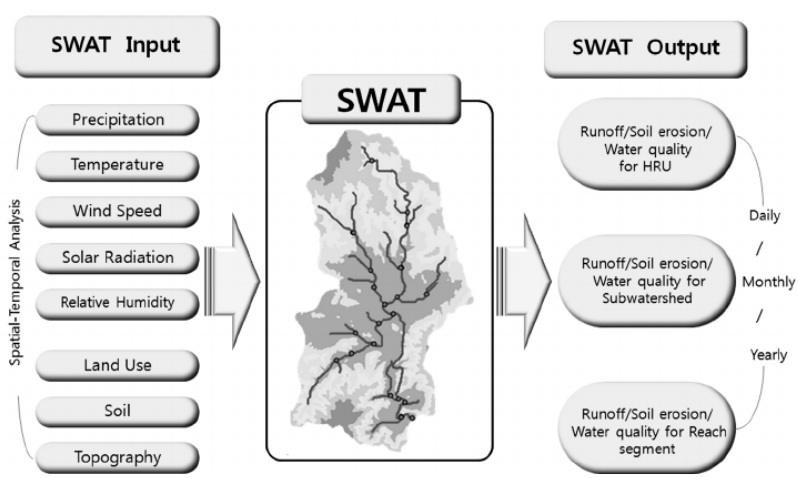
SWAT 모델의 입력자료는 크게 네 가지 유형으로 구분된다. 기상자료, 토지이용도, 토양도, DEM (Digital Elevation Model)이 기본 데이터 유형이며, 유형별로 장기 일 강우, 일 평균풍속, 일 평균 상대습도 등 세부 입력자료를 사용한다.

[표 2-4] SWAT 모델 입력자료

구분	세부 입력자료
기상	장기 일 강우, 일 평균풍속, 일 평균 상대습도, 일 최고온도, 일 최저온도, 일 수평면일사량
토지이용	산림, 농업지역, 시가화지역, 초지, 습지, 나지, 수역 등
토양	자갈, 모래, 실트, 점토, 사양토 등 함량정보
DEM	수치표고모형(DEM: Digital Elevation Model)

출처: 허성구, 김남원, 박윤식, 김종건, 김성준, 안재훈, 김기성, 임경재. 2008. SWAT ArcView GIS Extension Patch를 이용한 소유역 분할에 따른 수문 및 유사 거동에 미치는 영향 평가. 한국물환경학회지. 한국수질학회. 24(2) pp.149-150

분석을 통해 분석 대상지의 유출량, 유출특성, 토양 수분함량, 퇴적물 유출, 수질을 모의할 수 있다. 이 중 홍수 분석을 위해 주로 사용되는 것은 유출량으로, 일별, 월별, 연간 유출량의 형태로 도출된다. 홍수 발생과 관련된 분석을 위해 모델의 설정과 입력자료를 조정하여 시간 단위로 유출량을 분석하는 경우도 있다. 또한, SWAT모델을 GIS 소프트웨어와 연동하여 유출량 분석을 수행할 수 있다[19].



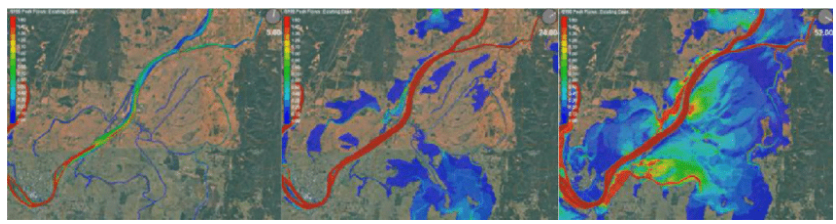
[그림 2-21] SWAT의 입력 및 출력 데이터

출처: 허성구, 김남원, 박윤식, 김종건, 김성준, 안재훈, 김기성, 임경재. 2008. SWAT ArcView GIS Extension Patch를 이용한 소유역 분할에 따른 수문 및 유사 거동에 미치는 영향 평가. 한국물환경학회지. 한국수질학회. 24(2) p.149

19) 허성구, 김남원, 박윤식, 김종건, 김성준, 안재훈, 김기성, 임경재. 2008. SWAT ArcView GIS Extension Patch를 이용한 소유역 분할에 따른 수문 및 유사 거동에 미치는 영향 평가. 한국물환경학회지. 한국수질학회. 24(2) p.151

③ TUFLOW[20)]

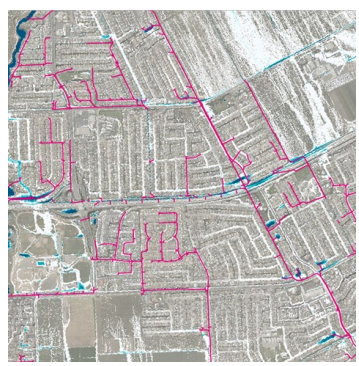
TUFLOW (Two-dimensional Unsteady FLOW)는 하천, 해안 및 도시지역의 홍수와 수질을 분석하는 시뮬레이션 소프트웨어이다. TUFLOW는 2D 모델링 소프트웨어로, 범람원, 연안 해역, 하구, 강, 도시 지역의 침수 패턴 분석에 강점이 있다. ArcGIS와 같은 GIS 소프트웨어와 연동한 분석이 가능하고, 홍수 발생 분석 시 홍수발생 범위와 침수심을 포함하는 데이터 맵핑이 가능하다. 이에 따라 범람원 맵핑, 홍수 예측 및 위험성 평가, 도시 배수 설계, 환경영향평가 등 다양한 분야에서 사용되고 있다.



[그림 2-22] TUFLOW의 홍수 분석 예시

출처: ACKERE, S. VAN. *et al.*. 2016. DEVELOPMENT OF A 3D DYNAMIC FLOOD WEB GIS VISUALISATION TOOL. International Journal of Safety and Security Engineering. 6(3). p.564

TUFLOW모델은 1차원 파이프 모델을 사용하여 하수관거를 적용하며, 지면의 물 흐름은 2차원 수문 모델을 사용하여 분석한다. 2D 도메인과 1D 하수관거를 동적으로 연결해 시뮬레이션하여 정확도가 높고, 컴퓨터의 그래픽 처리 장치(GPU)를 병렬로 사용할 수 있어 시뮬레이션 속도가 빠르다는 장점이 있다.



[분석 대상지 하수관거 입력 예시]

[홍수 분석 결과 예시]

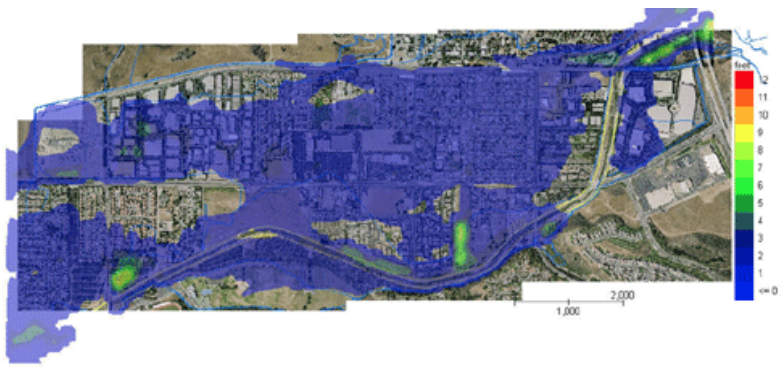
[그림 2-23] TUFLOW의 홍수 입출력 데이터 예시

출처: TUFLOW. https://www.tuflow.com/products/tuflow/. 2023.06.14. 접속

20) BMT. 2018. TUFLOW Classic/HPC User Manual

④ FLO-2D

FLO-2D는 2차원 홍수 모델링 소프트웨어로, 도시환경과 자연환경을 모두 고려하여 시뮬레이션 한다. 격자단위 분석을 수행하며, 물의 흐름을 8개 방향으로 분석함으로써 강우 유출수에 대한 정밀한 분석이 가능하다. FLO-2D를 구동하기 위해 DTM (Digital Terrain Model)과 강우데이터 등 유입 수문곡선이 필요하며 이를 통해 홍수발생을 모의할 수 있다. 세부적으로는 토양 침투와 증발, 강우의 지표 흐름과 하수관거 흐름, 도로와 건물에 의한 도시 내 흐름(Street Flow), 제방 및 둑과 같은 수리구조물의 효과 등 수리·수문학적 분석을 수행한다.



[그림 2-24] FLO-2D의 홍수 분석 예시
출처: 우림텍. http://www.woorimtech.com/page/flo211. 2023.06.14. 접속

FLO-2D는 관거 내 물의 흐름과 지표유출 및 시가지 유출을 모두 고려하는 장점으로 홍수문제 뿐 아니라 다양한 분야에서 활용되고 있다. 특히 복잡한 지형지물이 많은 도시 구조를 반영하는 능력이 탁월하여 도시 내 강우 흐름, 침수심, 유속 분석에 유용하다. 이에 따라 일반적으로 홍수위험지역 맵핑, 범람원 관리, 도시 배수 설계, 홍수 및 해일에 대비한 비상 대응 계획에 활용된다.

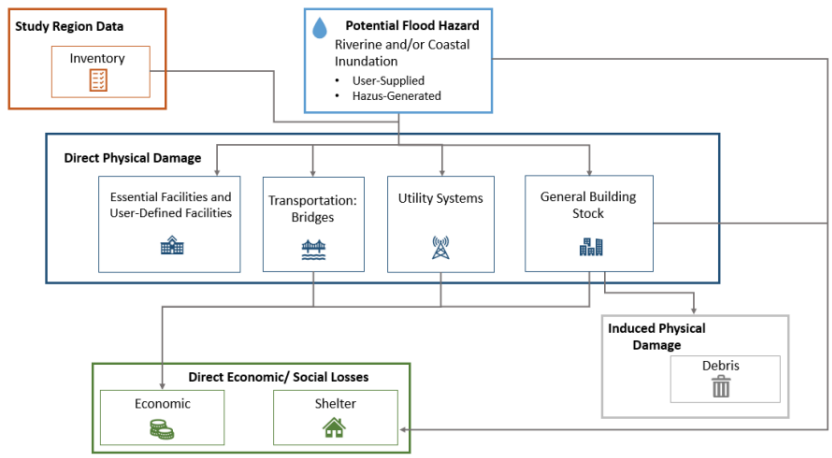
2) 리스크 분석 모형

① Hazus-MH[21)]

Hazus-MH는 미국 연방재난관리청(FEMA)에서 개발한 복합 자연재해에 대한 리스크 평가 도구이다. "Hazards US Multi-Hazard"의 약자인 Hazus-MH는 지진, 홍수, 허리케인 등의 자연재해로 인해 발생할 수 있는 잠재적 손실 추정을 지원한다. 지리 정보 시스템(GIS) 데이터를 이용하여 특정 지역의 건물, 인프라 및

21) FEMA. https://www.fema.gov/flood-maps/products-tools/hazus. 2023.06.14. 접속

인구에 대해 지진, 홍수, 허리케인 등 다양한 유형의 자연재해가 미치는 영향을 시뮬레이션 할 수 있다. 시뮬레이션을 위해 건축물 특성, 인구 밀도, 교통, 통신을 포함한 중요 도시계획 시설을 입력데이터로 사용하여 분석의 정확성을 높인다.



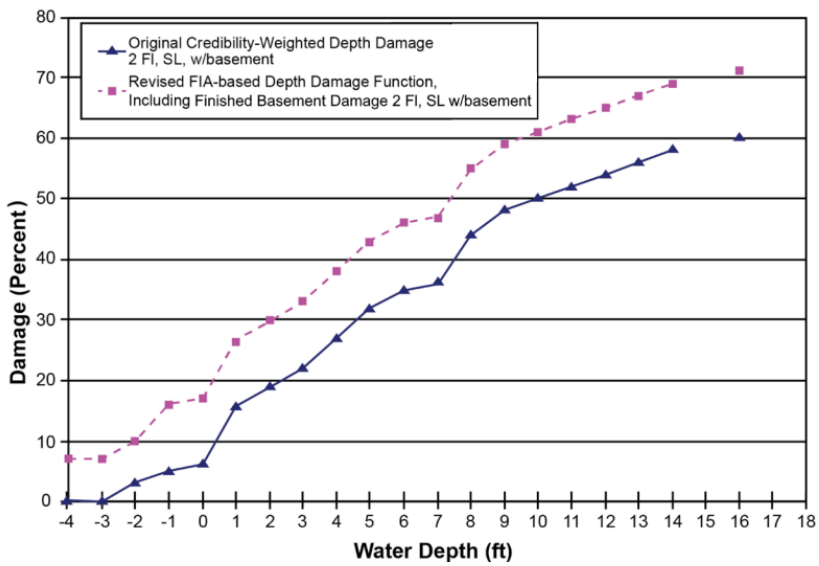
[그림 2-25] Hazus flood model의 분석 구조
출처: FEMA. 2022. Hazus Flood Technical Manual: Hazus 5.1. p.2-2

Hazus-MH는 자연재해로 인한 잠재적 피해와 손실에 대한 종합적인 리스크 평가가 가능하며, 이를 이용하여 자연재해 리스크 지도를 생성한다. Hazus-MH를 이용하여 도시 계획가 및 지자체 의사 결정자가 고위험 지역을 식별하고 영향을 받는 사람과 건물의 수를 추정할 수 있는 등 자연재해 피해를 줄이기 위한 대응방안 마련의 근거로 사용된다. 미국 연방, 주, 지방 수준은 물론 캐나다, 멕시코, 도미니카공화국 등 다른 해외 국가에서도 미국과 유사한 방식으로 활용되고 있다. 건축물의 구조나 용도, 연령 및 거주자 수, 인구 밀도 및 인구 통계, 주요 도시 인프라, 강 및 하천 데이터, 해안선 등의 정보를 입력데이터로 활용하여 분석을 수행한다.

Hazus-MH 모델은 잠재적 리스크 추정을 위해 확률적 접근 방식을 사용한다. 자연재해의 강도, 발생 위치 및 기타 입력데이터를 고려하여 각 재해 발생에 대한 수천 번의 시뮬레이션을 수행하고 이를 바탕으로 리스크를 추정한다. 이 같은 확률론적 접근 방식을 사용함으로써 분석결과의 정책 활용을 위한 자연재해의 불확실성과 변동성 정량화가 가능하다.

특히 홍수의 경우, 확률론적 홍수 위험 분석(PFHA: probabilistic flood hazard analysis) 방법을 적용한다. 홍수는 강우 강도, 강우 지속시간, 지형 공간 패턴에

따라 발생하며, 이에 따라 침수 정도 및 피해 수준이 달라진다. 이에 대응하기 위해 Hazus-MH는 과거 강우데이터부터 지형, 토지 피복, 건축물 특성 등 다양한 입력자료를 사용함으로써, 다양한 강우 강도와 지속 시간의 확률의 추정, 다양한 수준의 침수 리스크 분석을 제공한다.



[그림 2-26] Hazus flood model의 손상함수
출처: FEMA. 2022. Hazus Flood Technical Manual: Hazus 5.1. p.5-16

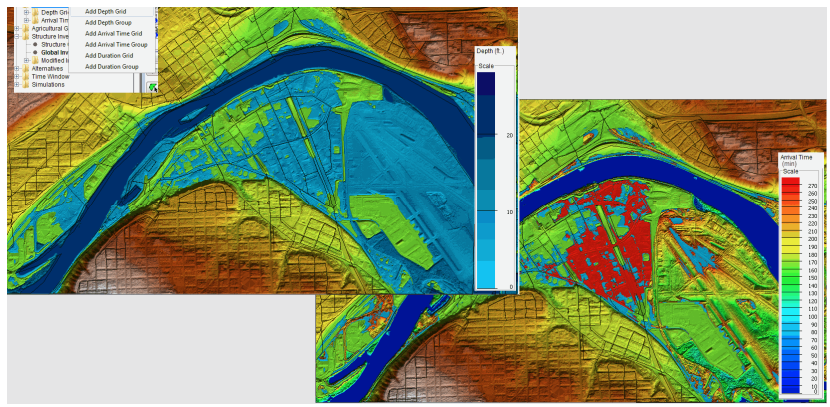
특히, Hazus-MH는 손상 함수를 이용하여 자연재해 영향에 대한 피해금액을 산정할 수 있다. 손상 함수는 자연재해 영향과 건축물이나 자동차 등 다양한 객체에 대한 예상 피해 관계를 설명하는 함수로 설명 가능하다. 홍수의 경우 건물 유형, 건축 자재, 연식, 거주자 수 등이 홍수 피해에 대한 건물의 피해금액에 영향을 미친다. Hazus-MH의 손상함수는 과거 홍수 피해 데이터와 홍수 발생 시 건축물의 홍수 대응 수준을 고려하여 보정되고, 이를 기반으로 피해금액 예측이 수행된다. 이 같은 방법으로 Hazus-MH는 복합적인 자연재해 피해를 예측하고 대응책마련을 위한 근거자료를 제공한다.

② HEC-FIA[22)]

HEC-FIA는 미국 육군 공병단의 수문공학센터(HEC)에서 개발된 고급 모델로, 홍수 침수 시뮬레이션 및 피해 예측을 위해 설계되었다.

22) U.S. Army Corps of Engineers Hydrologic Engineering Center. https://www.hec.usace.army.mil/software/hec-fia/. 2023.06.15. 접속

Hydrologic Engineering Center's Flood Impact Analysis의 약자인 HEC-FIA는 홍수의 발생과 그로 인한 침수를 시뮬레이션한다. HEC-FIA는 지형 정보와 다양한 지리 정보 시스템(GIS) 데이터를 사용하여 범람원 주변의 지형, 식생 및 인프라를 모델링 하며 이를 고려한 침수 시뮬레이션을 수행한다.

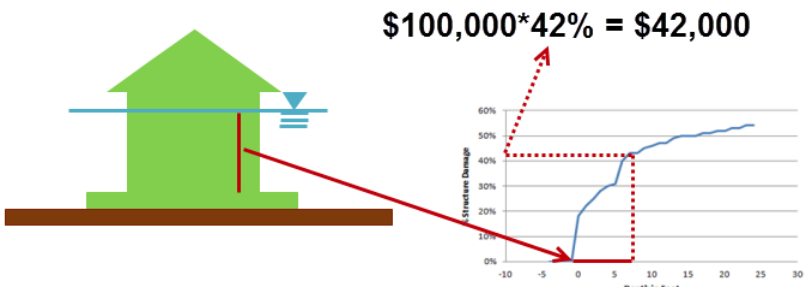


[그림 2-27] HEC-FIA 모델 입력자료 예시

출처: U.S. Army Corps of Engineers Hydrologic Engineering Center. https://www.hec.usace.army.mil/software/hec-fia/features.aspx. 2023.02.27. 접속

HEC-FIA는 그 사용 범위가 광범위하여, 도시 뿐만 아니라 농촌 지역과 같은 다양한 환경에 적용될 수 있다. 이에 따라 범람원 관리나 홍수 발생 시 대응 전략 수립에 유용한 도구로 알려져 있으며, 홍수 위험 관리와 재난 대응과 관련된 정부기관과 도시계획 및 설계를 수행하는 엔지니어링 회사에서 재해 모델링을 위해 사용되고 있다.

HEC-FIA는 홍수 피해 예측을 위해 세 가지 주요 모듈을 사용한다. 첫 번째 모듈인 HEC-GeoRAS는 지리 데이터 처리 시스템으로, 복잡한 지형 및 지리 데이터를 가공하여 시뮬레이션에 필요한 데이터로 변환한다. 두 번째 모듈인 HEC-RAS는 하천 분석 시스템으로, 지형정보를 기반으로 강의 수리학적 특성과 수로 흐름을 시뮬레이션한다. 마지막으로 HEC-HMS는 수문 모델링 시스템으로, HEC-GeoRAS를 통해 도출된 지형정보 및 강우량을 바탕으로 강우 유출 과정과 그에 따른 하천 흐름을 시뮬레이션 한다. 이러한 모듈들은 도시와 농촌 등 복합적이고 다양한 환경에서의 홍수 피해를 신속하고 정확하게 예측할 수 있도록 지원한다.



[그림 2-28] HEC-FIA 모델의 피해비용 예측 모식도

출처: U.S. Army Corps of Engineers Hydrologic Engineering Center. 2018. HEC-FIA Flood Impact Analysis: Technical Reference Manual. p.3-2

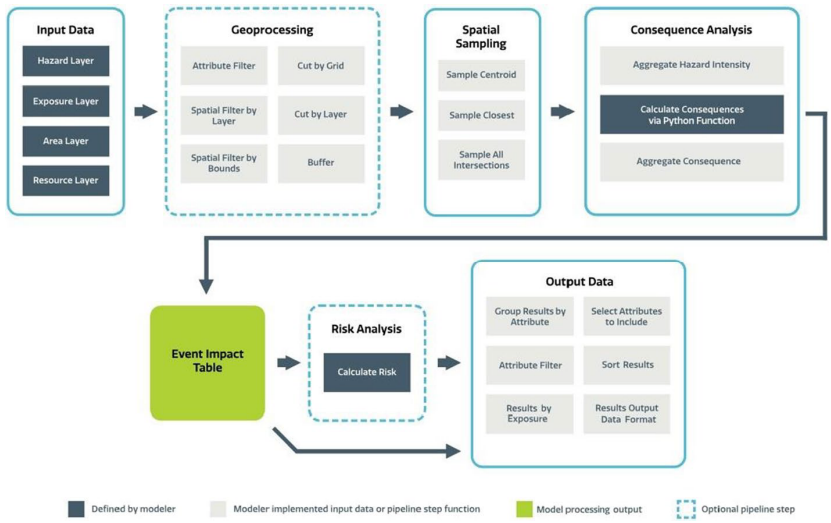
마지막으로, HEC-FIA는 구체적인 피해액 예측을 위해 손상함수를 개발하여 적용하고 있다. 침수심에 따른 구조물의 손상 정도를 손상률로 정의하며, 이를 통해 각 구조물의 피해액을 산정한다. 피해액은 손상함수와 구조물의 비용 가치(금액)에 따른 함수로 결정되며, 건축물과 같은 유형별 구조물의 가치 데이터를 입력데이터로 활용한다. 이를 통해 건물, 인프라, 기타 자산에 대한 피해를 포함하여 홍수로 인한 직간접적인 경제적 피해를 추정할 수 있다.

③ Riskscape[23)]

Riskscape는 뉴질랜드 매시 대학교의 재해 공동연구센터(JCDR)에서 개발한 리스크 분석 모델이다. GIS를 기반으로 지진, 홍수, 산사태, 화산 폭발 등 자연 재해에 대한 리스크 분석이 가능하며, 자연재해가 인프라, 건물, 환경 등 도시의 다양한 물리환경요소에 미치는 영향을 분석한다. Riskscape는 공간 데이터와 지리정보를 결합하여 분석하기 때문에 상대적으로 정확도가 높고 분석결과에 대한 근거가 명확하다고 평가받는다.

Riskscape는 다양한 데이터를 수집하고, 내부 모델을 연계하여 리스크 분석을 수행한다. 위험 모델, 노출 데이터, 취약성 모델 등 다양한 소스에서 얻은 데이터를 통합하여 지오 프로세싱, 샘플링, 손상정도 예측 등 과정을 거쳐 종합적인 리스크 평가를 수행한다. 건물 유형, 토지 이용, 인구분포 등 취약성에 영향을 미칠 수 있는 인프라 및 인구통계 관련 변수를 고려함으로써, 특정 지역의 재해 리스크를 보다 정확하게 평가한다.

23) Riskscape. https://riskscape.org.nz/. 2023.06.15. 접속



[그림 2-29] Riskscape 분석 프로세스 및 모델 파이프라인 개념도

출처: Paulik, Ryan. *et al.*. 2023. RiskScape: a flexible multi-hazard risk modelling engine. Natural Hazards. 119. p.1078

Riskscape를 통해 자연재해의 잠재적 영향을 시뮬레이션 할 때 지진, 홍수, 산사태, 화산 폭발 등 다양한 시나리오를 반영할 수 있다. 예를 들어, 지진이 다양한 유형의 건물이나 인프라에 미치는 영향을 모델링하거나 홍수가 다양한 토지 이용에 미치는 잠재적 영향 시뮬레이션이 가능하다. 특히, 리스크 분석에는 재해, 노출, 취약성 개념을 활용한다. 이를 위해 재해 발생 데이터, 구조물의 재해 노출 정도, 취약성(손상) 함수를 입력자료로 활용하여 구조물의 리스크를 산정한다. 리스크는 최종적으로 피해액을 산정하여 제시한다.

3. 소결

□ 재난·재해 리스크 분석

화재 및 홍수 관련 분석모델과 연구들은 화재와 홍수 발생확률이나 발생강도를 도출하거나, 이를 기반으로 위험도 및 리스크 분석을 하는 연구로 구분할 수 있다. 화재 및 홍수와 같은 재난·재해는 과거부터 도시민들에게 거대한 피해를 입혀왔고, 이에 대응하기 위해 연구자들이 다양한 분석모델을 개발해 왔다. 화재와 홍수의 발생 가능성과 규모를 예측하여 발생강도를 분석하였고, 이에 더하여 피해금액을 포함한 예측을 수행하고 있다.

화재피해 예측 모델은 시뮬레이션 보다 통계기반 분석이 주를 이루며, 화재발생을 가정할 시 발생강도나 피해액은 통계기반 모델링을 통해 분석하고 있는 것을 확인할 수 있다. 이는 건축물 화재발생의 특성상 화재발생 자체를 예측하기는 어려우나, 구조물의 특성정보를 활용하여 화재발생 시 화재의 규모나 피해금액 분석은 가능하기 때문인 것으로 판단된다.

홍수피해 예측 모델은 강우량 및 수리·수문 모델링을 기반으로 홍수발생을 시뮬레이션하고, 이에 기반하여 건축물의 피해정도를 산정한다. 화재발생 모델과는 다르게 지형정보와 강우량 등의 정보를 이용하여 물의 흐름과 집적을 시뮬레이션 할 수 있다. 이를 통해 홍수가 나는 지역과 수심을 예측할 수 있으며, 홍수가 나는 지역에 있는 구조물의 특성과 손상함수를 이용하여 피해액 단위의 리스크를 분석할 수 있다.

[표 2-5] 재난·재해 분석 모델 및 연구 정리

분석방법	리스크 산정 방식	관련 모델 및 연구
통계기반, 딥러닝	미산정	- NeuroFire
	피해액	- Canadian Wildland Fire Information System - A building fire risk prediction validation project
시뮬레이션 기반 모델링	미산정	- SWMM - SWAT - TUFLOW - FLO-2D
시뮬레이션 기반 모델링, 발생강도 × 손상함수	피해액	- Hazus-MH - HEC-FIA - Riskscape

출처: 연구진 작성

화재와 홍수 분석 모델들 중 리스크를 산정하는 연구들은 리스크 분석 결과를 피해액을 중심으로 제시하고 있으며, 리스크 분석은 크게 두 가지 방법으로 수행되고 있다. 첫째는 시뮬레이션 기반 모델링을 통해 도출된 재해 발생강도와 손상함수를 이용하여 피해액을 산정하는 방식이며, 둘째는 통계기반의 딥러닝 분석을 활용하여 피해액을 분석하는 방법이다. 과거 피해비용 자료와 건축물 등 도시공간의 속성정보를 이용하여 피해액을 분석한다. 딥러닝 기반의 리스크 산정 방식은 딥러닝 모델이 발생확률과 손상함수를 모두 포함하도록 설계되었다고 볼 수 있다.

[그림 2-30] 리스크 산정 방식
출처: 연구진 작성

본 연구에서는 통계기반 딥러닝 모델을 사용하여 리스크 분석을 수행하고자 한다. 홍수 피해와 관련하여서는 시뮬레이션 기반의 모델이 많으나, 본 연구에서 다루는 홍수피해 데이터에는 기존 시뮬레이션 기반 모델이 다루지 못하는 하수도 역류 피해 데이터가 10%를 상회한다. 따라서 시뮬레이션 기반 모델로는 분석할 수 없을 것으로 판단되어 딥러닝 모델 기반의 리스크 분석을 수행하고자 한다. 또한 구득한 화재와 홍수 데이터 모두 피해액 정보가 기입되어있어, 건축물 특성정보 및 기타 물리환경 정보와 연계 시 보다 정확도 높은 리스크 분석이 가능할 것으로 판단된다.

□ 재난·재해 분석 관련 변수

화재 발생 및 리스크 분석을 위해 사용되는 입력변수는 선행연구 검토결과 크게 네 가지로 구분할 수 있다. 첫째로 도시 환경 특성과 관련된 변수이다. Wang *et al.* (2019)은 도시 내 POI들의 용도를 입력변수로 사용하고, 조영진 외(2022)는 도시 내 화재발생 건축물 간 거리를 입력변수로 사용하는 등 도시 환경과 관련된 공간정보를 입력변수로 활용하였다. 두 번째로 사회환경 특성과 관련된 변수이다. 선행연구들에서 인구통계, 사회경제 데이터, 범죄데이터 등 다양한 인구·사회학적 특성을 공간데이터화 하여 분석에 사용하였다. 셋째로 기상특성 관련 변수이다. 선행연구들에서 온도, 습도, 풍속, 강수 등 기상관련 변수가 분석을 위해

사용하고 있음을 확인하였다. 마지막으로 건축물 특성 관련 변수가 사용되고 있다. 건축물의 층수, 엘리베이터 유무 등 건축물이 가지고 있는 속성정보가 화재 발생과 리스크 분석에 적용되고 있다.

홍수발생 및 리스크 분석에는 선행연구들에서 크게 여섯 가지 유형의 입력변수가 사용되고 있다. 첫째로, 지형관련 변수이다. 지형은 강우시 물의 흐름을 결정짓는 가장 중요한 변수로 모든 모델에서 이를 채택하여 사용하고 있다. 같은 이유로 배수 관련 변수 및 토지이용과 토지피복 관련 변수 또한 모든 모델에서 채택하여 사용하는 것으로 확인하였다. 토지특성 관련 변수나 기상 관련 변수는 홍수발생을 예측하기 위해 주로 사용되나 일부 리스크 분석 모델에서는 사용하지 않는 것으로 확인하였다. 마지막으로, 건축물 단위 리스크 분석을 하는 경우 혹은 강우 유출수 흐름을 분석하는 경우 건축물 관련 형상정보나 속성정보를 사용하는 모델이 있는 것을 확인하였다.

[표 2-6] 재난·재해 분석 변수 정리

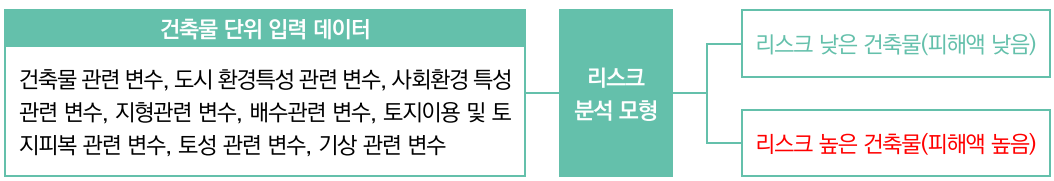
구분	입력변수	관련 모델 및 연구										
		1	2	3	4	5	6	7	8	9	10	11
화재	도시 환경특성 관련 변수		●	●	●							
	사회환경 특성 관련 변수		●	●								
	기상 특성 관련 변수	●	●	●								
	건축물 특성 관련 변수			●	●							
홍수	지형관련 변수					●	●	●	●	●	●	●
	배수 관련 변수					●	●	●	●	●	●	●
	토지이용 및 토지피복 관련 변수					●	●	●	●	●	●	●
	토성 관련 변수					●	●	●	●	●		
	기상 관련 변수					●	●	●	●	●	●	
	건축물 관련 변수								●	●	●	●

주) 1.CWFIS, 2.Wang *et al.* (2019), 3.McGuinness *et al.* (2019), 4.조영진 외(2022), 5.SWMM, 6.SWAT, 7.TUFLOW, 8.FLO-2D, 9.Hazus-MH, 10.HEC-FIA, 11.Riskscape
출처: 연구진 작성

본 연구는 이를 바탕으로 국내 데이터 보유 현황을 검토하고, 관련 연구를 추가 분석하여 건축물 화재 및 홍수 리스크 분석 모델의 입력자료 구축에 활용하고자 한다.

□ 분석 모형의 개념 도출

선행연구를 토대로 정립한 리스크 개념정의와 관련 연구 및 모형 검토결과를 토대로 건축물 화재 및 홍수 리스크 분석 모형의 개념을 설정하였다. 본 연구에서 개발하고자하는 리스크 분석 모형은 건축물 단위의 입력데이터를 받아 건축물의 재난·재해 리스크가 낮은지, 높은지 구분할 수 있도록 설계하고자 한다.

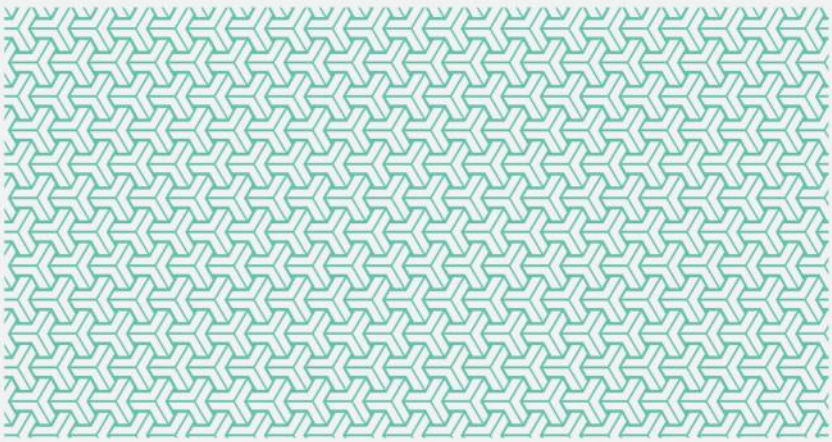


[그림 2-31] 리스크 분석 모형의 개념
출처: 연구진 작성

리스크가 낮은지, 높은지의 여부는 재난·재해 발생 시 피해액을 기준으로 피해액이 낮은 경우 리스크가 낮은 건축물, 피해액이 높은 경우 리스크가 높은 건축물로 구분하고자 한다. 피해액의 크고 작음은 명확한 기준이 없어 본 연구에서 조작적으로 정의하여 진행할 필요가 있다. 본 연구에서 다루는 피해액의 높고 낮음은 상대적인 수치로 보고 현재까지 발생한 건축물 화재 피해액과 홍수 피해액을 근거로 통계적으로 구분하여 연구를 진행하고자 한다.

Chapter.3

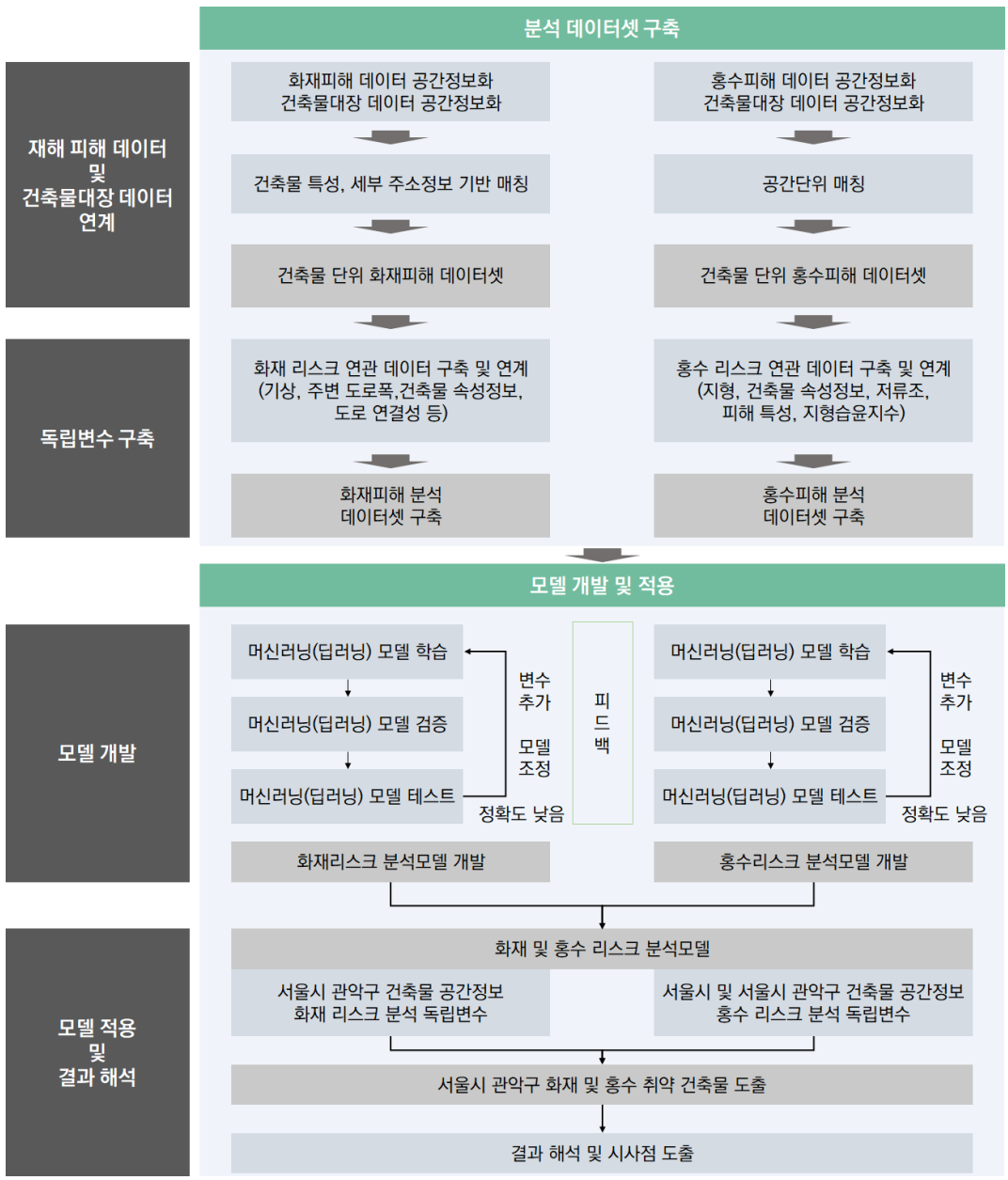
제3장 건축물 화재 및 홍수 리스크 분석 모델 개발



1. 모델 개발 절차 및 방향
2. 건축물 및 재해 데이터 연계
3. 건축물 화재 리스크 분석 모델 개발
4. 건축물 홍수 리스크 분석 모델 개발
5. 소결

1. 모델 개발 절차 및 방향

1) 모델 개발 절차



[그림 3-1] 건축물 화재 및 홍수 리스크 분석 모델 개발 프로세스

출처: 연구진 작성

건축물 화재 및 홍수 리스크 분석 모델은 크게 두 단계에 걸쳐 진행된다. 첫 번째 단계는 분석 데이터셋 구축이다. 분석모델 개발을 위한 데이터셋은 화재와 홍수 피해데이터를 각각 건축물대장 데이터와 연계하는 작업, 그리고 화재와 홍수 리스크 분석을 위한 독립변수를 생성하고 화재 및 홍수 피해 데이터와 연계하는 작업으로 구분할 수 있다.

두 번째 단계는 모델 개발과 적용이다. 모델 개발 과정은 분석 데이터셋을 학습, 검증, 테스트용 데이터셋으로 구분하고, 해당 데이터를 이용하여 모델을 학습, 검증, 테스트하는 단계를 거친다. 모델의 테스트 결과가 낮다면 변수를 추가하거나 모델을 조정하여 모델을 학습하는 과정으로 되돌아가 다시 학습을 시작하는 피드백 과정을 거친다. 이 과정을 통해 건축물 화재 및 홍수 리스크에 영향을 미치는 주요 원인 또한 도출된다. 충분한 정확도를 갖는 모델을 개발한 후 개발된 모델을 이용하여 서울시 건축물의 화재 및 홍수 리스크를 분석하는 과정을 거쳐 모델 개발을 완료한다.

2) 데이터 연계 방향

본 연구는 향후 건축·도시공간 데이터의 활용성 제고, 이기종 데이터의 효율적 연계 및 활용방법에 대해 검토하기 위해 두 가지 방법으로 데이터 연계를 수행하고자 한다.

첫 번째 방법은 2022년에 실시한 선행연구인 조영진 외(2022)의 데이터 연계 방법이다. 해당 방법을 화재발생 데이터와 건축물대장 데이터 연계에 적용하였다. 데이터 연계를 위해 1) 화재발생 데이터 지오코딩, 2) 오류데이터 정제, 3) 건축물대장 데이터 지오코딩, 4) 지오코딩 데이터 기반 자동 데이터 연계, 5) 연계되지 않은 데이터의 수동 연계의 다섯 단계 프로세스를 수행하였다. 이 방법론은 상대적으로 높은 정확도를 얻을 수 있다고 판단되나, 수동연계 프로세스에 과도한 수작업이 필요하여 데이터 연계의 효율성은 높지 않다. 따라서 최대한 자동화 프로세스를 도입하는 두 번째 방법을 고안하였다.

두 번째 방법은 홍수피해 데이터와 건축물대장 데이터 연계에 적용하고자 하였다. 각 데이터는 모두 지번주소와 도로명주소 정보를 기입하고 있으나, 주소정보 기입 주체가 달라 주소의 기입방식이 일부 다르고 지오코딩의 정확성도 서로 다를 것으로 판단하였다. 이에 본 연구는 각 데이터의 지번주소와 도로명주소 중 지오코딩 정확도가 높은 정보를 이용하여 각각 지오코딩 후 GIS 위치정보를 기반으로 데이터를 매칭하는 자동매칭 방식을 적용하고자 한다.

본 연구에서 두 가지 서로 다른 데이터 연계 방법을 적용해봄으로써 향후 구축될 다양한 이기종 데이터의 연계방안 마련의 지침이 될 수 있을 것으로 기대한다.

3) 대상지 선정 및 특성

본 연구에서 개발하고자 하는 리스크 분석 모델은 다양한 도시 및 지역에 적용이 가능해야 한다. 따라서 연구 대상지는 다양한 국내 도시 특성을 포함하는 지역으로 선정하고자 한다. 범용성 확보를 위해 다양한 인문·사회적 특성 및 물리·환경적 특성을 고려할 필요가 있다. 즉, 거주인구 규모가 충분히 크고, 다양한 유형의 가구로 구성될 필요가 있다. 또한, 다양한 용도, 규모, 형태, 구조, 노후도의 건축물이 있고, 주거·상업·공업지역이 다양하게 분포하는 등 복합적 특성을 모두 포함할 필요가 있다.

본 연구에서는 이같은 특성을 나타내고, 분석을 위한 가용 데이터가 풍부하며, 분석 결과의 범용적 적용이 가능할 것으로 판단되는 서울시를 대상으로 연구를 수행하였다. 서울시는 약 1,000만 명의 인구가 거주하고 약 60만동의 건축물이 있는 우리나라의 수도로, 건축물 화재 및 홍수 피해 발생 이력을 다수 보유하고 있어 본 연구를 수행하기 적합한 대상지로 선정되었다.

□ 서울시 건축물 화재 피해 발생 현황

서울시 건축물 화재피해 데이터는 세부적으로 사망 및 부상 등 인명피해, 화재발생 재산피해액, 화재발생 년, 월, 일 및 시각, 화재발생 건축물 주소, 발화열원 대분류 및 소분류, 발화요인, 연소확대물, 소실면적, 장소, 발화지점 등 항목이 기입되어있다. 본 연구에서는 2017년부터 2021년까지 5년간의 데이터를 사용하였다.

5년간 서울시 화재피해가 접수된 데이터는 총 28,266건이다. 연간 평균 5,653.2건이 접수되고 있으며, 2018년이 6,368건으로 피해접수 건수가 가장 많다. 발화요인은 전기적 요인과 자연적 요인을 포함하여 총 10가지로 구분하고 있으며, 부주의가 57.4%로 가장 높고, 자연적인 요인이 가장 낮게 나타나고 있다. 두 번째와 세 번째로 높은 발화 요인은 건축물의 노후화와 관련 있는 전기적 요인 및 기계적 요인이며, 이로 인한 화재는 전체의 29.38%를 차지하였다.

[표 3-1] 2017-2021년 서울시 발화요인별 화재발생 건수

구분	전기적 요인	기계적 요인	화학적 요인	가스 누출	교통 사고	부주의	기타	자연적 요인	방화 (명확/의심)	발화 요인 미상	합계
건수	6,789	1,515	208	108	135	16,224	202	39	611	2,435	28,266

출처: 서울 열린데이터 광장. https://data.seoul.go.kr/dataList/10167/S/2/datasetView.do. 2022.06.22. 접속

화재발생 장소별로 구분해보면 전체 28,266건 중 주거공간에서 11,717건, 비주거공간에서 10,327건, 운송 및 임야/기타 공간에서 6,222건이 발생한 것으로 나타났다. 즉, 주거공간의 비중이 상대적으로 높으며, 주거공간 중에서는 공동주택의 비율이 가장 높고, 비주거공간 중에서는 음식점의 화재발생 비율이 높은 것으로 나타난다.

[표 3-2] 2017-2021년 용도별 서울시 화재피해 건수

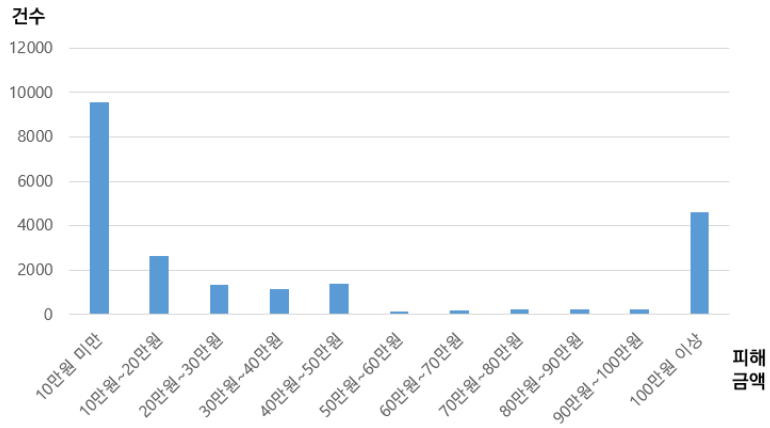
구분	비주거										
	학교	일반 업무	판매 시설	숙박 시설	종교 시설	의료 시설	공장/창고/작업장	위락 오락 시설	음식점	일반 서비스 시설	기타
건수	177	1,933	984	233	175	205	915	265	3,036	1,190	1,214

구분	주거			운송 (차량철도 등)	임야/기타	합계
	단독주택	공동주택	기타주택			
건수	4,666	6,912	139	1,998	4,224	28,266

출처: 서울 열린데이터 광장. https://data.seoul.go.kr/dataList/318/S/2/datasetView.do. 2022.06.22. 접속

피해금액의 경우 평균 약 699만원 수준으로 나타나며, 최대 금액은 716억원[24]으로 나타난다. 대부분의 피해금액은 0원에서 20만원 사이에 있는 것으로 나타나며, 해당 범위 내에 총 9,551건이 포함되어있다. 전반적으로 10만원 초과 피해액은 피해금액이 커질수록 감소하는 추세를 나타내며, 피해금액 100만원 이상의 화재피해도 다수 발생하는 것으로 나타났다.

24) 2019년 제일평화시장 화재



[그림 3-2] 피해금액대별 화재피해발생 건수
출처: 연구진 작성

□ 서울시 건축물 홍수 피해 발생 현황

서울시 건축물 홍수피해 데이터는 세부적으로 지번주소, 도로명주소, 세대원수, 피해종류, 피해물량, 피해금액, 피해구분, 피해일시 등 항목이 기입되어있다. 본 연구에서는 2016년부터 2022년까지 7년간의 데이터를 사용하였다.

7년간 홍수피해가 접수된 건수는 총 30,184건이다. 홍수 피해요인은 강풍 등 13가지로 구분된다. 가장 많은 피해요인은 침수로, 일반 침수를 포함한 6가지 유형의 침수는 총 29,468건, 일반 침수는 22,604건으로 나타난다. 일반 침수 22,604건은 홍수 피해 사례의 기입 담당자가 달라 세부적인 침수 유형을 구분하지 않았을 것으로 판단된다. 또한, 침수 유형 중 공공하수도 역류 및 개인하수도 역류 등 저지대 침수가 아닌 침수가 약 3,100건으로 전체의 최소 10%정도를 차지하는 것으로 나타난다.

[표 3-3] 2016-2022년 피해요인별 서울시 홍수피해 건수

구분	강풍	건물붕괴	산사태	유실	축대붕괴	기타	하천범람
건수	8	8	2	6	12	675	5

구분	침수					일반 침수	합계
	공공하수도역류	개인하수도역류	도로범람유입	저지대침수	기타		
건수	3,047	82	1,635	854	1,246	22,604	30,184

출처: 연구진 작성

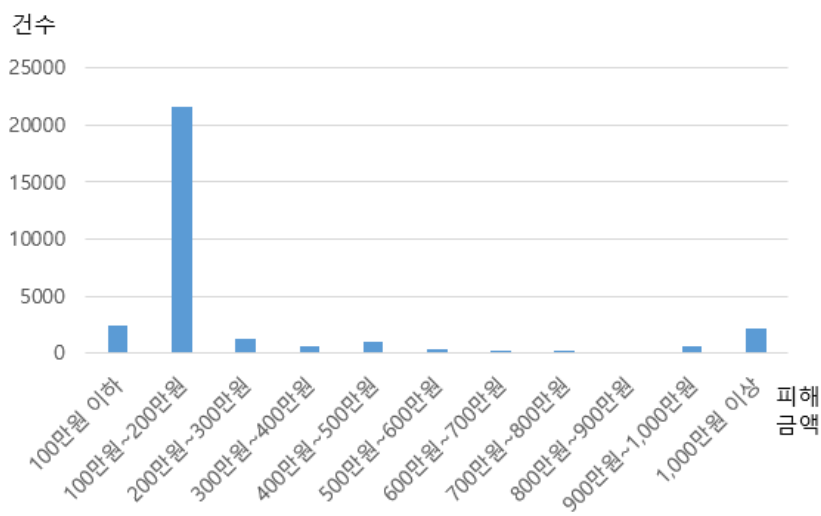
피해금액의 경우 평균 약 654만원 수준으로 나타나며, 최대 금액은 80억원으로 나타난다. 대부분의 피해금액은 100만원에서 200만원 사이에 있는 것으로 나타나며, 해당 범위 내에 총 21,580건이 포함되어있다.

또한, 전체적인 추세를 보면 피해금액이 200만원을 넘어선 이후 800만원에서 900만원 사이까지는 접수 건수가 줄어들고 있으나, 900만원 이후부터 다시 증가하고 있는 것으로 나타난다.

[표 3-4] 2016-2022년 피해금액대별 서울시 홍수피해 건수

구분 (만원)	0~ 100	100~ 200	200~ 300	300~ 400	400~ 500	500~ 600	600~ 700	700~ 800	800~ 900	900~ 1000	1000 ~
건수	2,363	21,580	1,273	645	959	302	259	260	117	626	2,126

출처: 연구진 작성



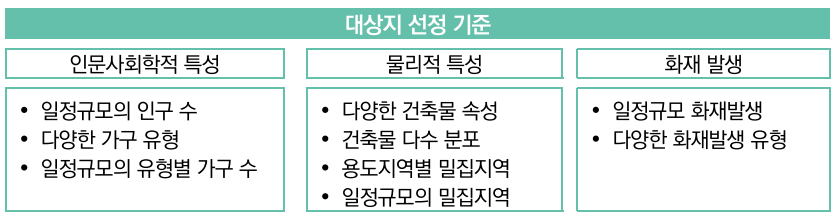
[그림 3-3] 피해금액대별 홍수피해발생 건수

출처: 연구진 작성

2. 건축물 및 재해 데이터 연계

1) 건축물 화재 발생 데이터 공간정보화[25)]

2022년 진행한 선행연구(빅데이터 기반 건축물 화재 예측 모델 개발 연구)에서는 인문사회학적 특성, 물리적 특성이 도시의 범용성을 적용하기 충분하고, 빅데이터 분석을 수행을 위해 화재발생 건수가 충분한 지역을 대상지로 선정하여 화재 리스크 분석 모델을 개발하였다.



[그림 3-4] 대상지 선정기준

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.53

서울특별시는 다양한 유형의 일정규모 이상의 인구와 가구로 구성되어 있으며, 노후건축물과 신축건축물이 혼재되어 있고, 다양한 용도지역과 주거·상업·공업 등 용도별 밀집지역을 포함하여 도시의 범용성을 확보하고 있다. 그리고 분석에 충분한 화재발생 건수를 보유하고 있어 대상지로 적합하다.

최근 5년(2017~2021년)의 서울시의 화재 발생 현황을 살펴보면, 총 28,266건으로, 발화요인으로는 부주의에 의한 화재발생이 가장 많고 건축물 노후화에 따른 설비의 노후화, 작동이상으로 추정되는 전기적 및 기계적 요인에 의한 화재는 8,304건으로 29.38%에 달한다.

[표 3-5] 2017-2021년 서울시 발화요인별 화재발생 건수

구분		2017	2018	2019	2020	2021	합계	비율
실화	전기적요인	1,315	1,525	1,322	1,265	1,362	6,789	24.02%
	기계적요인	364	379	283	241	248	1,515	5.36%
	화학적요인	39	43	43	36	47	208	0.74%
	가스누출	23	27	22	14	22	108	0.38%
	교통사고	29	25	37	18	26	135	0.48%
	부주의	3,555	3,639	3,502	2,890	2,638	16,224	57.40%
	기타	21	24	42	59	56	202	0.71%
자연적인요인		10	12	4	5	8	39	0.14%
방화(명확/의심)		123	167	136	98	87	611	2.16%
발화요인(미상)		499	527	490	462	457	2,435	8.61%
합계		5,978	6,368	5,881	5,088	4,951	28,266	100%

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.56

25) 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구를 선행연구로 재작성

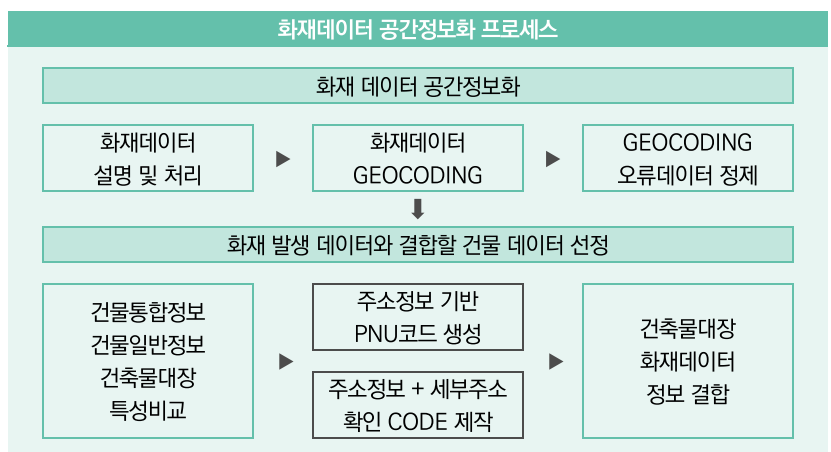
장소별로는 주거에서 발생한 화재가 11,717건(41.15%)으로 가장 많고, 비주거 건축물에서는 음식점이 3,036건(10.74%)으로 가장 많은 화재가 발생했다.

[표 3-6] 2017-2021년 서울시 장소별 화재발생 건수

구분		2017	2018	2019	2020	2021	합계	비율
주거	단독주택	1,016	1,073	983	815	779	4,666	16.51%
	공동주택	1,520	1,669	1,409	1,177	1,137	6,912	24.45%
	기타주택	21	22	30	40	26	139	0.49%
비주거	학교	39	30	32	35	41	177	0.63%
	일반업무	411	438	383	329	372	1,933	6.84%
	판매시설	160	207	205	205	207	984	3.48%
	숙박시설	33	63	58	34	45	233	0.82%
	종교시설	39	48	34	22	32	175	0.62%
	의료시설	45	50	40	29	41	205	0.73%
	공장/창고/작업장	208	185	209	151	162	915	3.24%
	위락오락시설	51	65	46	58	45	265	0.94%
	음식점	681	682	608	571	494	3,036	10.74%
	일반서비스시설	255	294	271	176	194	1,190	4.21%
	기타	279	261	259	201	214	1,214	4.29%
운송(차량철도 등)		468	416	394	338	382	1,998	7.07%
임야/기타		752	865	920	907	780	4,224	14.94%
합계		5,978	6,368	5,881	5,088	4,951	28,266	100%

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.56

화재 리스크 분석 모델 개발을 위해 GIS 건물통합정보에 건축물 속성정보 데이터와 화재발생 데이터의 정보 제공 수준을 분석하여 연계 최소단위를 선정하고 연계하여 화재발생 데이터를 공간정보화 하였다.

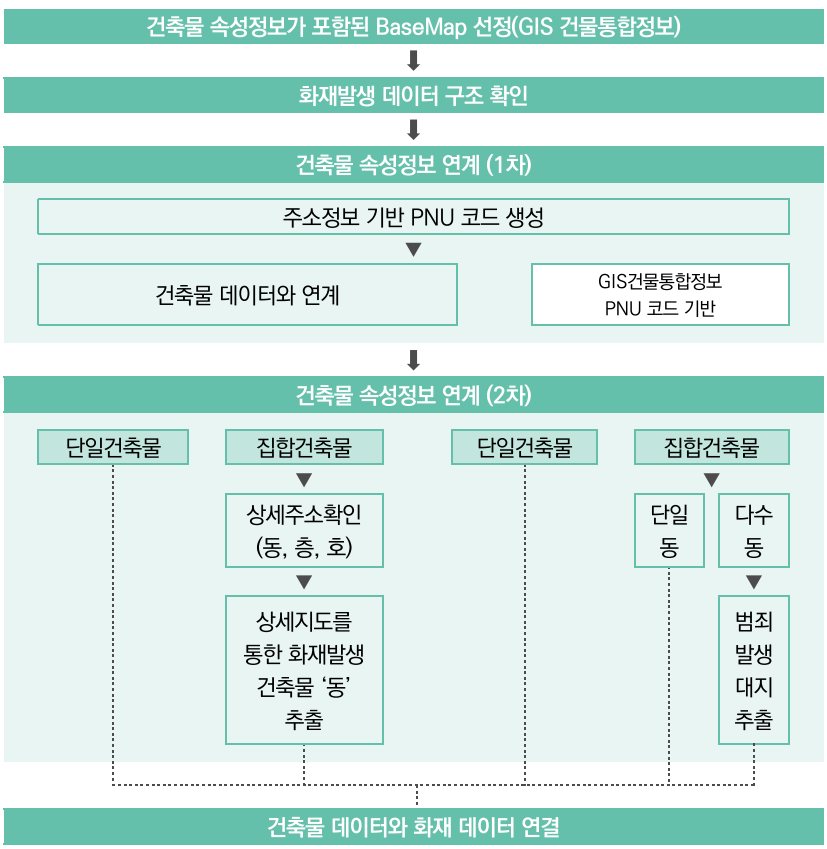


[그림 3-5] 건축물 화재 발생 데이터 구축 프레임워크

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.59

건축물 속성정보는 국토교통부의 건축행정시스템(세움터)의 건축물대장 데이터 중 안전사고와 연관성을 보일 것으로 판단되는 용도, 연면적, 높이, 층수, 노후도, 건폐율 등을 활용하였다. 화재발생 데이터에는 발생시점(연도, 일시), 발생지점(시도, 시군구, 법정도, 상세주소), 발화열원, 발화요인, 화재장소(용도), 발화지점, 소실면적(㎡), 인명피해(사망, 부상자 수), 재산피해액이다.

화재발생 데이터의 공간정보화는 먼저, 화재발생 데이터의 주소정보를 기반으로 PNU코드[26]를 생성하여 GIS 건물통합정보와 연계하고, 집합건축물 또는 다수의 동이 하나의 지번에 있는 건축물(학교, 아파트 등)은 상세지도를 이용하여 화재가 발생한 건축물 '동'과 연계한다. 이후 공간정보화된 화재발생 데이터와 건축물 속성정보를 결합하여 하나의 통합 데이터를 구축하였다.



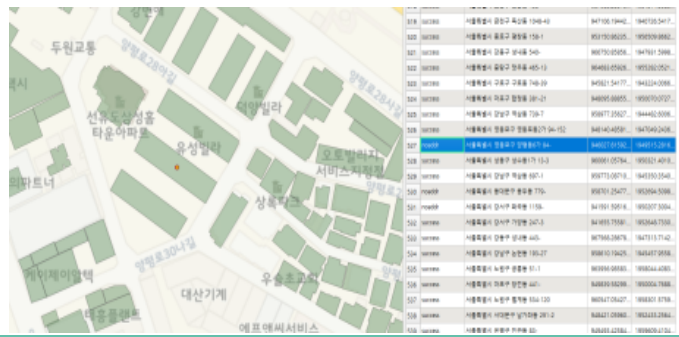
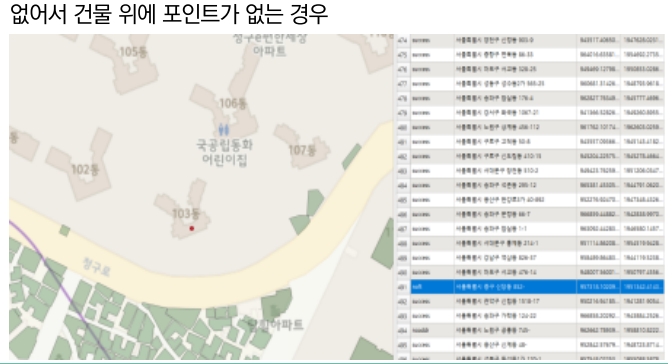
[그림 3-6] 건축물 데이터와 화재 데이터 연계 절차

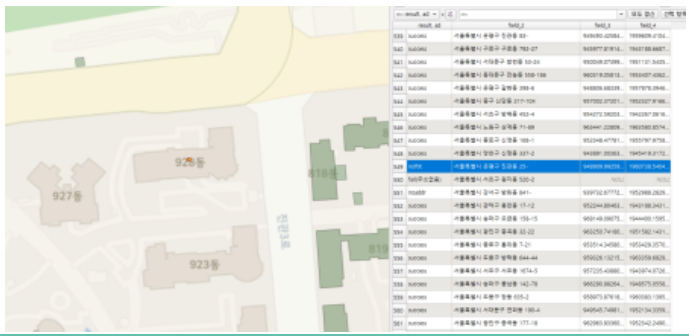
출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.57

26) 총 19자리의 숫자로 구성된 코드로 00(시·도)000(시·군·구)000(읍·면·동)00(리)0(대장구분)0000(본번)0000(부번)으로 이루어져 있다. 앞의 8자리는 행정안전부에서 제공하는 주민등록 주소 코드로 PNU코드 생성 이전 변경 사항을 확인하여야 한다.

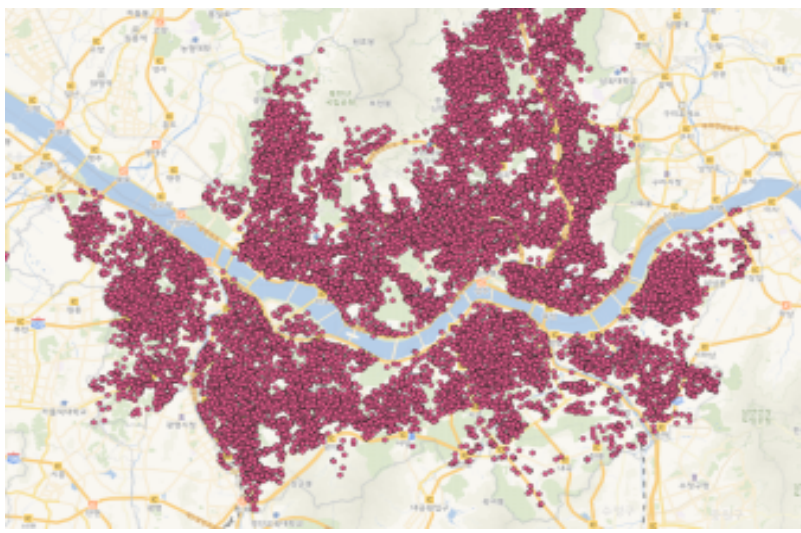
화재발생 데이터를 지오코딩 한 결과 총 21,055개 데이터 중 4,307개 데이터에서 오류가 발생하였으며, 4개의 유형으로 분류된다. 오류 포인트는 직접 확인하여 위치를 정확한 건축물 동으로 옮겨주는 Point relocation 작업을 수행하여 약 93%의 화재발생 데이터를 건축물 단위의 공간정보 데이터로 구축했다. 이후 건축물 데이터와 결합하였다.

[표 3-7] 화재데이터 지오코딩에 따른 오류구분 및 예시

오류구분	오류예시
noaddr	화재 데이터의 주소 지번이 끝까지 나오지 않거나, 주소가 정확하지 않아서 지오코딩이 되었지만 다른 곳에 표시된 경우 
fail	도로 위 혹은 공사장 등 주소가 있지만 지오코딩 오류로 포인터가 생성되지 않은 경우
noft	주소는 정확하여 포인트는 본 위치에 있으나 건물통합정보 데이터의 건물 shp이 없어서 건물 위에 포인트가 없는 경우 

noftx	주소도 불분명하고, 건물통합정보 데이터의 건물 shp도 없는 경우 

출처: 조영진 외(2022). 빅데이터 기반 건축물 화재 예측 모델 개발 연구. pp.60-61



[그림 3-7] 화재 데이터의 공간 정보화

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.62

화재 데이터와 건축물대장 데이터 연계시 건축물의 멸실을 고려하여 화재발생시기 기준, 전년도 4분기 건축물대장 데이터와 매칭을 진행하였다.

화재 데이터와 건축물 속성정보 데이터 연계 프로세싱

1) 단일건물인 경우, 바로 화재데이터에 건축물대장 데이터 결합
2) 집합건물에서 해당 주소의 건물이 한 동인 경우, 데이터 결합
3) 집합건물/건축물대장의DONG_NM(동이름)이 화재 세부주소와 일치하는 경우 해당 데이터 결합
4) 부수적으로 건축물대장 상 동 이름이 '제00동' 형식, 'A동'을'에이동'으로 기재한 경우 등 의미에 부합하는 해당 동을 찾아 결합
5) 화재데이터 세부주소에서 동 이름을 숫자만 기재한 경우 해당 동이 있는지 찾아서 결합
6) 연결되지 않을시 화재발생 당해연도 건축물 연결
7) 비매칭결과 에러 처리코드

출처 : 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.63

화재 데이터와 건축물대장 데이터의 매칭 결과 총 21,055개 데이터 중 18,624개 데이터가 결합되었다. 일부 표기 방법에 따른 차이 등과 같은 원인에 따른 매칭 실패 데이터는 Revision으로 분류하여 모든 데이터를 직접 확인하여 수정했다. 최종적으로 18,814개(약 89.4%)의 데이터가 정제 및 매칭에 성공했다.

[표 3-8] 화재 건축물 연계 데이터 최종 매칭 결과

분류	재확인 분류	개수	퍼센트
Success	Success(성공)	18,624	88.5%
	r_Success(재매칭 성공)	190	0.9%
성공 총 합계		18,814	89.4%
Fail	NoDJ(건축물 대장X)	595	2.8%
	r_NoDJ(재매칭 – 건축물 대장X)	16	0.07%
	r_Remove(재매칭 – 주소 오류등)	174	0.82%
	필드값 없음(화재 데이터 오류 등)	1,456	6.9%
실패 총 합계		2,241	10.6%
전체 합계		21,055	100%

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.64

[표 3-9] 화재 데이터 및 건축물 대장 매칭 결과 칼럼Code 및 설명

칼럼Code	설명
success	건축물대장과 주소가 잘 매칭된 경우
NoDJ	건축물 대장에 해당 주소 건물이 없는 경우
R_NoDJ	재확인 (건축물 대장 앞뒤 년도) 시에도 건축물 대장에 정보가 없는 경우
R_success	재확인 시 건축물 대장의 정보를 발견한 경우 (예 : 세부주소에서는 'A동' 건축물대장에서는 '에이동'으로 나와 매칭이 안된 경우 등을 모두 다 확인)
R_remove	noaddr, noftX 등 세부 주소가 잘못 되어 있는데 데이터셋에 포함되어 있어서 건축물 대장과 매칭이 어려운 경우로 분석 진행시 삭제하면 되는 칼럼
fail	필드값 없음
noaddr, noftX	주소 이상 혹은 지오코딩이 되지 않은 데이터 등은 건축물 대장과 매칭할 수 없기 때문에 이런 데이터들이 모두 비어 있는 채로 있는 경우

출처: 조영진 외. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. p.65.

화재발생 데이터와 건축물 속성정보 데이터를 연계하여 구축한 화재발생 건축물의 공간단위 통합데이터는 위치 좌표(x, y), 화재데이터 정보, PNU코드, 건축물 대장 정보 등으로 구성된다. 이 때, 위치 좌표는 정확한 주소를 기반으로 포인트를 갱신한 좌표이다.

2) 건축물 홍수피해 데이터 공간정보화

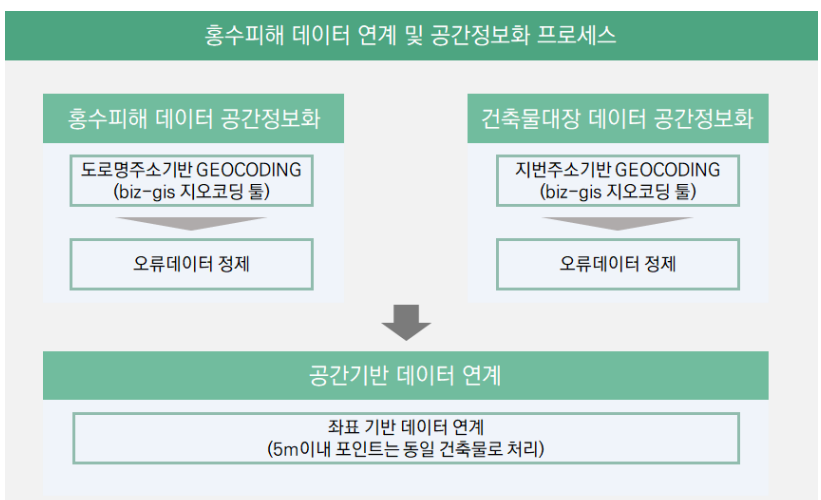
① 건축물 홍수피해 데이터 구축 프로세스

건축물 홍수피해 데이터의 공간정보화 및 건축물대장과의 연계는 크게 세 단계로 구분하여 진행하였다.

첫째로, 홍수피해 데이터의 공간정보화를 수행하였다. 홍수피해 데이터는 도로명주소와 지번주소를 모두 포함하고 있으며, 지번주소보다는 도로명주소가 지오코딩 시 높은 정확도를 나타내는 것으로 확인되어 도로명주소를 기반으로 지오코딩하였다.

둘째로, 건축물대장 또한 지번주소를 이용하여 지오코딩하였다. 건축물대장은 도로명주소와 지번주소가 모두 포함되어있다. 그러나 도로명주소의 정확성을 담보하기 어렵고, 지오코딩에 사용한 biz-gis 지오코딩 툴이 건축물 동 명칭과 같은 세부 주소정보를 활용하여 정확도를 높일 수 있는 것으로 확인되어 지번주소를 이용하였다.

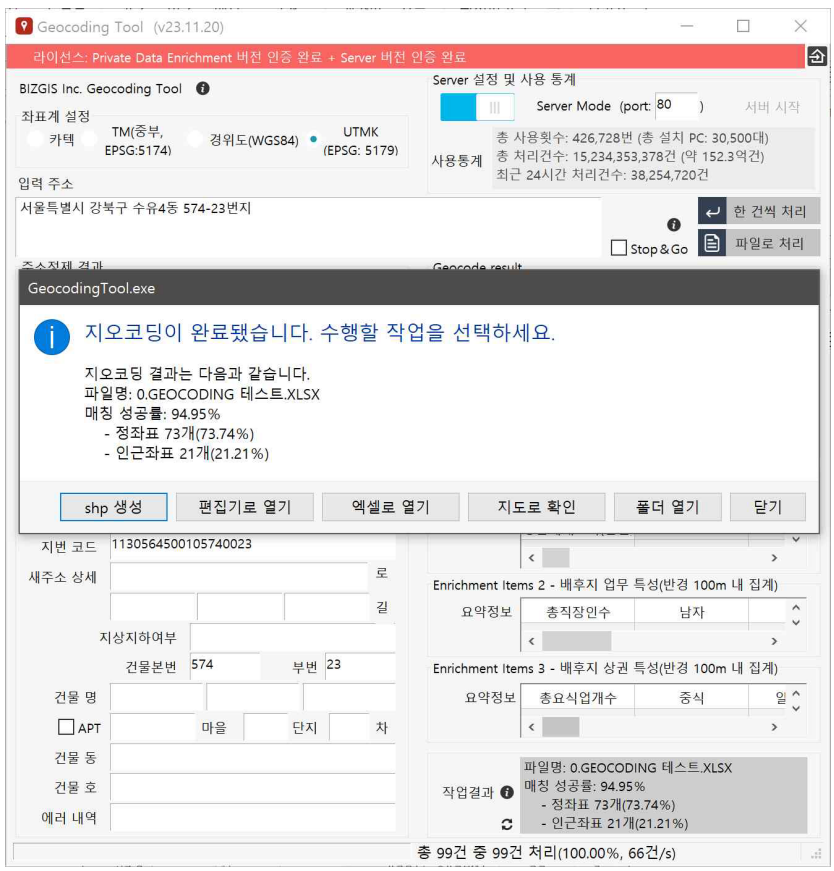
마지막으로, 지오코딩된 홍수피해 데이터와 건축물대장 데이터를 GIS 툴을 이용하여 매칭하였다. 지오코딩된 각 데이터는 좌표를 갖는 점 데이터 형태이다. 홍수피해 건축물과 건축물대장 상의 건축물이 동일한 건축물일 경우 각 점의 위치가 동일하거나 가까울 것이라는 가정하에 거리기반 데이터 연계를 수행하였다. 데이터 연계 후 최종 검토를 통해 오류 데이터는 제거하였다.



[그림 3-8] 건축물 홍수피해 데이터 구축 프레임워크
출처: 연구진 작성

② 홍수피해 데이터 주소정보 정제 및 지오코딩

홍수피해 데이터의 지오코딩을 위해 biz-gis의 지오코딩툴을 이용하였다. 지오코딩툴은 주소정보를 X, Y 좌표값으로 변환해주며, 이를 위해 주소정보의 정제가 필요하다. 홍수피해 데이터의 주소정보 항목은 주소(지번), 주소(도로명), 피해위치(지번), 피해위치(도로명)의 네 개로 구성되어있다. 주소 열에 적힌 주소와 와 피해위치 열에 적힌 주소는 대부분 일치한다. 그러나 일치하지 않는 데이터를 검토한 결과, 주소 열에 적힌 주소 중에는 서울시가 아닌 것도 있어 이는 피해자의 거주지인 것으로 추측되며, 피해위치 열에 적힌 주소가 건축물 홍수피해 발생 위치인 것으로 확인할 수 있다. 이에 따라 홍수피해 데이터 지오코딩에는 피해위치 열에 기입된 주소정보를 사용하였다.



[그림 3-9] biz-gis 지오코딩툴의 지오코딩 결과 화면 예시

출처: BIZ-GIS Inc.. nd. Geocoding Tool 매뉴얼. manual ver1. p.8

피해위치 지번주소 정보는 서울특별시까지 모두 동일하며 뒤이어 구 명칭, 동 명칭, 본번, 부번으로 이루어져있다. 다만 뒤이어 건축물 동명칭정보, 호실정보, 층수정보 및 기타정보가 무작위로 연결된 경우가 있어 이에 대한 정제가 필요하다. 도로명주소 또한 서울특별시까지 모두 동일하며 뒤이어 도로명과 건물번호로 이루어져있으나 지번주소와 동일하게 뒤이어 동명칭정보, 호실정보, 층수정보 및 기타정보가 무작위로 연결된 데이터가 존재한다. 이에 따라 지오코딩을 위한 주소 정제 작업을 수행하였다.

도로명주소는 도로명과 건물번호로 이루어져있어 지번주소보다 개별 건축물의 위치를 더 정확하게 나타낸다. 이에 최종적으로 지오코딩은 피해위치(도로명) 컬럼의 주소를 정제하여 수행하였다. 지오코딩에 사용된 biz-gis의 지오코딩툴은 입력된 주소 위치의 X, Y 좌표를 반환하며, 좌표체계를 선택할 수 있다. 또한 지오코딩 수행 후 매칭 성공률(%)과 정좌표 및 인근좌표 선택 여부를 제공한다.

홍수피해 데이터의 지오코딩 결과 총 30,184개 홍수피해 데이터 중 약 99%가 매칭에 성공하였다. 정좌표 29,728건, 도로중심좌표 14건, 인근좌표 1건, 정제실패 2건으로 나타났다.

[홍수피해 정보] [건축물 대장]

[그림 3-10] biz-gis 지오코딩툴의 지오코딩 결과
출처: 연구진 작성

③ 건축물대장 데이터 지오코딩

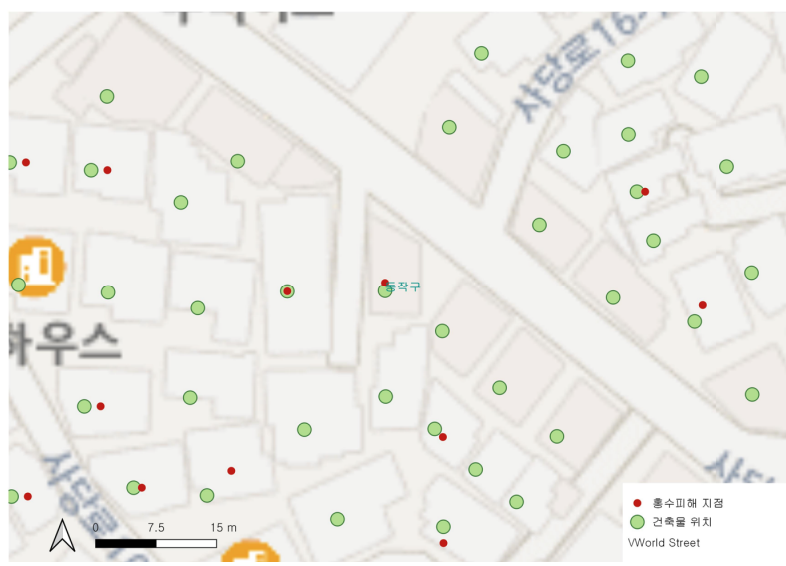
건축물대장 데이터의 지오코딩 또한 biz-gis의 지오코딩툴을 이용하였다. 건축물대장에는 지번주소와 도로명주소가 모두 포함되어있다. 지오코딩을 위한 건축물대장 자료는 2022년 4분기 시점의 자료를 사용하였다.

건축물대장의 주소정보는 지번주소를 사용하였다. 건축물대장은 도로명주소가 도입되기 이전 사용승인된 건축물의 경우 도로명주소가 없거나 부정확한 경우가 많아 지번주소를 사용하였다. 건축물대장의 주소정보는 시도, 시군구, 법정동, 본번, 부번의 형태로 저장되어있으며, 추가적으로 건축물 동 명칭 정보가 있다. 지오코딩을 위해 주소정보를 하나로 구성하고 biz-gis의 지오코딩툴에 입력하였다.

2022년 4분기 기준 건축물대장의 서울시 전체 건축물은 596,244동이며, 지오코딩에 따른 매칭성공률은 약 94%로 나타났다. 결과적으로 정좌표 559,307동, 인근좌표 69동, 도로중심좌표 99동, 정제실패 36,769동으로 나타났다.

④ 데이터 연계

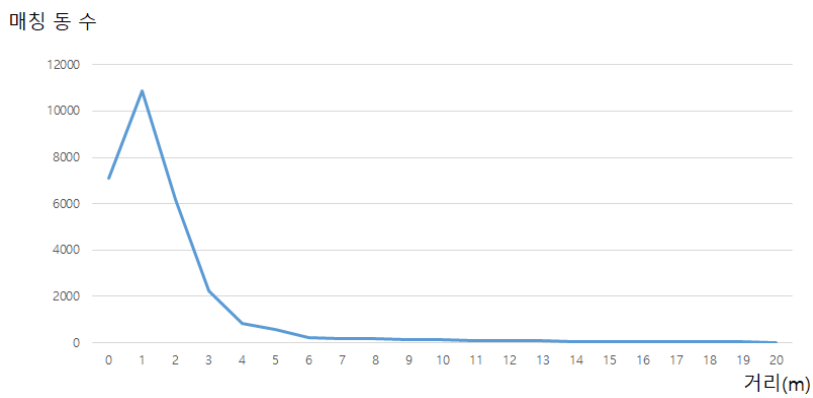
지오코딩이 완료된 홍수피해 데이터와 건축물대장 데이터를 확인하기 위해 홍수피해 데이터는 빨간색으로, 건축물대장 데이터는 초록색으로 표시하였다. 도로명주소를 이용하여 지오코딩한 홍수피해 데이터와 지번주소를 이용하여 지오코딩한 건축물대장 데이터는 같은 건물이라도 거리가 이격되는 경우가 발생하는 것을 확인할 수 있다. 이 같은 경우 같은 건물임에도, 지오코딩 툴에 저장된 도로명주소와 지번주소의 좌표 차이로 인해 정확히 매칭되지 않는다. 이에 따라, 데이터 연계를 위하여 공간기반 데이터 조인 방법을 적용하였다.



[그림 3-11] 홍수피해 및 건축물대장 지오코딩 결과
출처: 연구진 작성

홍수피해 데이터와 건축물대장 데이터를 연계하기 위해 QGIS의 최근접거리 기반 속성 결합(Join Attributes by Nearest) 기능을 사용하였다. 해당 기능은 입력한 두 벡터 레이어에서 가장 근접한 객체를 찾아 속성정보를 결합시켜준다. 즉, 홍수피해 지점을 기준으로 가장 가까운 건축물을 찾고, 홍수피해 데이터에 건축물의 속성정보를 연계시켜준다.

총 30,184개 홍수피해 지점 중 27,231개 지점이 건축물과 5m 이내 거리에 있는 것으로 나타났으며, 1,272개 지점은 5m 초과 10m 이내, 1,681개 지점은 10m 초과된 거리의 건축물과 매칭되었다. 본 연구는 5m 이내에 있는 건축물과 매칭된 홍수피해 지점을 매칭성공으로 보고 총 27,231개 데이터를 이용하여 분석을 수행하고자 하였다.

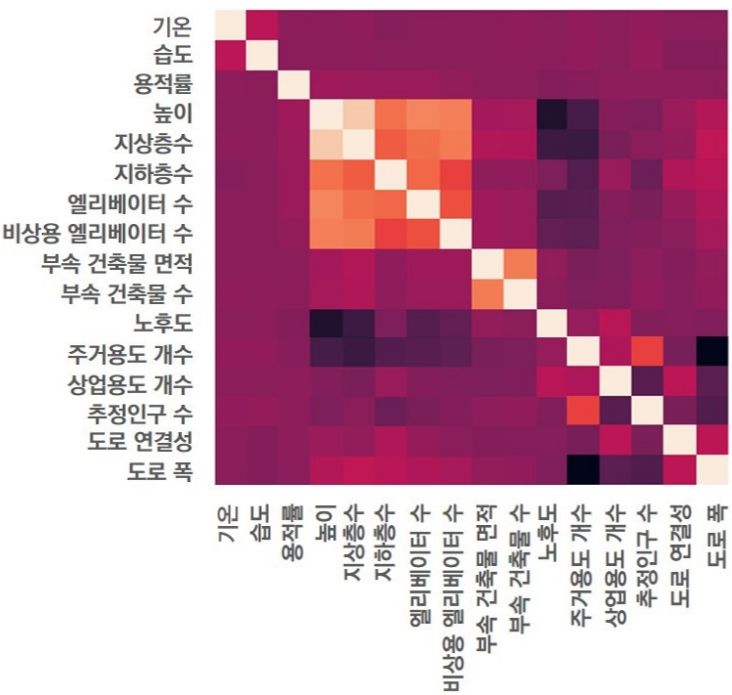


[그림 3-12] 홍수피해 및 건축물대장의 공간기반 연계 결과
출처: 연구진 작성

3. 건축물 화재 리스크 분석 모델 개발

1) 입력 데이터 선정

선행연구를 기반으로 화재에 영향을 미치는 변수들을 추출하였으며, 상관관계 분석에 따라 상관성이 0.7 이상인 변수들을 제거하였다. 보편적으로 상관성이 0.7 이상이면 매우 높은 상관성을 보인다고 해석하기 때문에, 본 연구에서는 상관성 0.7 이상인 변수를 제거하고 진행하였다. 상관성이 높은 변수들을 제거하는 이유는 중복되거나 매우 유사한 정보를 담고 있는 변수들을 제거하여 모델의 복잡성을 줄이고 분석 성능을 개선하기 위함이다.



[그림 3-13] 건축물 화재 변수 간의 상관관계 분석
출처: 연구진 작성

"건폐율-용적률"과 "전체 추정 인구 수-연령별 추정인구 수"는 상관성이 0.7 이상으로 도출되어 변수 선택에서 건폐율과 연령별 추정인구 수는 제거하였다. 건폐율과 용적률 두 변수 모두 도시 계획 및 건물 설계에서 중요한 역할을 하는 변수이다. 건폐율은 대지면적에 건물을 지을 수 있는 비율을 백분율로 나타낸 것으로, 건물의 바닥면적을 의미한다고 해석된다. 반면 용적률은 전체 대지면적에 대한 건물 연면적의 비율을 의미한다. 용적률을 선택한 이유는 용적률이 건폐율을 포함하는 개념으로 보다 포괄적인 정보를 제공하며, 건물의 높이와 밀도 정보를

포함할 가능성이 높기 때문이다. 또한, 각각의 연령별 인구수 대신 전체 추정 인구수 변수를 채택한 이유는 각 연령 변수들을 모두 사용한다면 특정 연령 그룹의 효과를 더 상세히 파악할 수 있지만 데이터의 차원이 높아지게 되는 단점이 있고, 지역의 전반적인 인구 규모와 동향을 이해하기 위해서는 전체 추정인구 수 변수를 사용하는 것이 더욱 효과적일 것으로 판단하였기 때문이다.

최종적으로 연구에 사용한 독립변수들은 크게 세 가지의 특성으로 구분할 수 있다. 첫째로 건축 및 도시환경 특성에 속하는 변수, 둘째로 건축물 자체 특성에 속하는 변수, 셋째로 인구 통계 및 날씨 특성에 속하는 변수이다. 종속변수로는 건축물 화재 리스크를 고려하기 위해 화재피해액을 활용하였다. 또한, 모든 독립변수는 화재 발생 관련 변수, 피해금액 관련 변수, 화재발생과 피해금액 모두 관련된 변수로 구분할 수 있다.

건축 및 도시환경 특성에 속하는 변수는 도로 연결성, 도로 폭, 주거 건물 수, 상업 건물 수가 사용되었다. 도로연결성은 건물이 도로 세그먼트에 얼마나 연결되어 있는지를 나타내는 변수로, 1,000m 반경 내 도로 수를 고려한다. 변수 구축에는 공간분석 툴 중 Depthmap X를 활용하였다. 도로 폭은 주변 도로의 평균 폭을 나타내는 변수로, 400m 반경 내의 도로 폭을 고려하여 변수로 사용하였다. 주거용도 수는 건물 중 주거 용도의 건물 수를 나타내는 변수로, 400m 반경 내 주거 건물 수를 고려하여 사용하였다. 상업용도 수는 인접 건물 중 상업용도 건물 수를 나타내는 변수로, 400m 반경 내 상업 건물 수를 고려하여 변수화 하였다.

건축물 자체 특성에 속하는 변수는 건축물의 노후도, 지상 및 지하 층 수, 용적률, 승강기 수, 비상용 승강기 수를 사용하였다. 노후도는 건물 사용승인일로부터 경과한 년수를 나타내는 변수이다. 지하 및 지상 층 수는 건물의 지상 및 지하층수를 나타내는 변수로, 건물의 규모를 반영하는 변수로 채택하였다. 용적률은 건물의 용적률을 나타내는 변수로, 건물의 사용 목적과 크기를 반영하기 위해 사용하였다. 승강기 수는 건물에 설치된 승강기의 수를 나타내는 변수로, 건물 내 이동수단을 고려하기 위해 사용하였다. 비상용 승강기 수는 건물에 설치된 비상용 승강기의 수를 나타내는 변수로, 비상 상황 대비를 고려하기 위해 사용하였다.

인구 통계 및 날씨 특성에 속하는 변수는 기온, 습도, 풍속, 추정인구 수가 최종 채택되었다. 기온, 습도, 풍속은 화재 리스크 분석에 영향을 미치는 날씨 변수 중 채택된 것으로 날씨 및 대기 정보를 반영하는 변수로 최종 선택하였다. 추정인구 수는 주변 지역의 추정 인구 수를 나타내는 변수로, 주변 환경의 인구밀도를 반영하는 변수로 채택되었다.

[표 3-10] 화재발생 건축물 데이터셋의 독립변수

변수 특성	변수 명	변수 설명	단위	발생 관련	피해액 관련
건축 및 도시 환경 특성	도로 연결성	도로 세그먼트에 연결된 도로의 수 (버퍼 1,000m)	count		○
	도로 폭	평균 도로 폭(버퍼 1,000m)	m		○
	주거용도 수	주거 용도의 건물 수(버퍼 400m)	count		○
	상업용도 수	상업 용도의 건물 수(버퍼 400m)	count		○
건축물 자체 특성	노후도	건물 사용승인일로부터 경과한 년수	count	○	
	지상 층 수	건물의 지상 층 수	count	○	○
	지하 층 수	건물의 지하 층 수	count	○	○
	용적률	건물의 용적률	%		○
	승강기 수	건물의 승강기의 수	count		○
	비상용 승강기 수	건물의 비상용 승강기의 수	count		○
인구 통계 및 날씨 특성	기온	화재가 발생한 시각의 온도	˚C	○	
	습도	화재가 발생한 시각의 습도	%	○	
	풍속	화재가 발생한 시각의 풍속	m/s		○
	추정인구 수	추정 인구 수(버퍼 400m)	count		○

출처: 연구진 작성

2) 데이터 가공

데이터 가공은 화재 리스크 분석 모델을 개발하는 핵심 단계 중 하나로, 선정된 변수들을 정제하고 가공하여 모델 학습에 적합한 형태로 변환하는 과정이다. 이러한 작업은 모델의 분석 성능을 향상시키고 모델이 데이터를 더 잘 이해하도록 도와준다.

□ 변수 스케일 조정

화재 리스크 분석 모델에서 다양한 종류의 변수가 사용된다. 예를 들어, 건물 면적, 연령, 온도, 인구 밀도 등 각 변수의 단위와 크기는 다를 수 있다. 이러한 다양한 변수들을 일관된 스케일로 조정해야 한다. 주로 사용되는 스케일 조정 방법으로는 표준화(Standardization)와 정규화(Normalization)방법이 있다. 이러한 스케일 조정은 변수 간에 비교 가능한 상태로 만들어주며, 모델 학습 과정에서 수렴 속도와 분석 성능을 개선한다.

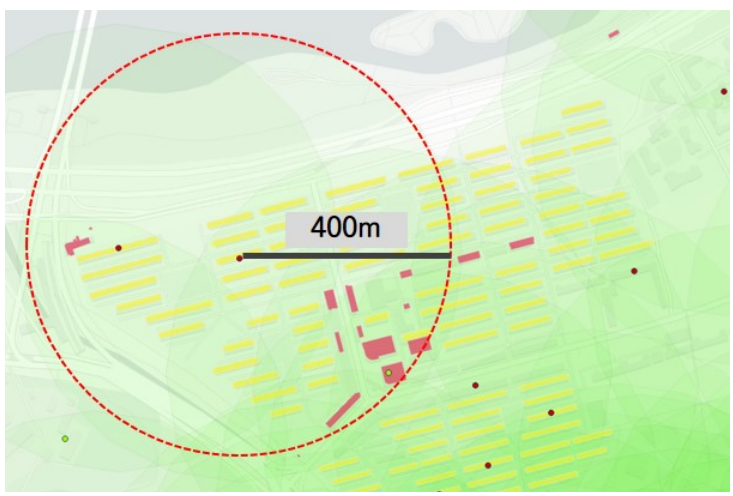
□ 분석단위 설정

본 연구에서 사용되는 분석단위(Unit of Analysis)는 개별 건축물을 중심으로 설정된다. 개별 건물을 분석단위로 설정함으로써 건물 간의 다양한 특성과 주변 환경 요소를 고려할 수 있으며, 더욱 정확한 분석 모델을 개발할 수 있다.

[그림 3-14] 건축물 화재 리스크 분석 모델 변수별 분석 단위
출처: 연구진 작성

이러한 분석단위 설정은 건물 안전성을 평가하고 개선하는 데 있어 중요한 차원을 제공한다. 특히 건축물 주변환경 반영을 위해 지리 정보 시스템(GIS)을 활용하여 각 개별 건축물 주변의 버퍼를 활용하여 분석하였고, 변수의 특성에 따라 다양한 거리 단위로 주변 환경을 반영했다.

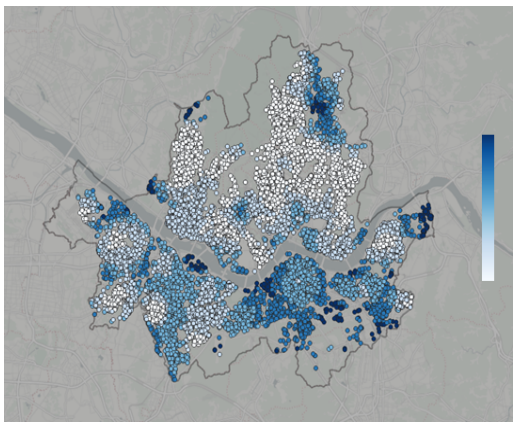
400m버퍼를 활용하여 분석하는 과정을 시각화할 경우 아래와 같은 양상이 나타난다. 분석지점을 중심으로 반지름 400m의 원을 설정하고 해당 원 내부에 포함되는 특성을 더하거나, 평균을 내는 등 산술적으로 분석할 수 있다. 같은 방식으로 도로 연결성 및 도로 폭 변수의 경우, 개별 건축물을 중심으로 1,000m의 버퍼를 설정하여 활용하였다. 또한 인근 건물의 특성을 고려하기 위해 개별 건축물을 중심으로 400m의 근린주구 단위로 버퍼를 설정하여, 건물 주변의 주거 및 상업시설 수를 파악하고, 주변 건물 용도가 화재피해에 미치는 영향을 분석했다.



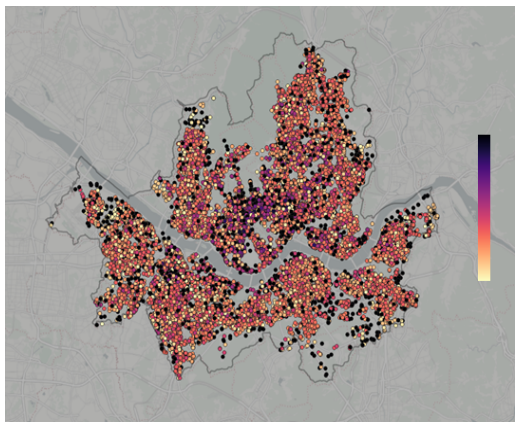
[그림 3-15] 버퍼를 활용한 공간분석(400m 기준)
출처: 연구진 작성

도로 연결성은 주변 도로 네트워크와의 접근성을 고려하는 데 이용될 수 있으며, 이동 및 대피의 용이성을 평가하는 데 활용할 수 있다. 건물의 노후도 및 지상 층수 등의 건물 자체 정보는 개별 건물의 고유한 특성으로 고려되었다. 따라서 별도의 버퍼 설정 없이 각 건물의 자체 정보를 직접 활용하여 분석에 반영했다. 이러한 맞춤형 분석 단위 설정은 건물 화재 리스크 분석 모델을 개발할 때, 다양한 건물 특성과 주변 환경을 종합적으로 고려하는 데 도움을 주며, 정확한 모델 개발과 화재 예방 및 대응에 필요한 정보를 제공한다.

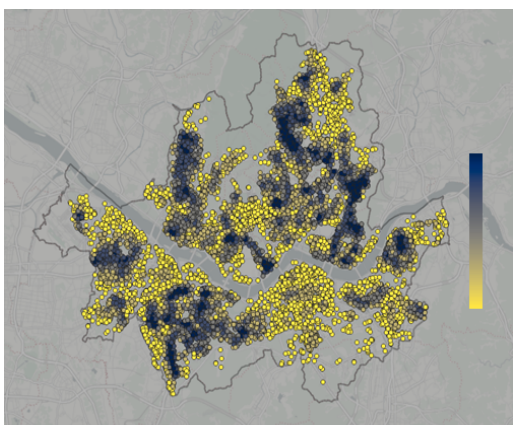
모든 변수는 개별적으로 구축한 후 건축물 단위 분석을 위해 건축물 형상정보와 연계하였다. 이 같은 데이터를 건축물 화재 리스크 분석모델 학습에 종합적으로 이용하였다.날씨특성을 제외한 변수별로 지도에 시각화한 결과는 다음과 같다.



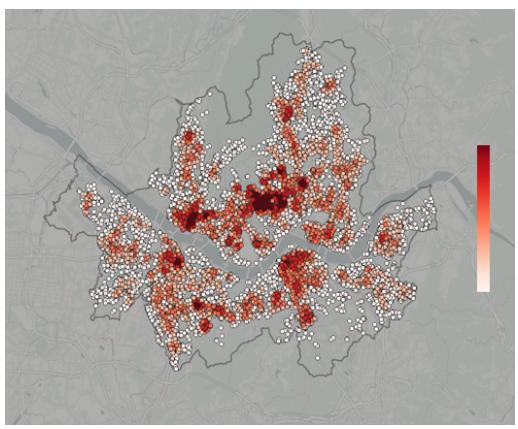
도로 폭



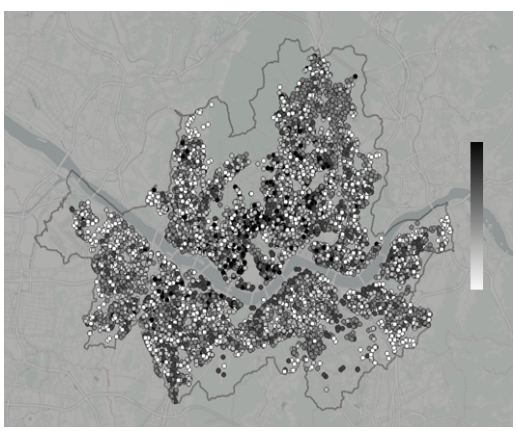
도로 연결성



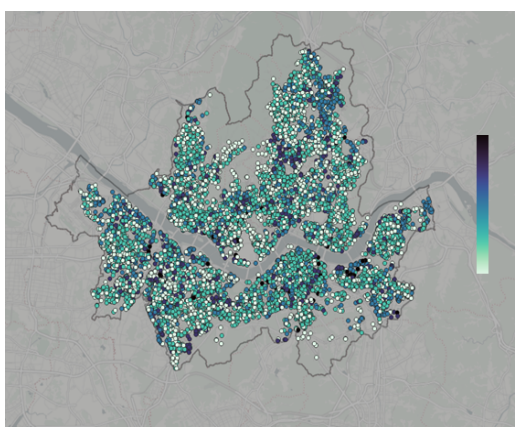
주거용도 수



상업용도 수



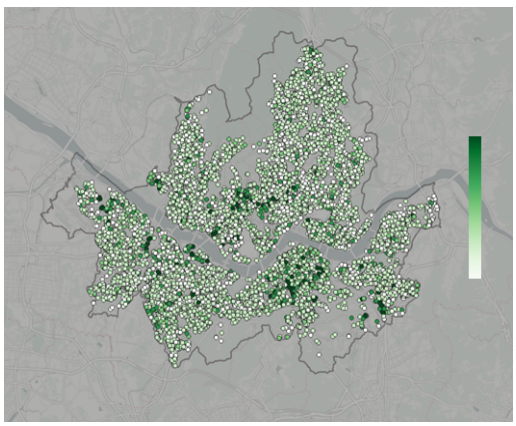
노후도



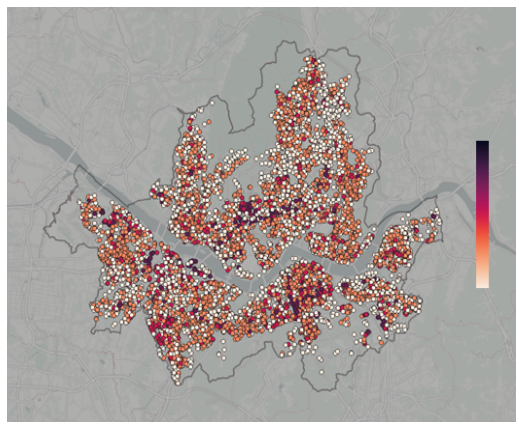
지상층수

[그림 3-16] 건축물 화재 리스크 분석 입력 변수 구축 결과

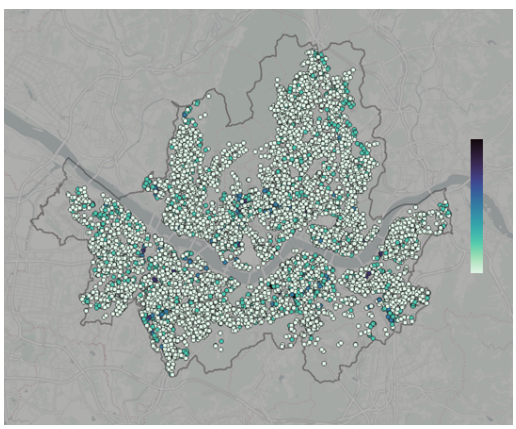
출처: 연구진 작성



지하 층수

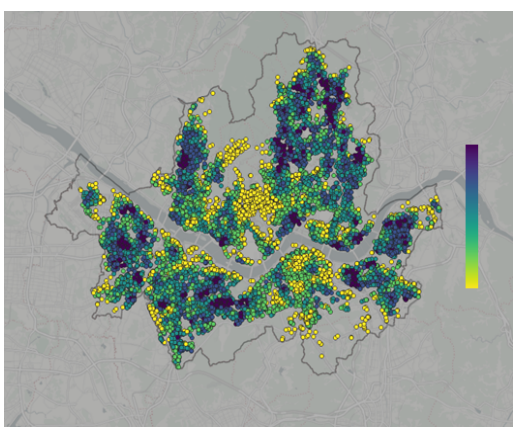


용적률



승강기 수

비상용 승강기 수



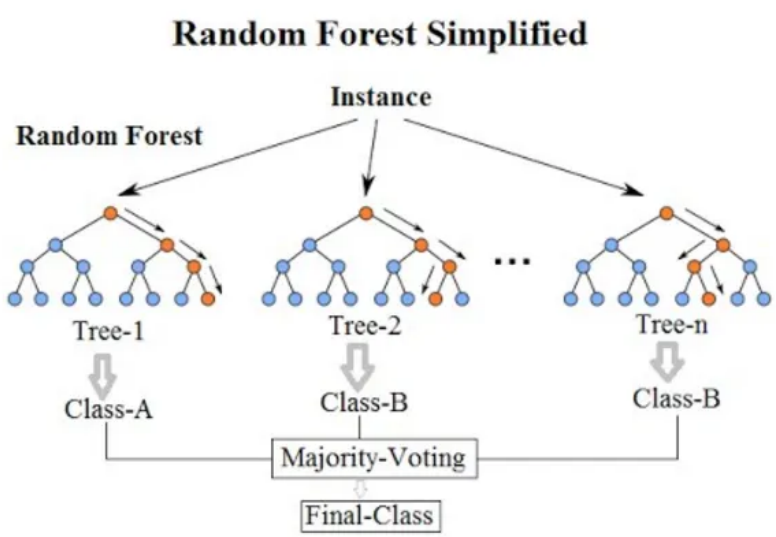
추정인구 수

3) 분석 모델 선정 및 학습

□ 모델 선정

높은 정확도의 모델을 구축하기 위해 다양한 머신러닝 알고리즘을 검토해보았다. 건축물 화재 리스크 분석 모델은 조영진 외(2022)의 연구에서 나타난 것처럼 높은 정확도를 확보하기 어렵다. 이에 다양한 알고리즘을 적용 후 정확도가 높은 알고리즘을 채택하는 방식을 선택하였다. 랜덤포레스트, 로지스틱 회귀 분석, LightGBM, XGBoost 등 알고리즘을 선정하여 각각 학습을 진행하였고, 각각 78%, 65%, 70%, 71%의 정확도를 나타냄을 확인하였다. 결과적으로 본 연구는 가장 뛰어난 성능을 나타낸 랜덤포레스트를 분석모델로 선정하여 연구를 진행하였다.

랜덤 포레스트는 앙상블 학습 기법 중 하나로, 의사결정트리(Decision Tree)의 모음이다. 여러 개의 의사결정트리를 생성하고 각 트리의 분석 결과를 조합하여 최종 분석을 수행하는 방식으로 동작한다(Ali *et al.* 2012, p.272). 이를 통해 과적합(Overfitting) 문제를 감소시키고 분석 성능을 향상시킬 수 있다(Hastie *et al.* 2009, p.596). 랜덤 포레스트는 변수 중요도를 계산하여 어떤 변수가 분석에 중요한 역할을 하는지를 파악할 수 있는데, 이는 모델의 해석과 설명력을 높이는 데에 도움을 준다.



[그림 3-17] 랜덤 포레스트 모델

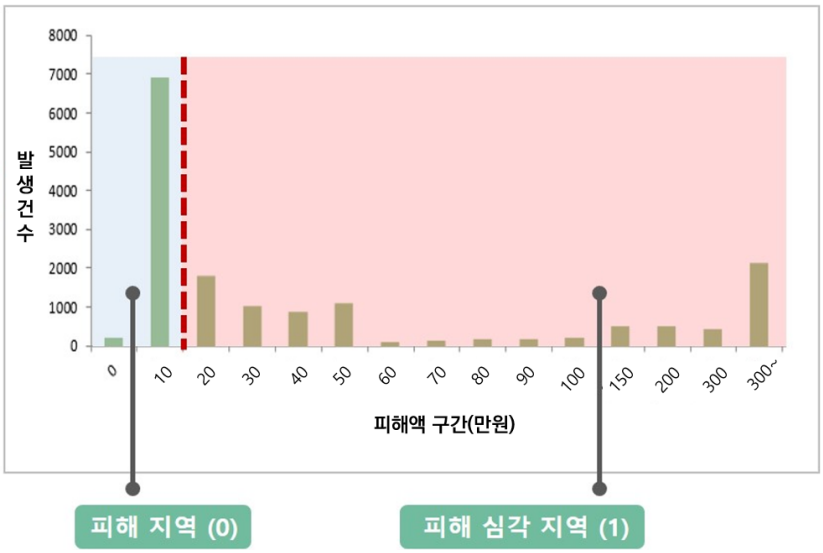
출처: Hoffman, Ken. 2020. https://ken-hoffman.medium.com/random-forest-hyperparameters-explained-8081a93ce23d. 2023.10.10. 접속

□ 입력데이터 전처리

첫째로 이상치 제거 과정을 수행하였다. 데이터에 이상치(Outlier)가 포함되어 있다면, 모델의 분석 성능에 부정적인 영향을 미칠 수 있다. 이상치는 일반적인 데이터 패턴과 동떨어진 값을 가지는 데이터 포인트를 의미한다. 변수 프로세싱에서는 이러한 이상치를 감지하고 처리하는 방법을 적용한다. 대표적인 이상치 처리 방법으로는 이상치 제거, 이상치 대체 등이 있다. 이상치 처리를 통해 모델이 더 안정적으로 학습할 수 있다.

둘째로 결측치 제거 과정을 수행하였다. 실제 데이터 수집 과정에서는 결측치(Missing Value)가 발생할 수 있다. 결측치는 모델 학습에 방해가 되므로 적절한 대체방법을 통해 처리해야 한다. 대표적인 결측치 대체 방법은 평균값, 중앙값, 최빈값 등을 활용하는 것이다. 이러한 변수 프로세싱 단계에서는 데이터의 품질을 향상시키고 모델의 분석력을 향상시키기 위해 다양한 통계 및 머신러닝 기법을 활용한다. 모델 학습에 적절한 형태로 데이터를 가공함으로써, 최종적으로 높은 분석 성능을 가진 화재 리스크 분석 모델을 개발할 수 있다.

화재피해액은 건축물 화재발생 시 피해금액을 기입해 놓은 것으로 본 연구에서는 피해금액을 임계값으로 구분하여 피해지역과 피해 심각지역으로 구분하여 활용하였다. 피해금액은 최소 0원부터 최대 약 716억원까지 기입되어있다. 본 연구는 해당 금액을 10만원 단위로 구분하여 빈도분석을 수행한 결과 피해금액 20만원 미만과 20만원 이상에서 변곡점이 나타나는 것을 확인할 수 있었다. 이에 피해금액 20만원 미만인 화재발생 데이터는 피해가 적은 건축물로 분류하였으며, 피해금액 20만원 이상인 화재발생 데이터는 피해가 큰 건축물로 분류하여 종속변수로 사용하였다.



[그림 3-18] 건축물 화재발생 피해금액 및 발생빈도
출처: 연구진 작성

4) 모델 검증 및 영향 요인 도출

□ 분석모델 정확도 평가지표 선정

건축물 화재 리스크 분석모델의 정확도 평가를 위해 혼돈행렬(Confusion matrix)을 도출하고, 이에 기반하여 정확도 및 F1점수(F1-Score)를 산출하였다. 또한 ROC 곡선(Receiver Operating Characteristic Curve)을 적용하여 모델의 정확도 평가를 수행하였다. 정확도 및 F1점수와 같은 평가지표들을 통해 모델의 성능을 객관적으로 평가하고, 필요한 경우 모델을 개선하는 데 도움을 줄 수 있다(Novaković *et al.* 2017, pp.41-44; 노주형 외 2022, pp.72-73). ROC 곡선의 경우 분석모형의 정확도 측정을 위해 일반적으로 사용되는 평가방법 중 하나로, 본 연구에서도 채택하여 사용하였다(Japkowicz 2006, pp.7-9).

혼돈행렬(Confusion Matrix)은 모델의 분류 결과를 나타내는 표이다. 실제 클래스와 분석한 클래스에 따라 True Positives, True Negatives, False Positives, False Negatives를 나타낸다. 이를 통해 모델이 어떤 오류를 어떻게 발생시켰는지 시각적으로 파악할 수 있다. Confusion Matrix는 정확도, 정밀도, 재현율 등 다양한 평가지표의 계산에 사용된다(노주형 외 2022, pp.72-73).

정확도(Accuracy)는 모델의 전체적인 분석 정확도를 나타내는 지표로, 정확히 분석한 샘플의 비율로 계산되며, 공식적으로는 다음과 같이 표현된다.

정확도(Accuracy)= TP+TN/TP+TN+FP+FN

- TP (True Positives): 실제 화재가 발생하고 모델도 화재로 분석한 경우의 샘플 수
- TN (True Negatives): 실제 화재가 발생하지 않고 모델도 화재가 아닌 것으로 분석한 경우의 샘플 수
- FP (False Positives): 실제 화재가 발생하지 않지만 모델이 화재로 분석한 경우의 샘플 수
- FN (False Negatives): 실제 화재가 발생하였으나 모델이 화재가 아닌 것으로 분석한 경우의 샘플 수

출처 : 노주형 외. 2022. pp.72-73

정확도는 모델의 전반적인 분석 성능을 나타내는 지표로, 정확도가 높을수록 모델의 성능이 좋다고 평가된다. 그러나 데이터 불균형이 심한 경우에는 정확도가 왜곡될 수 있으므로 다른 평가지표와 함께 고려하는 것이 중요하다.

F1점수(F1-Score)는 정밀도(Precision)와 재현율(Recall)의 조화평균으로 계산되는 지표로, 불균형한 데이터셋에서 모델의 성능을 평가하는 데 유용하다. 특히, F1점수는 정밀도와 재현율을 모두 고려하기 때문에 모델의 분석 성능을 평가하는 데 유용하다.정밀도(Precision)와 재현율(Recall)을 이용한 F1점수는 다음과 같이 정의된다.

F1점수(F1-Score)

F1점수(F1-Score) = 2 · Precision · Recall / Precision+Recall
TP 정밀도: Precision= TP/TP+FP
재현율: Recall= TP/TP+FN

출처 : 노주형 외. 2022. pp.72-73

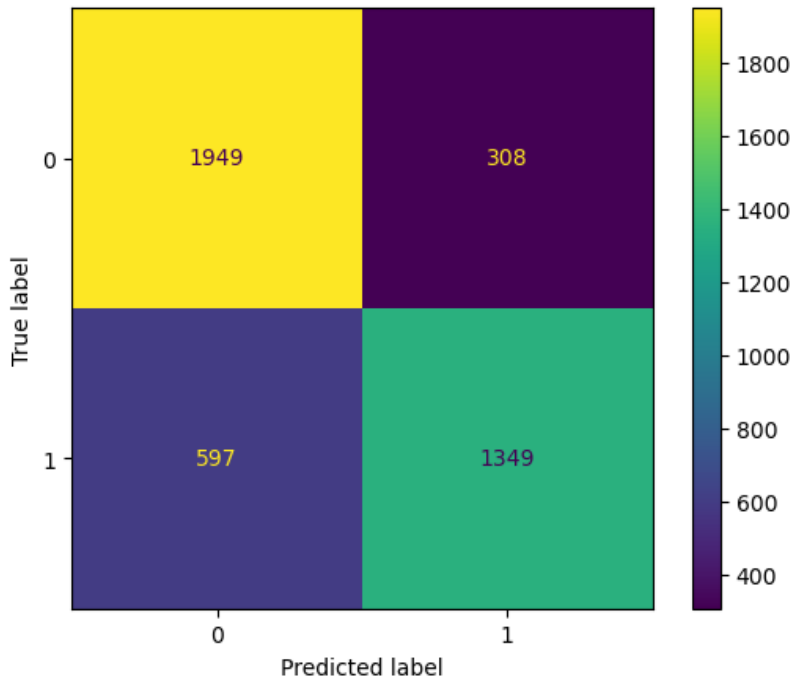
추가적으로 사용한 ROC 곡선은 모델의 분류 성능을 시각적으로 평가하는 데 사용되는 중요한 도구이다. ROC 곡선은 분류 모델의 민감도(Sensitivity)와 1-특이도(Specificity) 사이의 관계를 나타내며, 이를 통해 모델의 성능을 다양한 임계값(threshold)에서 평가할 수 있다.

ROC 곡선은 왼쪽 상단 모서리로 향할수록 모델의 성능이 더 우수하며, 대각선에 가까울수록 성능이 좋지 않다고 해석된다. ROC 곡선 아래 영역인 AUC (Area Under the Curve) 값은 ROC 곡선 아래의 면적을 나타내며, 이 값이 1에 가까울수록 모델의 성능이 우수하다(Japkowicz 2006, pp.7-9). ROC 곡선은 다양한 임계값에서 모델의 성능을 평가할 수 있으며, 특히 분류 모델의 성능을 조정할 때 임계값을 설정하는 데 도움을 줄 수 있어 모델 간의 성능 비교나 최적 모델의 선택에도 ROC 곡선과 AUC 값이 유용하게 활용될 수 있다(Bradley 1997, p.1150).

□ 분석모델 정확도 평가

화재 리스크 분석 모델의 평가 결과를 종합적으로 살펴보면, 다양한 평가 지표를 통해 모델의 성능이 우수함을 확인할 수 있다. 이러한 평가 결과는 모델이 화재 리스크 분석 분야에서 실용적으로 활용될 수 있는 높은 분석 능력을 가지고 있음을 시사한다.

아래 그림은 화재 리스크 분석 모델의 혼돈행렬을 시각화하여 나타내고 있으며, 이를 통해 정확도, 정밀도, 재현율 등 지표값을 계산할 수 있다.



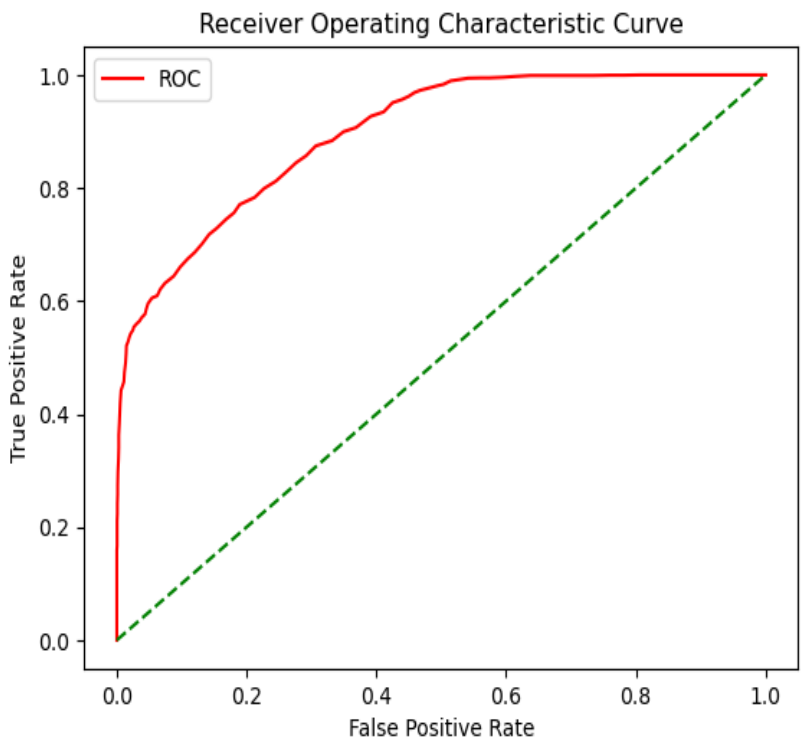
[그림 3-19] 화재 리스크 분석 모델 혼돈행렬
출처: 연구진 작성

분석 모델의 정확도는 약 0.78로, 전체 데이터 중 약 78%를 정확하게 분류하였다. 이는 모델이 대부분의 경우에 정확한 분석을 수행한다는 것을 의미한다. 또한, 정밀도는 0.81로 나타났으며 모델이 화재피해를 분석할 때 그 분석이 실제로 화재 발생과 관련이 있는 경우가 81%임을 의미한다. 이는 모델의 화재 리스크 분석이 상당히 신뢰할 만하다는 것을 보여준다.

정밀도와 재현율의 조화평균을 나타내는 F1점수는 0.75로 나타났다. 이는 모델의 균형 잡힌 분석 능력을 평가하는 데 사용된다.

분석모델의 F1점수가 0.75로 나타났으므로, 모델은 정확도와 정밀도를 균형 있게 유지하면서 화재 리스크 분석을 수행하고 있다고 평가할 수 있다.

ROC 곡선 아래 면적인 AUC 값은 0.89로 나타났다. AUC 값이 1에 가까울수록 모델의 분류 능력이 우수함을 나타내며, 이는 모델이 긍정 및 부정 클래스를 구분하는 능력이 뛰어남을 의미한다.

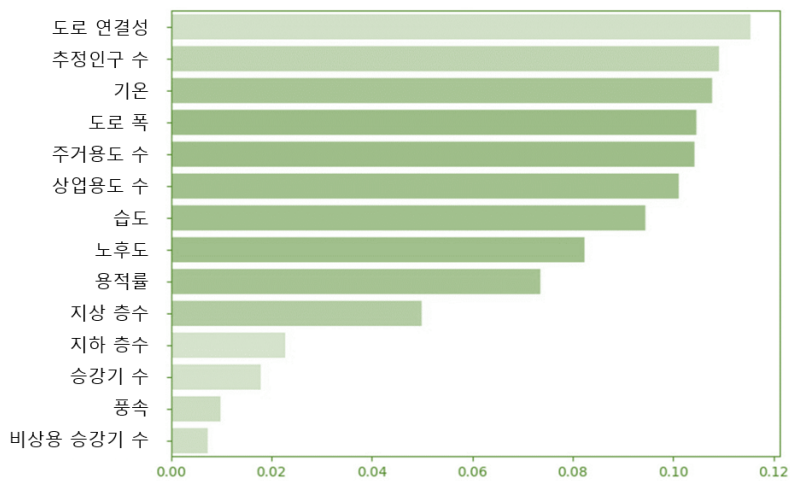


[그림 3-20] 화재 리스크 분석 모델 ROC 곡선
출처: 연구진 작성

평가지표들을 종합적으로 고려할 때, 개발한 분석모델은 화재 리스크 분석에 효과적이며 안정적인 성능을 보인다. 따라서 본 모델은 화재 예방 및 대응을 위한 중요한 도구로 활용될 수 있을 것으로 기대된다. 다만 정확도와 정밀도 간의 트레이드오프를 고려하여 모델을 조정하는 것도 중요할 것이다. 또한 향후 데이터의 추가 수집과 모델의 계속적인 업데이트를 통해 모델의 성능을 높이는 것이 화재 리스크 분석 분야에서 중요한 과제가 될 것이다.

□ 건축물 화재 리스크 분석 영향요인 도출

랜덤 포레스트는 트리 결정 방식을 사용하기 때문에 'feature_importances' 속성을 이용해 알고리즘을 통해 피처의 중요도를 알 수 있다(Menze *et al.* 2009). 또한, 이에 따라 모형에 영향을 미치는 변수중요도를 도출할 수 있다.



[그림 3-21] 화재 리스크 분석 모델 변수중요도 시각화
출처: 연구진 작성

건축물 화재 리스크 분석 모델에서 중요도가 높게 나타난 상위 5개의 변수는 도로 연결성, 인구 수, 온도, 도로 폭, 주거용도 건물 수이다. 화재 리스크 분석 모델 중요도 도출 결과에 따르면, 도로 연결성이 화재 리스크에 가장 큰 영향을 미치는 것으로 나타난다.

4. 건축물 홍수 리스크 분석 모델 개발

1) 입력 데이터 선정

□ 모형 입력데이터 선정

딥러닝 기반 홍수 리스크 분석을 위한 데이터셋은 종속변수와 독립변수로 구분할 수 있다. 종속변수는 홍수 피해금액으로 설정하고, 독립변수는 홍수 리스크에 영향을 미칠 것으로 판단되는 변수들로 선정하였다. 독립변수 중에서 앞서 기계학습 모형 기반 입력데이터와 동일한 데이터는 그대로 사용하였고, 추가로 필요한 변수들을 보완하였다.

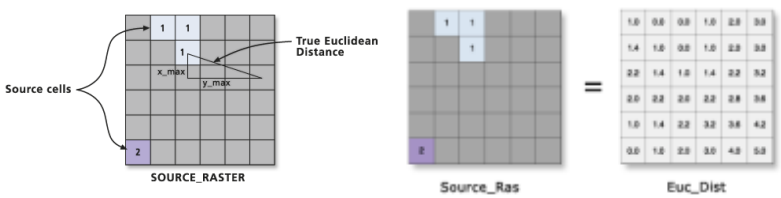
딥러닝 기반 모델의 독립변수들은 피해발생가능성의 증가 및 감소 혹은 홍수 피해액 증가 및 감소에 영향을 줄 것으로 판단되는 요소들이다. 크게 홍수 피해 특성, 건축물 특성, 지형 특성으로 구분할 수 있다. 홍수 피해 특성 변수는 피해 종류, 피해물의 수, 피해유형, 피해일로 구성된다. 건축물 특성 변수는 건축물 구조, 건축물 주용도, 건축물 지붕구조, 건축물 면적으로 구성된다. 지형특성은 고도, 향, 경사, 지형습윤지수, 소유역 정보, 하천으로부터의 거리, 홍수 민감 지도로 구성된다.

[표 3-11] 딥러닝 기반 건축물 홍수 리스크 분석을 위한 종속변수와 독립변수 리스트

구분	데이터 명	설명	피해발생	피해액
종속변수	피해금액	건축물 홍수 피해금액		
독립변수	연면적	하나의 건축물 각 층의 바닥 면적의 합계		○
	건축면적	건축물 외벽(기둥)의 중심선으로 둘러싸인 부분의 수평투영면적		○
	대지면적	대지의 수평투영 면적	○	○
	건폐율	건축면적의 대지면적에 대한 비율		○
	용적률	건축물 총면적의 대지지면적에 대한 백분율		○
	용적률 산정용 연면적	-		○
	건물높이	-		○
	지상층수	-		○
	지하층수	-	○	○
	건물수	-	○	
	구조	건축물의 구성구조		○
	주용도	건축물의 주 사용용도		○
	향	건축물이 위치한 지점의 대지의 방향	○	○
	경사도	건축물이 위치한 지점의 경사도	○	
	지형습윤지수	통수가능량(경사)을 통한 지점의 습윤도	○	
	하천으로부터의 거리	건축물이 위치한 지점과 하천과의 거리	○	
	고도	건축물이 위치한 지점의 고도값	○	
	고도차이	평균고도(소유역) - 건축물 지점별 고도값	○	
	소유역	건축물이 위치하고 있는 유역	○	
	평균고도(소유역)	소유역별 평균 고도값	○	
	홍수 민감 지도	기계학습 기반 모델(MaxEnt)을 통해 도출된 건축물 별 홍수피해 민감도	○	

출처: 연구진 작성

종속변수인 홍수발생 피해금액과 독립변수 중 피해 특성과 건축물 특성에 대해서는 데이터 구축 시 조사된 정보를 그대로 활용하였다. 소유역 정보는 국토교통부에서 제공하는 표준권역 정보 중 소유역 정보를 이용하여 건축물 위치의 소유역 정보를 추출하였다[27]. 소유역 고도값은 유역별 고도 평균 값으로 계산하였다. 고도차이는 유역평균 고도값에서 건출물 지점별 고도값을 뺀 값을 활용하였다. 하천으로부터의 거리는 ESRI ArcGIS10.5를 이용하여 토지피복지도의 하천을 기준으로 Euclidean distance를 분석하고, 건축물 위치의 값을 추출하였다.



[그림 3-22] Euclidean distance의 거리 계산 방법

출처: esri. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/understanding-euclidean-distance-analysis.htm. 2023.06.20. 접속

□ 홍수민감지역(홍수 민감 지도) 입력데이터 선정 및 가공

건축물 홍수 리스크 분석을 위한 독립변수 중 홍수 민감 지도 구축은 추가적인 분석을 통해 데이터를 구축하였다. 홍수 민감 지도는 기계학습 모형기반의 분석을 수행하였다. 먼저 기계학습 모형으로는 MaxEnt (Maximum Entropy Model)를 사용하였다. 분석에 사용된 종속변수는 건축물의 피해유무 정보이며, 독립변수는 15개로 구성된다. 관련 독립변수는 선행연구 검토를 통해 최종적으로 선정하였다.

[표 3-12] 기계학습 기반 홍수 민감지역 분석을 위한 종속변수와 독립변수

변수 구분			설명	선행연구
종속변수	홍수피해	피해건축물	홍수피해 건축물의 위치 정보	-
독립변수	지형 및 토양	고도	건축물의 해발고도(m)	- Zuo *et al.* 2023
		경사도	건축물이 위치한 지점의 경사(도)	- Siahkamari *et al.* 2018
		유효토심	건축물이 위치한 지점의 유효토심	- 김종계 외 2008
		토성	건축물이 위치한 지점의 토성	- Karmakar *et al.* 2010

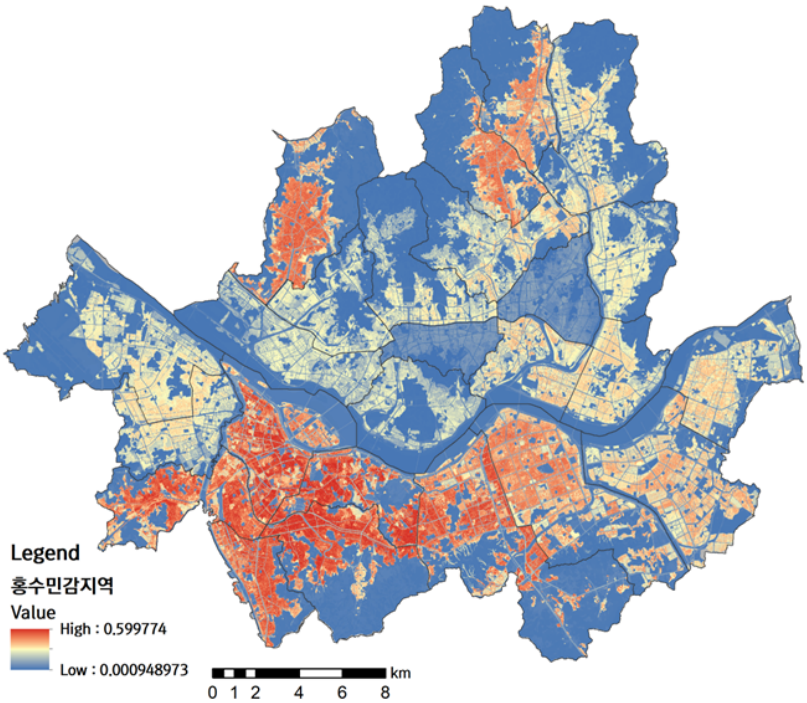
27) 공공데이터포털. https://www.data.go.kr/data/15056790/openapi.do?recommendDataYn=Y. 2023.06.20. 접속

변수 구분			설명	선행연구
		토양배수	건축물이 위치한 지점의 토양배수등급	- Kim *et al*. 2016
		지형습윤지수	건축물이 위치한 지점의 지형습윤지수	- Shafizadeh-Moghadam *et al*. 2018
		하천으로부터의 거리	하천으로부터 건축물 위치까지의 거리	- Pourghasemi *et al*. 2023
	인공 환경	그린인프라 면적	건축물이 위치한 행정구역(구단위) 내 비오톱 등급 1, 2등급에 해당하는 지역의 면적 비율	- 김효민 외 2013
		불투수포장율	건축물이 위치한 지점의 불투수포장율	- Iqbal *et al*. 2022
		오픈스페이스	건축물이 위치한 행정구역(구단위) 내 비오톱 지도 기준 오픈스페이스의 면적 비율	- Brody & Highfield 2013
		토지이용	건축물이 위치한 지점을 주거지, 공업지, 상업지, 문화·체육·휴양지, 교통지, 공공시설, 농업지, 산림, 초지, 습지, 나지, 수역으로 분류	- Brody *et al*. 2014
		구면적 대비 빗물 저류조 용량	건축물이 위치한 행정구역(구단위)의 면적 대비 빗물 저류조 용량	- 김효민 외 2013 - Freni & Liuzzo 2019
		구면적 대비 빗물 펌프장 용량	건축물이 위치한 행정구역(구단위)의 면적 대비 빗물 펌프장 용량	- 김효민 외 2013 - Choo *et al*. 2021
		시가화 면적 대비 하수관거 연장	건축물이 위치한 행정구역(구단위)의 면적 대비 하수관거연장	- 김효민 외 2013 - Jang *et al*. 2019
		구면적 대비 유수지 면적	건축물이 위치한 행정구역(구단위)의 면적 대비 유수지 면적	- 김효민 외 2013 - Li *et al*. 2019

출처: 연구진 작성

구축된 15종의 환경변수를 독립변수로 하고, 종속변수로 홍수 피해 건축물 위치 정보를 설정하여 MaxEnt 모델을 구동하였다. 5-fold CV을 통해 모델 구동 과정에서 발생하는 불확실성을 보완하고자 하였으며, ROC curve 분석을 통해서 모델의 통계적 신뢰성을 도출하였다. 그 외 모델 구동을 위한 hyper parameter는 기본값으로 설정하였다. 홍수 민감 지도 제작 시에는 모델 구동 후 도출되는 Threshold 값을 적용하여 이항지도로 변경하였다.

홍수 민감 지역 분석결과, 피해 건축물의 위치가 집중되어 있는 한강 남쪽의 홍수 민감성이 높게 도출되었다. 피해 건축물의 위치 정보가 편향되는 문제점을 해소하기 위해서 분석단위 격자별 위치정보 샘플링을 수행했음에도 불구하고, 편향성이 나타나는 것으로 파악된다. 서울시의 남부에서는 구로구, 영등포구, 동작구, 금천구, 관악구, 서초구, 강남구에 이르는 지역이 홍수 발생 확률이 가장 높게 나타나는 민감지역으로 파악되었다. 서울시의 북부에서는 은평구, 강북구, 도봉구의 홍수 발생 확률이 높은 것으로 파악되었다.



[그림 3-23] 건축물 홍수 민감 지역
출처: 연구진 작성

독립변수로 활용된 개별 환경변수의 특성을 살펴보면 다음과 같다. 고도는 0~100m에서 홍수 발생 확률이 높았으며, 100m 이후에는 홍수 발생 확률이 10% 미만으로 감소했다. 경사는 0~5%에서 홍수 발생 확률이 높았으며, 경사가 급해질수록 홍수 발생 확률이 점차 감소하였다. 토성은 공극의 크기가 가장 작은 미사질식양토에서 홍수 발생 확률이 가장 높게 나타났다. 토양 배수의 경우 등급이 보통을 기준으로 양호해질수록 홍수 발생 확률이 감소하였다. 지형습윤지수는 3이상부터 홍수 발생 확률이 급격하게 증가하여 13에 이르러 확률이 최고점에 이르렀다.

또한, 하천으로부터의 거리가 1km까지는 홍수 발생 확률이 증가하는 양상이 보이며 그 이후로는 점차 감소하는 양상을 보였다. 그린인프라 면적 비율은 15% 미만인 경우 홍수 발생 확률이 높았으며, 35% 이상부터 홍수 발생 확률이 점차 감소하였다. 토지이용의 경우, 종속변수로 활용된 홍수 피해 건축물이 위치하고 있는 주거지, 공업지 상업지에서 홍수확률이 높게 나타났다. 구면적 대비 빗물 저류조 용량이 크면 클수록, 펌프용량이 많을수록, 관거 비율이 증가할수록, 유수지 면적이 클수록 발생확률이 감소하는 것으로 나타났다.

도출된 결과에 대해 신뢰도 검증을 수행하였다. ROC (Receiver Operating Characteristic) 곡선 분석을 통해 도출된 모델의 신뢰도(AUC: Area Under the ROC Curve) 값은 0.710로 보통 수준의 신뢰도를 나타냈다. 이는 홍수 피해 발생 건축물 데이터 위치의 편향성으로 인해 높은 신뢰도가 도출되지 못한 것으로 판단된다.

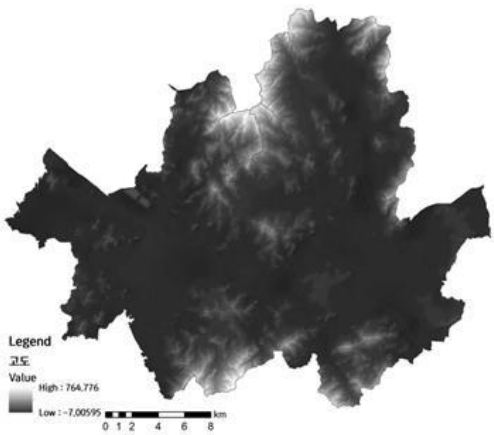
2) 데이터 가공

지형 독립변수 구축을 위해서 일부 데이터의 가공이 필요하였다. 국토교통부 국토지리정보원에서 제공하는 수치표고모델(DEM: Digital elevation model)을 이용하여 건축물의 고도 정보를 추출하였다[28]. 향과 경사는 ESRI ArcGIS10.5를 활용하여 DEM을 이용하여 분석하고, 건축물 위치의 값을 추출하였다. 지형습윤지수는 한국지질자원연구원에서 제공하는 데이터를 바탕으로 건축물 위치의 값을 추출하였다[29].

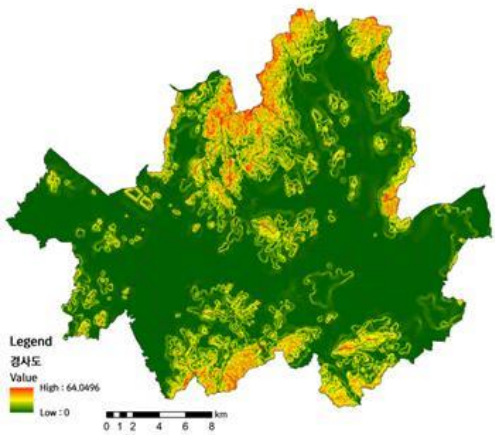
그린인프라 면적 변수는 비오톱 등급 1, 2등급에 해당되는 면적을 그린인프라 면적 비율 정보를 기반으로 구축하였다. 인공적인 환경변수는 인간의 영향이 미치는 영역을 대상으로 구축하였으며, 불투수포장율, 구면적 대비 빗물 저류조 용량 등 데이터를 구축하였다. 오픈스페이스는 서울시 도시생태현황도(비오톱지도)의 기준을 적용하였으며, 토지이용은 주거지, 공업지, 상업지, 문화·체육·휴양지, 교통지, 공공시설, 농업지, 산림, 초지, 습지, 나지, 수역으로 구분하여 분석을 진행하였다. 배수구역이나 배수분구별 처리가 이루어지는 빗물저류조 및 유수지 면적 등은 면적당 시설용량이 홍수 발생 저감에 중요하기 때문에 분석에 활용하였다.

28) 공공데이터포털. https://www.data.go.kr/data/15059920/fileData.do. 2023.06.15. 접속

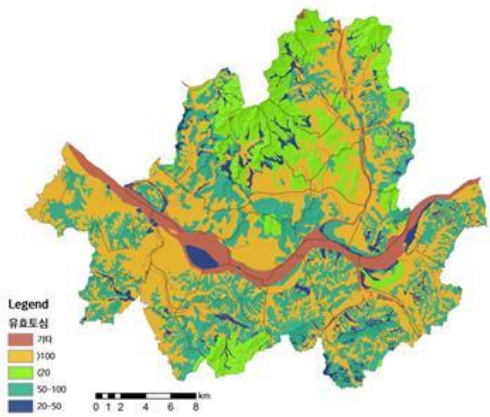
29) 환경 빅데이터 플랫폼. https://www.bigdata-environment.kr/user/data_market/detail.do?id=aaa8c7b0-313f-11ea-adf5-336b13359c97. 2023.06.15. 접속



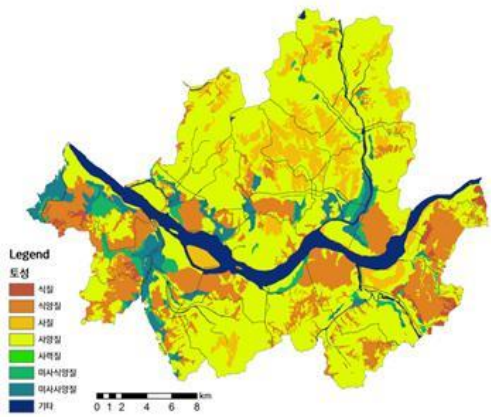
고도



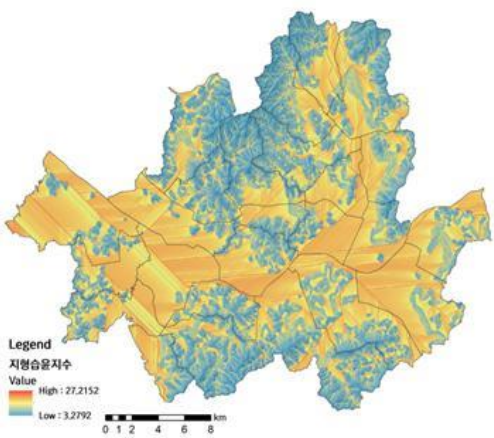
경사도



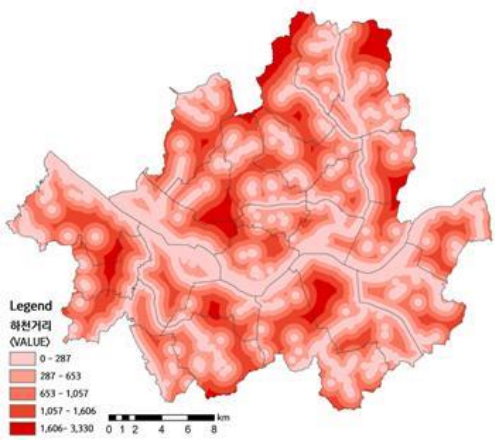
유효토심



토성



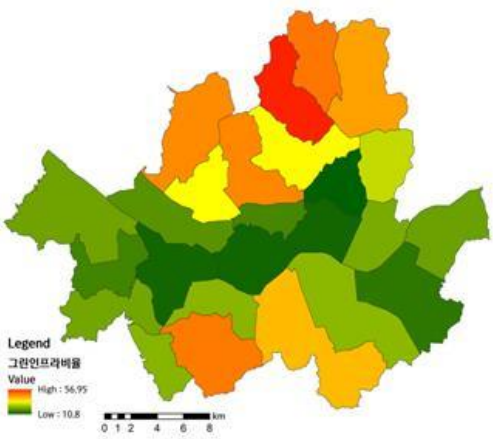
지형습윤지수



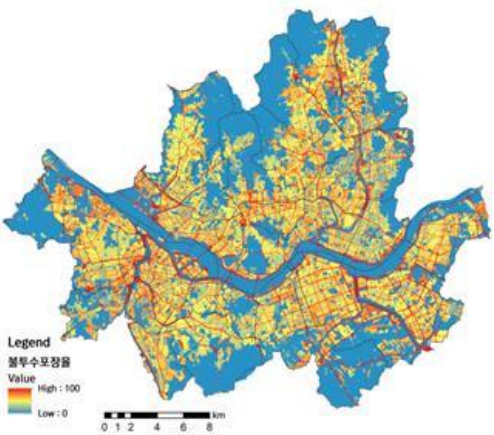
하천으로부터의 거리

[그림 3-24] 홍수 민감 지도 분석 환경 변수 시각화
출처: 연구진 작성

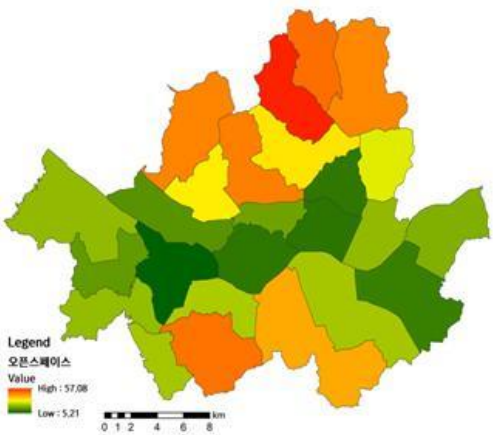
토양배수



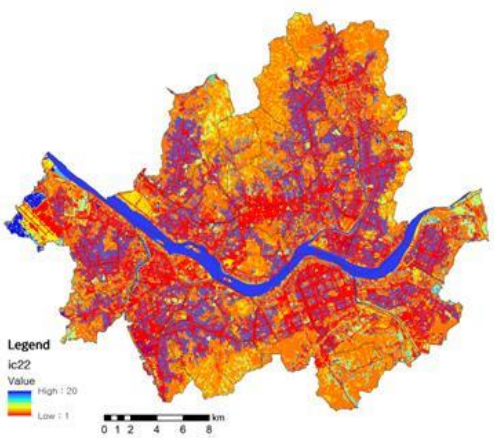
그린인프라 면적



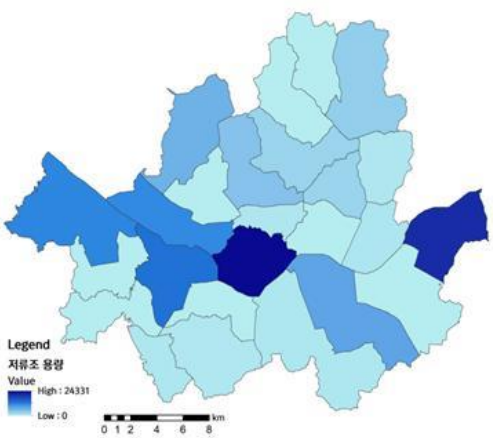
불투수포장율



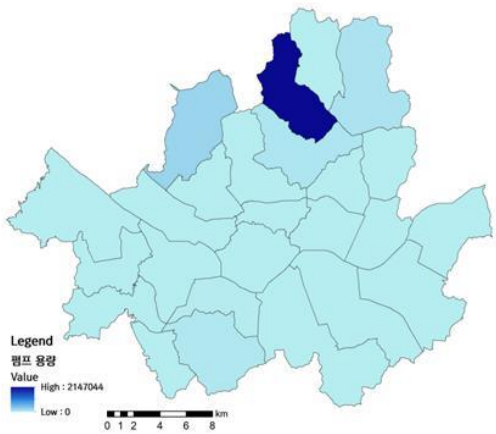
오픈스페이스



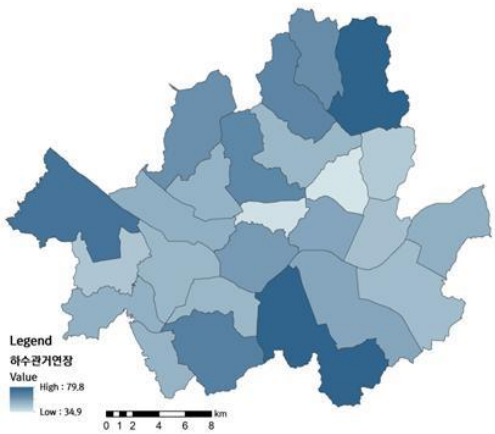
토지이용



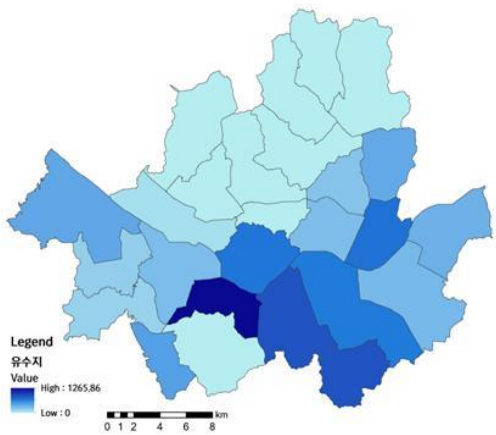
구면적 대비 빗물 저류조 용량



구면적 대비 빗물 펌프장 용량



시가화 면적 대비 하수관거 연장



구면적 대비 유수지 면적

[그림 3-24] 홍수 민감 지도 분석 환경 변수 시각화(이어서)
출처: 연구진 작성

3) 분석 모델 선정 및 학습

□ 모델 선정

건축물 홍수 리스크 분석을 위한 딥러닝 모델로는 TabNetClassifier를 선정하였다. 입력 데이터의 변수 중 가장 중요한 변수들을 스스로 학습하여 선택하므로 복잡한 데이터에서도 유의미한 패턴과 상관관계를 자동으로 학습하고 활용하여 좋은 성능을 보인다. TabNetClassifier는 딥러닝 모델인 TabNet을 사용하여 분류(classification) 문제를 해결하는 데 사용되는 모델로, 특히 표 형식의 데이터(탭형 데이터)에 특화되어 있으며 자동으로 특징의 중요도를 고려하여 데이터를 학습하고 분석할 수 있는 모델이다[30].

□ 입력데이터 전처리

Distance가 5를 초과하거나 민감확률이 -9999인 데이터를 삭제하고 총 27,438개 데이터를 사용하여 학습하였다. 주소, 번지, 반복되는 특징 등 부적합한 특징을 제거하여 총 23개의 특징을 사용하였다. 원-핫 인코딩으로 인해 특징의 수가 늘어나 최종적으로 113개 특징이 학습에 사용되었다.

[표 3-13] 데이터 전처리방법

전처리방법	특징(변수)명	개수
Min-max scaling	위도, 경도, 대지면적, 건축면적, 연면적, 건폐율, 용적률 산정용 연면적 , 용적률, 건물높이, 지상층수, 지하층수, 건물수, 향, 경사도, 지형습윤지수, 하천으로부터의 거리, 건축물 지점별 고도값, 고도차이, 민감확률, 평균고도(소유역)	20
One-hot encoding	소유역 단위, 구조, 주용도	3

출처: 연구진 작성

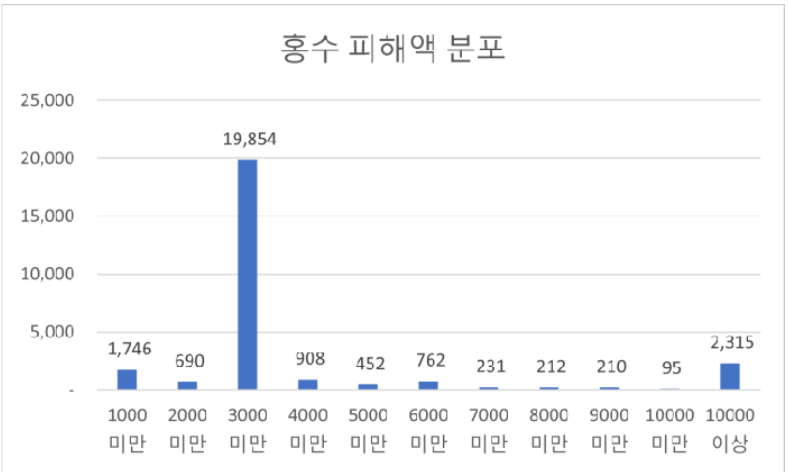
건축물 홍수 리스크 분석 모델을 학습시키기 위해 pytorch_tabnet 라이브러리의 TabNetClassifier를 활용하여 분석 모델을 구현하였다. 학습데이터와 테스트데이터를 9:1의 비율로 나누어 학습을 진행하였다. 피해액의 분포가 200~300만원 미만 사이에 편중되어 있으므로, 학습시 SMOTE[31][32]를 적용하여 불균형 문제를 완화시켰다.

30) DreamQuark. https://github.com/dreamquark-ai/tabnet. 2023.06.20. 접속

31) SMOTE (Synthetic Minority Over-sampling Technique) : 분류 문제에서 소수 클래스(적은 빈도로 발생하는 클래스)의 샘플을 늘려 모델의 성능 향상

32) Maklin, Cory.. 2022. https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c. 2023.06.20. 접속

(단위: 천원)



[그림 3-25] 건축물 홍수 피해액 분포
출처: 연구진 작성

건축물의 홍수 피해액을 기준으로 총 5개의 경우로 나누어 학습하고 모델의 정확도를 평가하였다.

[표 3-14] 피해액 기준 단위별 클래스 학습

유형 1. 피해액 기준 100만원 단위로 나누어 총 11개 클래스로 학습											
구간	≤100	≤200	≤300	≤400	≤500	≤600	≤700	≤800	≤900	≤1000	〉1000
클래스	0	1	2	3	4	5	6	7	8	9	10
유형 2. 피해액 기준 200만원 단위로 나누어 총 6개 클래스로 학습											
구간	≤200		≤400		≤600		≤800		≤1000		〉1000
클래스	0		1		2		3		4		5
유형 3. 피해액 기준 300만원 단위로 나누어 총 4개 클래스로 학습											
구간	≤300			≤600			≤900			〉900	
클래스	0			1			2			3	
유형 4. 피해액 기준 500만원 단위로 나누어 총 3개 클래스로 학습											
구간	≤500				≤1000				〉1000		
클래스	0				1				3		
유형 5. 피해액 기준 300만원 이하, 초과 2개의 클래스로 학습											
구간	≤300							〉300			
클래스	0							1			

출처: 연구진 작성

4) 모델 검증 및 영향 요인 도출

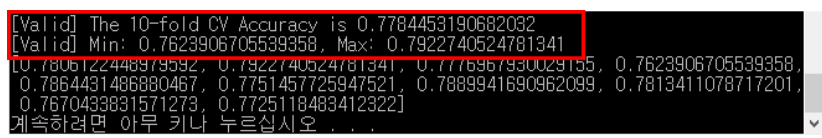
□ 모델 정확도 평가지표 선정

학습된 모델의 정확도는 검증(validation) 데이터에 대한 정확도로 평가하였다. 10-fold CV (cross validation)을 통해 정확도를 측정하였다. K-fold CV는 데이터를 K개의 폴드(fold)로 나누고, K-1개의 폴드를 훈련(train) 데이터로, 나머지 1개의 폴드를 검증(valid) 세트로 사용하여 성능을 평가한다. 이 과정을 K번 반복함으로써 데이터가 적을 때에도 신뢰할 수 있는 성능 지표를 얻을 수 있다[33].

□ 분석모델 정확도 평가

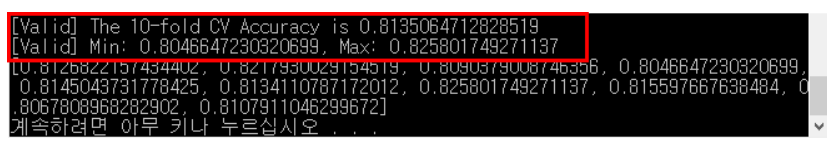
실제 학습 결과 화면을 캡쳐하고, 빨간색 사각형으로 검증 데이터에 대한 평균 정확도와 최대, 최소 정확도를 표시하였다. 각 CV 회차별 정확도도 확인할 수 있다.

- (유형 1) 피해액 기준 100만원 단위로 나누어 총 11개 클래스로 학습을 진행하였으며, 10-fold CV 결과, 평균 정확도는 77.84%(최대 79.23%, 최소 76.24%)로 나타났다.



[그림 3-26] 피해액 기준, 클래스 별 학습 진행 결과: 100만원, 총 11개 클래스
출처: 연구진 작성

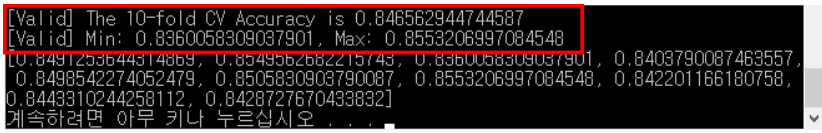
- (유형 2) 피해액 기준 200만원 단위로 나누어 총 6개 클래스로 학습을 진행하였으며, 10-fold CV 결과, 평균 정확도는 81.35%(최대 82.58%, 최소 80.47%)로 나타났다.



[그림 3-27] 피해액 기준, 클래스 별 학습 진행 결과: 200만원, 총 6개 클래스
출처: 연구진 작성

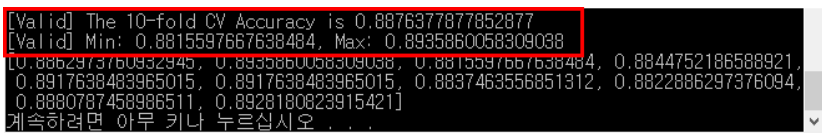
- (유형 3) 피해액 기준 300만원 단위로 나누어 총 4개 클래스로 학습을 진행하였으며, 10-fold CV 결과, 평균 정확도는 84.66%(최대 85.53%, 최소 83.60%)로 나타났다.

33) Scikit-learn. https://scikit-learn.org/stable/modules/cross_validation.html. 2023.06.20. 접속



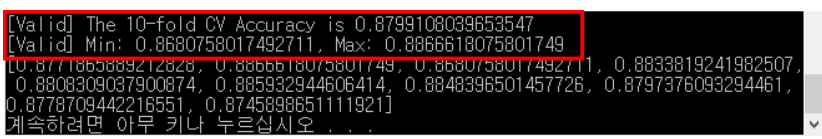
[그림 3-28] 피해액 기준, 클래스 별 학습 진행 결과: 300만원, 총 4개 클래스
출처: 연구진 작성

- (유형 4) 피해액 기준 500만원 단위로 나누어 총 3개 클래스로 학습을 진행하였으며, 10-fold CV 결과, 평균 정확도는 88.76%(최대 89.36%, 최소 88.16%)로 나타났다.



[그림 3-29] 피해액 기준, 클래스 별 학습 진행 결과: 500만원, 총 3개 클래스
출처: 연구진 작성

- (유형 5) 피해액 기준 300만원 이하와 300만원 초과 2개의 클래스로 학습을 진행하였으며, 10-fold CV 결과, 평균 정확도는 87.99%(최대 88.67%, 최소 86.81%)로 나타났다.



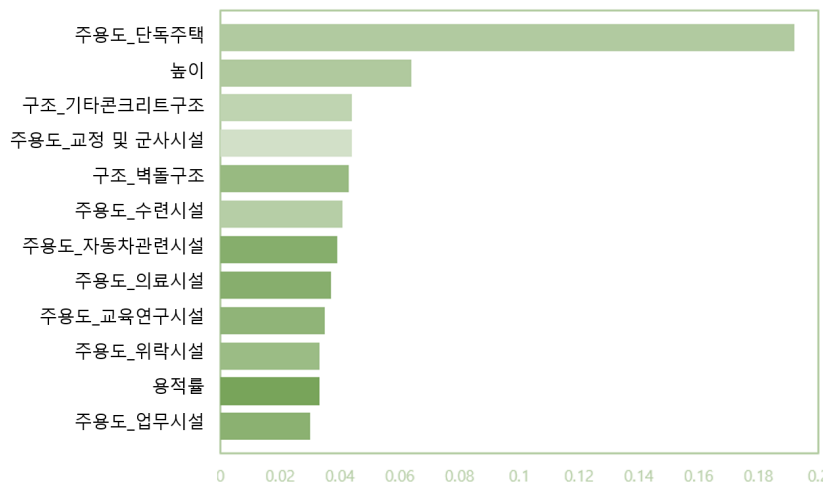
[그림 3-30] 피해액 기준, 클래스 별 학습 진행 결과: 300만원 이하, 초과 2개 클래스
출처: 연구진 작성

본 연구는 건축물 화재 리스크 분석 모델과의 정합성을 위하여 최종적으로 피해액 기준 300만원 이하와 초과의 2개 클래스로 학습한 모델을 선정하여 향후 분석에 활용하였다. 해당 모델의 정확도는 평균 약 88% 수준으로 모델의 정확도 측면에서 높은 정확도를 나타낸다고 판단할 수 있다.

□ 건축물 홍수 리스크 분석 영향요인 도출

건축물 홍수 피해 리스크 분석 결과, 총 10회 반복 딥러닝 분석을 하였을 때 변수별 기여도 및 중요도를 산출하였다. 모델 학습을 위해 독립변수 중 일부는 원-핫 인코딩을 진행하였기 때문에 주용도 변수, 구조 변수, 소유역 관련 변수가 다수 존재한다. 즉, 원-핫 인코딩을 진행함에 따라 위와 같이 독립변수가 여러 개의 변수로 분리되었다.

모델의 홍수 리스크 분석에 사용되는 독립변수 수는 총 71개이며 변수별 영향정도를 기준으로 영향을 많이 미치는 변수들을 추출하였다. 건축물 홍수 리스크 분석에 높은 영향을 주는 상위 12개의 독립변수는 주용도가 단독주택인 경우, 건축물의 높이, 기타콘크리트 구조, 주용도가 교정 및 군사시설인 경우, 벽돌구조, 주용도가 수련시설인 경우, 주용도가 자동차 관련 시설인 경우, 주용도가 의료시설인 경우로 나타났다. 종합하면 건축물의 용도와 구조 그리고 높이와 용적률이 건축물 홍수 리스크를 분석하기 위한 주요 변수인 것으로 판단할 수 있다.



[그림 3-31] 홍수리스크 분석 모델 변수별 중요도
출처: 연구진 작성

가장 중요도가 높게 도출된 변수는 건축물의 용도 중 단독주택으로 중요도가 0.192로 나타났다. 다수의 변수로 구분된 건축물의 용도와 구조 이외의 변수 중 건축물의 높이 변수는 중요도가 0.064, 용적률은 0.033으로 나타났다.

[표 3-15] 홍수리스크 분석 모델 변수별 중요도

변수명	중요도	변수명	중요도
주용도_단독주택	0.192	주용도_자동차관련시설	0.039
높이	0.064	주용도_의료시설	0.037
구조_기타콘크리트구조	0.044	주용도_교육연구시설	0.035
주용도_교정및군사시설	0.044	주용도_위락시설	0.033
구조_벽돌구조	0.043	용적률	0.033
주용도_수련시설	0.041	주용도_업무시설	0.03

출처: 연구진 작성

5. 소결

본 연구는 서울시 건축물 화재 및 홍수에 의한 피해액 데이터를 기반으로 건축물 화재 리스크 분석 모델과 건축물 홍수 리스크 분석 모델을 개발하였다. 화재와 홍수는 건축물에서 발생하는 주요 재해·재난으로, 기존에도 화재나 홍수 발생 예측 연구가 다수 수행되었다. 본 연구는 건축물 화재나 홍수를 예측하는 모델이 아닌 건축물에 화재나 홍수가 발생한 경우 피해액의 규모를 분석하는 연구를 수행하였다. 이를 위해 서울시를 대상으로 수집한 건축물 화재 피해액과 홍수 피해액 자료를 활용하였으며, 피해액을 기준으로 리스크가 낮은 건축물과 리스크가 높은 건축물로 구분하여 모델을 개발하였다.

건축물 화재 및 홍수 리스크 분석 모델의 학습을 위해 학습데이터 구축이 무엇보다 중요하다. 기존 조영진 외(2022)의 연구에서는 화재발생 데이터와 건축물대장 데이터의 연계를 위해 주소정보 텍스트 기반의 매칭을 수행하여 1:1 매칭이 되지 않는 주소정보를 매칭하기 위한 인력 및 시간 등 자원 투자가 과다하였다는 한계점이 있었다. 본 연구에서는 이 같은 한계를 극복하기 위해 주소정보를 이용한 지오코딩을 수행한 후 공간단위의 매칭을 진행하였으며, 매칭이 되지 않는 소수의 데이터는 학습데이터로 사용하지 않았다. 이를 통해 학습용 데이터 구축을 위해 소모되는 시간과 인력을 효율적으로 활용할 수 있었고, 모델의 정확도 개선을 위해 더 많은 자원을 투자할 수 있었다.

건축물 화재 리스크 분석 모델은 피해액 20만원을 기준으로 피해액 20만원 미만인 경우 리스크가 낮은 건축물, 20만원 이상인 건축물은 리스크가 높은 건축물로 구분하였다. 모델 정확도 향상을 위해 랜덤포레스트, 로지스틱 회귀 분석, LightGBM, XGBoost 등 다양한 알고리즘을 적용해보았으며, 그 중 정확도가 가장 높게 나타난 랜덤포레스트 모델을 채택하였다. 또한, 조영진 외(2022)의 건축물 화재 발생 예측 모델에서 사용한 변수에 더하여 건축물 화재 리스크에 영향을 미칠 것으로 예상되는 도로 연결성, 인구 수, 온도, 도로 폭, 주거 건물 수 등 도시환경 및 사회환경 변수를 독립변수로 추가하였다. 본 연구의 건축물 화재 리스크 분석 모델의 정확도는 78% 수준으로 나타났으며, 이는 조영진 외(2022)의 연구에서 개발한 건축물 화재 리스크 분석 모델의 65.8% 수준대비 10% 이상의 정확도 개선이 이루어졌음을 확인할 수 있다.

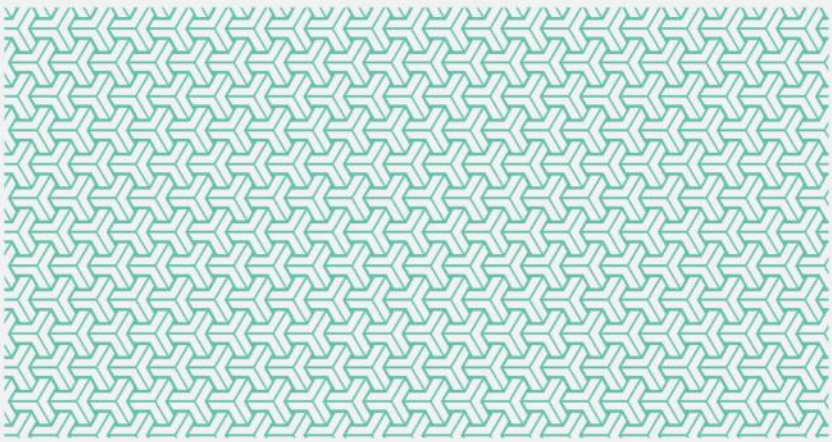
건축물 홍수 리스크 분석 모델은 피해액 300만원을 기준으로 피해액 300만원 이하는 리스크가 낮은 건축물 300만원 초과는 리스크가 높은 건축물로 구분하여 분석을 수행하였다. 건축물 홍수 리스크 분석 모델은 입력데이터로 건축물 속성

정보, 지형변수, 토지이용 및 토지피복 변수 이외에도 MaxEnt 모델을 활용한 홍수 민감도 지도를 입력자료로 활용하였다. 홍수 민감도 분석을 위해 기존 선행연구에서 사용한 지형관련 변수, 토성 관련 변수, 토지이용 및 토지피복 관련 변수, 배수 관련 변수 등 다양한 변수를 입력자료로 활용하였으며, 홍수 피해를 입은 건축물을 종속변수로 활용하였다. 건축물 홍수 리스크 분석을 위해 구축한 독립변수들의 전처리를 수행한 후 딥러닝 모델인 TabNetClassifier 모델에 적용하였다. 최종적으로 정확도는 88%수준으로 높은 정확도를 갖는 모델을 도출하였다.

건축물 화재 리스크 분석 모델과 홍수 리스크 분석 모델의 학습 결과, 각 모델의 분석에 높은 영향을 미치는 변수를 확인할 수 있었다. 건축물 화재 리스크 분석의 경우 도로 연결성, 인구 수, 온도, 도로 폭, 주거용도 및 상업용도 건물 수 등 도시환경, 사회환경, 건축물 속성 등 다양한 요인이 고르게 영향을 미치고 있는 것으로 확인되었다. 건축물 홍수 리스크 분석 모델의 경우 건축물의 용도와 구조 그리고 용적률과 같은 건축물의 속성이 모델에 중요한 영향을 미치는 것으로 나타났다.

Chapter.4

제4장 건축물 화재 및 홍수 리스크 분석 모델 적용



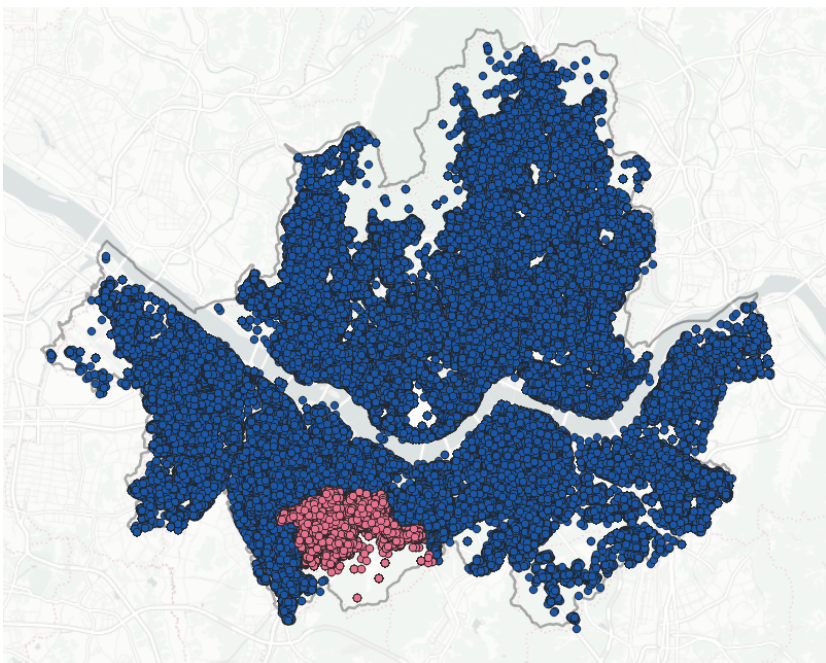
1. 건축물 화재 및 홍수 리스크 분석
2. 건축물 화재 및 홍수 리스크 통합 분석
3. 소결 및 시사점

1. 건축물 화재 및 홍수 리스크 분석

1) 건축물 화재 리스크 분석

□ 분석 모델 적용

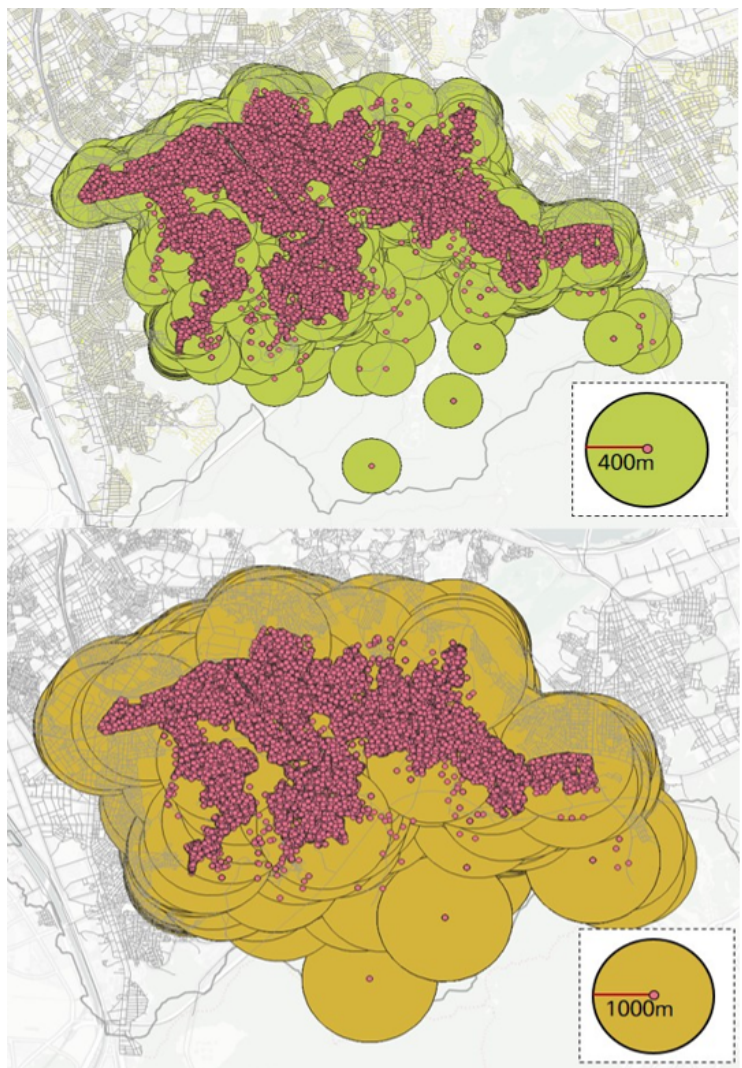
모델을 적용한 대상지는 최근 5년간 화재 발생이 가장 높은 상위 3개의 구 중 관악구를 대상으로 진행하였다. 총 32,079동의 건축물을 대상으로 모델을 적용하였다.



[그림 4-1] 관악구 건축물 포인트 지도
출처: 연구진 작성

분석 모델 적용은 화재 예방 및 리스크 분석에 대한 실질적인 유효성을 검증하기 위해 중요한 단계로, 독립변수 구축과 모델 적용 및 결과 분석의 총 두 단계로 진행되었다. 독립변수 구축의 경우 건축물 단위로 개별 분석을 진행하여 도로 연결성 및 인구수 관련 변수를 모델 학습데이터에 추가적으로 구축하는 과정을 필요로 하였다.

첫 번째 단계인 독립변수 구축은 데이터의 전처리 및 공간정보 연계를 통해 건축물 화재 리스크 분석을 위한 데이터셋을 구축하는 과정이다. 먼저, 모델의 입력값으로 각 건축물 주변의 공간정보를 적용하기 위해 데이터 전처리를 진행하였다. 주요 독립변수인 도로 연결성 관련 변수는 1,000m의 버퍼를 활용하여 데이터를 구축하였고, 도로 폭과 관련된 변수도 같은 방식으로 1,000m의 버퍼를 활용하였다. 또한, 건물용도와 추정 인구와 관련된 변수는 400m의 버퍼를 활용하여 데이터를 구축하였다. 날씨와 관련된 정보는 화재가 발생한 날의 데이터를 랜덤하게 추출하여 모델에 적용하였다.



[그림 4-2] 건축물 주변 공간정보 전처리 과정
출처: 연구진 작성

두 번째 단계인 모델 적용 및 결과 분석은 기존에 학습한 건축물 화재 리스크 분석 모델을 첫 번째 단계에서 구축한 관악구 건축물 데이터에 적용하는 과정과 이에 따라 도출된 결과의 해석 단계로 구분할 수 있다.

모델을 관악구의 32,079동 건물포인트에 적용한 결과, 피해가 큰 건물(1)로 분석된 건물포인트는 8,405동, 피해가 적은 건물(0)로 분석된 건물포인트는 23,674동으로 추출되었다. 즉, 약 31%의 건물이 화재 발생 시 피해금액이 클 것으로 분석되었으며, 69%의 건물이 피해금액이 적을 것으로 분석되었다.

□ 분석 결과 해석

랜덤포레스트 모델을 통해 중요도가 상위인 4개의 변수로 추출된 도로 연결성, 인구 수, 온도, 도로 폭을 중심으로 분석된 모델 결과를 해석하였다. 이를 통해 해당 변수들이 화재피해 정도에 미치는 영향을 해석하였다.

[표 4-1] 분석 모델 적용 결과: 화재

변수 명	리스크 낮은 건축물 평균 값	리스크 높은 건축물 평균 값
도로 연결성	2.68	2.61
추정인구 수	344.11	347.19
기온(°C)	16.46	9.95
도로 폭(m)	5.73	5.59

출처: 연구진 작성

도로 연결성은 화재 리스크 수준 분석에서 상위 4개 변수 중 하나로 나타났다. 이 변수는 한 도로에서 이어진 다른 도로의 개수를 나타내며, 연결성이 낮을수록 해당 도로로의 접근성이 감소한다. 피해가 큰 건물 주변의 도로 연결성 평균은 2.61로 피해가 적은 지역의 2.68보다 낮게 나타났다. 이를 통해 연결성이 낮은 지역에서 화재 발생 시 사후 대처에 어려움을 겪을 가능성이 높다는 것을 유추할 수 있다.

분석 결과에 따르면 인구 수 역시 상위 4개 변수 중 하나로 식별되었다. 피해가 큰 건물 주변의 추정인구 수 평균은 약 347명으로, 피해가 적은 건물 주변의 인구수(약 344명)보다 높게 나타났다. 이를 통해 인구가 많을수록 화재 발생 시 재산피해액이 커질 가능성이 있다는 해석을 할 수 있다. 인구가 많은 지역은 구조물 등이 더 밀집되어 있을 수 있으며, 이로 인해 화재 피해가 증가할 수 있다.

기온 변수를 살펴보면, 낮은 기온에서 피해가 더 취약하다는 해석이 가능하다. 기온이 낮을수록 화재 발생가능성 및 피해규모가 증가할 가능성이 있으며, 이는 낮은 기온 조건에서의 화재 대응 및 피해 감소 노력의 필요성을 시사한다.

마지막으로, 도로 폭 변수는 화재 리스크 분석에서 중요한 역할을 하는 것으로 확인되었다. 피해가 적은 건물 주변의 도로 폭 평균은 5.73m로 나타나고, 피해가 큰 건물 주변의 도로 폭은 5.59m로 나타났다. 피해가 큰 건물 주변의 도로 폭이 더 좁은 것을 확인할 수 있으며, 이는 화재 발생 이후 소방차의 접근성 등 사후 대처에 취약할 수 있음을 시사한다.

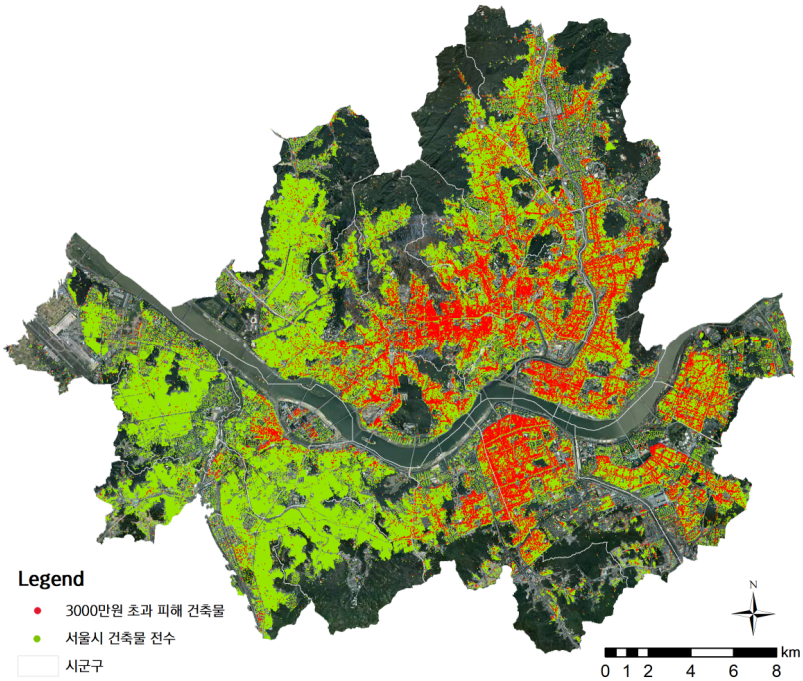
이러한 결과를 종합적으로 고려하면, 도로 연결성, 인구 수, 기온, 도로 폭 등의 변수가 화재 리스크 분석 및 화재 대응에 중요한 영향을 미치며, 개발한 화재 리스크 모델이 이러한 요소들을 고려한 효과적인 화재 예방 및 대응 정책을 수립하는 데 정량적 데이터 값을 제시하여 도움이 될 것으로 판단된다. 또한, 이러한 변수들을 관리하고 개선하는 방향으로 화재 위험을 감소시키는 노력이 필요하다.

2) 건축물 홍수 리스크 분석

□ 분석 모델 적용

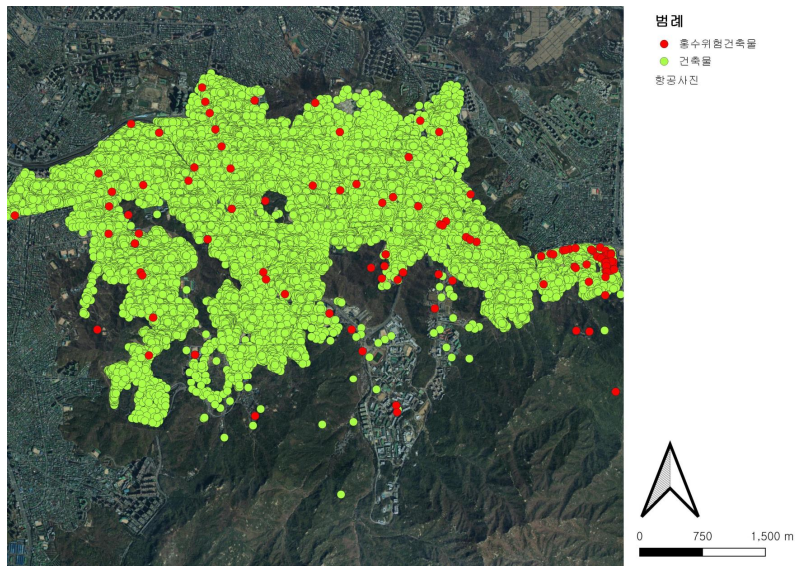
앞서 도출된 5개 유형의 모델 중에서 향후 정책적으로 적용하기에 가장 적합한 기준으로 예상되는 모델은 5번째 유형의 모델인 300만원 이하와 초과의 2개 클래스로 구축된 딥러닝 모델이다. 따라서 이 모델을 기준으로 건축물 홍수 피해 리스크 분석 결과를 살펴보았다. 서울시 내에서 홍수 피해 정보가 존재하지 않는 596,244개 건축물에 대해 10-fold CV를 이용한 10개의 세부모델로 평가한 결과에서 세부모델 5개 이상에서 300만원을 초과하는 피해를 받을 것으로 분석된 건축물을 추출하였다. 이는 앙상블 기법을 응용한 것으로, 서로 다른 데이터로 구축된 10개 세부모델에서 분석된 결과를 종합하여 불확실성을 줄이는 데 도움을 줄 수 있다.

5회 이상 300만원을 초과하는 홍수 피해를 받을 것으로 분석되는 건축물의 수는 전체 건축물(596,244개)의 약 12.11%에 해당하는 72,237개로 나타났으며, 서울시의 중심부, 북부, 동부에 많이 분포하는 것으로 확인되었다. 이는 홍수 민감 지도가 주로 서울시의 남부에 집중되었던 것과는 다른 양상이다. 홍수 민감 지도 구축에 사용된 기계학습 기반 모델이 위치정보를 종속변수로 하고, 독립변수로는 자연적인 요소와 인공적인 요소를 종합적으로 반영하는 반면에, 딥러닝 기반 모델은 피해금액을 종속변수로 하고, 독립변수로 건축물의 특성에 관련된 요소에 집중하고 있기 때문에 발생한 차이점이라고 판단된다. 종속변수와 독립변수가 서로 다르게 적용되었으므로 결과의 신뢰성 비교에 대해서는 검증 데이터의 확보를 통한 추가적인 고찰이 필요하다고 사료된다.



[그림 4-3] 서울시 건축물 홍수 리스크 지도
출처: 연구진 작성

관악구를 대상으로 분석한 경우, 32,079동의 건축물 중 홍수 리스크가 높은 건축물은 총 446동으로 나타났다. 이는 관악구 전체 건축물 중 약 1.5%에 해당하는 건축물로 서울시 전체와 비교할 때 관악구에 있는 건축물의 홍수 리스크는 상대적으로 낮음을 확인할 수 있다. 해당 결과는 홍수 리스크를 피해비용만을 이용하여 산출한 결과로 해석되며, 향후 인명피해를 리스크 산정에 포함할 경우 보다 개선된 결과를 얻을 수 있을 것으로 판단된다.



[그림 4-4] 관악구 건축물 홍수 리스크 지도
출처: 연구진 작성

□ 분석 결과 해석

건축물 홍수 리스크 모델에서 중요도가 높게 나온 변수 중 연속형 변수인 높이와 용적률를 우선적으로 살펴보았다. 평균적으로 높이가 높은 건축물이 홍수 리스크가 큰 것으로 나타났다. 다만, 용적률의 경우 관악구 건축물 중 용적률 항목이 기입되지 않은 경우가 많아 통계적인 의미를 확인하기 어려웠다. 추가적으로 건축면적, 연면적, 하천으로부터의 거리, 고도차 등의 변수에서 리스크가 낮은 건축물과 높은 건축물 사이의 평균의 차이가 클 것으로 나타났다. 상대적으로 연면적이 큰 건축물이 홍수 리스크가 높은 것으로 나타났으며, 하천으로부터 가까울수록 홍수에 대한 리스크가 높게 나타났다. 또한, 주변보다 상대적으로 낮은 지역에 위치한 건축물의 홍수 리스크가 그렇지 않은 건축물보다 높은 것으로 확인되었다.

[표 4-2] 분석 모델 적용 결과: 홍수

변수 명	리스크 낮은 건축물 평균 값	리스크 높은 건축물 평균 값
높이	8.3m	10.8m
건축면적	97㎡	898㎡
연면적	641㎡	3,276㎡
하천으로부터 거리	616m	395m
고도차	24m	-25m

출처: 연구진 작성

건축물 용도의 경우 전체 건축물 중 홍수위험 건축물의 비율인 약 1.5%보다 상대적으로 높은 5% 이상 비율로 나타난 용도를 확인해보았다. 다만, 관악구 내 해당 용도 건축물의 총 동수가 100동 미만인 경우는 제외하였다. 결과적으로 업무시설, 교육연구시설 등이 다른 용도 건축물에 비해 홍수 리스크가 높은 것을 확인할 수 있었다.

건축물 구조 또한 전체 건축물 중 홍수위험 건축물의 비율인 약 1.5%보다 높은 5% 이상 비율이며, 해당 구조의 건축물 동 수가 100동 이상인 것을 확인해보았다. 결과적으로 일반철골구조 건축물이 다른 구조 건축물에 비해 홍수 리스크가 높은 것으로 나타났다.

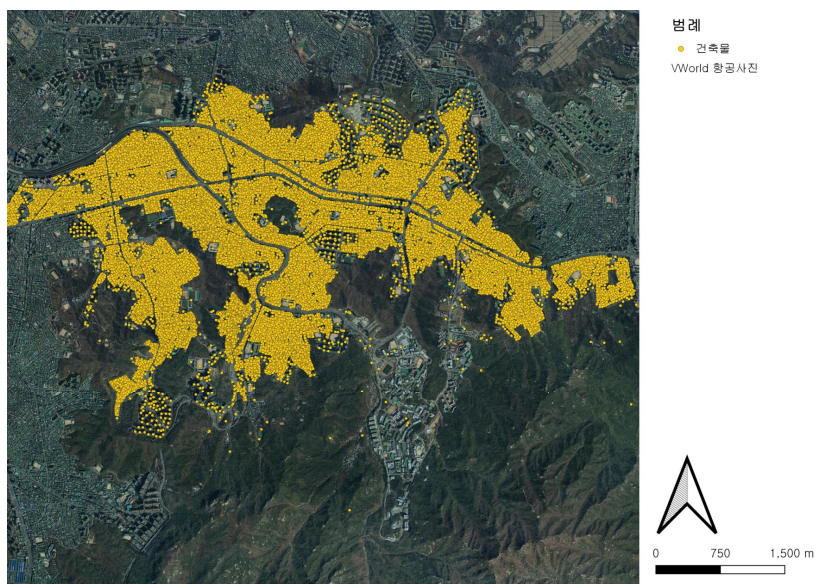
[표 4-3] 분석 모델 적용 결과: 홍수(건축물 용도 및 구조)

변수 명	리스크 낮은 건축물 비율	리스크 높은 건축물 비율
업무시설(용도)	6%	94%
교육연구시설(용도)	30%	70%
일반철골구조(구조)	27%	73%

출처: 연구진 작성

2. 건축물 화재 및 홍수 리스크 통합 분석

본 연구는 건축물의 화재 리스크와 홍수 리스크를 개별적으로 분석하였으나, 두 결과 모두 건축물 단위에서 이루어져 건축물 단위의 통합이 가능하다. 이에, 건축물 화재 리스크 분석 결과와 건축물 홍수 리스크 분석 결과를 중첩하여 건축물 화재 및 홍수 리스크를 복합적으로 분석하고자 하였다. 통합분석을 위해 개별적으로 분석된 건축물 화재 및 홍수 리스크 분석 결과를 관악구 건축물 포인트 데이터에 연계하였다. 관악구에는 총 32,079동의 건축물이 있으며, 건축물의 분포는 아래 그림과 같다.



[그림 4-5] 서울시 관악구 내 건축물 분포 현황
출처: 연구진 작성

건축물 화재 및 홍수 리스크 분석 결과를 통합하는 경우 총 4가지 유형의 건축물을 도출할 수 있다. 화재 리스크와 홍수 리스크가 모두 낮은 건축물, 화재 리스크는 높고 홍수 리스크는 낮은 건축물, 화재 리스크는 낮고 홍수 리스크는 높은 건축물, 그리고 화재 리스크와 홍수 리스크가 모두 높은 건축물로 구분할 수 있다.

[표 4-4] 건축물 리스크 유형 및 동 수

	화재 리스크 낮음	화재 리스크 높음
홍수 리스크 낮음	23,276동 (72.56%)	8,357동 (26.05%)
홍수 리스크 높음	398동 (1.24%)	48동 (0.15%)

출처: 연구진 작성

노란색: 화재 리스크 낮음, 빨간색: 화재 리스크 높음

[그림 4-6] 건축물 화재 리스크 지도

출처: 연구진 작성

노란색: 홍수 리스크 낮음, 파란색: 홍수 리스크 높음

[그림 4-7] 건축물 홍수 리스크 지도

출처: 연구진 작성

관악구 내 건축물 중 화재 리스크가 높은 건축물은 총 8,405동으로 분석되었으며, 홍수 리스크가 높은 건축물은 총 446동으로 분석되었다. 리스크가 높은 건축물을 시각화하기 위해 네 가지 구분을 사용하였다. 첫째는 화재 및 홍수 리스크가 모두 낮은 건축물이다. 둘째는 화재 리스크가 높은 건축물이고 셋째는 홍수 리스크가 높은 건축물이다. 마지막으로 네 번째는 화재와 홍수 리스크가 모두 높은 건축물을 구분하였다.

노란색: 화재 및 홍수 리스크 모두 낮음, 빨간색: 화재 리스크 높음, 파란색: 홍수 리스크 높음, 별표: 화재 및 홍수 리스크 모두 높음

[그림 4-8] 건축물 홍수 및 화재 리스크 지도
출처: 연구진 작성

결과적으로 관악구 내 건축물 중 화재 또는 홍수 리스크가 모두 낮은 건축물은 23,276동으로 전체 건축물 중 약 72.56%가 이에 포함된다. 화재 리스크만 높은 건축물은 8,357동으로 전체 건축물 중 약 26.05%가 이에 포함되고, 홍수 리스크만 높은 건축물은 398동으로 전체 중 약 1.24%가 이에 해당한다. 건축물 화재 및 홍수 리스크가 모두 높은 건축물은 총 48개 동으로 전체 건축물 중 0.15%가 해당되는 것으로 나타났다.

건축물 재난 및 재해 리스크와 공간분포를 비교해볼 때 화재발생 리스크가 높은 건축물은 상대적으로 균등하게 분포하고 있으나, 홍수 리스크가 높은 건축물은 특정 지역에 밀집하는 형태로 나타나고 있다.

3. 소결 및 시사점

본 장에서는 건축물 화재 리스크 분석 모델과 건축물 홍수 리스크 분석 모델을 이용하여 서울시 관악구 내 모든 건축물들에 대해 각각 화재 리스크와 홍수 리스크를 도출하였다. 또한, 건축물 단위로 도출된 결과물을 결합하여 건축물의 화재 및 홍수 리스크를 복합적으로 확인하는 방법을 검토하였다.

1) 리스크 분석 모델

기존 연구들이 재난·재해 발생 가능성에 집중하여 모델을 개발했다면, 본 연구는 피해금액의 높고 낮음을 기준으로 하는 리스크에 집중하여 모델을 개발하였다. 건축물 화재 리스크 분석 모델과 건축물 홍수 리스크 분석 모델은 건축물에 화재나 홍수가 발생한 경우 발생가능한 피해액을 분석하는 모델이다. 즉, 건축물에 화재나 홍수가 발생할 가능성을 분석하기보다는 발생을 가정했을 때 피해 금액이 클 것으로 분석되는 건축물을 선별할 수 있다.

건축물 홍수 발생가능성을 나타내는 홍수 민감 지도와 건축물 홍수 리스크 분석 지도를 비교해볼 때 발생확률과 리스크 분석이 서로 다름을 확인할 수 있다. 건축물 홍수 민감 지도에서는 관악구 및 강남구 등 한강 남쪽 지역이 대체적으로 홍수 발생 확률이 높은 것으로 나타나고 있다. 반면 건축물 홍수 리스크 분석에서는 관악구 등 서울 남서 지역의 리스크가 전반적으로 낮음을 확인할 수 있으며, 강북 지역의 리스크가 높음을 확인할 수 있다.

뿐만 아니라 홍수 발생 가능성 분석과 리스크 분석은 건축물 차원의 대응 측면에서도 차이가 있다. 기존의 재난·재해 피해 발생 예측 모델은 화재나 홍수가 발생할 가능성이 있는 건축물을 도출하고 있다. 건축물 화재의 경우 건축물 내부에서 발생하는 문제인 경우가 많아 건축물 차원의 대응이 가능하나, 건축물 홍수 피해의 경우에는 도시 및 지역단위에서 발생하는 문제로, 건축물 차원에서의 대응책 마련이 어려운 것이 사실이다.

리스크 분석 모델의 경우, 피해금액에 초점을 맞춤으로써 홍수 발생 자체에 대비하는 것이 아닌 홍수 발생 시 피해금액을 최소화 하는 대응방안을 제시할 수 있다. 즉, 이미 홍수가 발생한 경우 건축물의 피해금액이 증가하거나 감소하는 원인은 건축물 내부의 문제에 기인하는 경우가 많으며, 본 연구의 리스크 분석 모델은 이에 초점을 맞추었다.

결과적으로 건축물의 화재와 홍수 리스크가 높은 건축물의 위치를 확인할 수 있었으며, 리스크를 높이는 요인이 무엇인지 통계적으로 확인할 수 있었다. 다만, 화재의 경우 피해금액을 높이는 요인이 건축물 관련 변수보다는 도시환경 및 사회환경 관련 변수로 나타나 이에 대한 추가적인 연구가 필요할 것으로 판단된다. 건축물 홍수 리스크의 경우 건축물의 높이, 연면적, 용도 등 건축물 속성정보가 영향을 많이 미치는 것으로 나타났으며, 이외에도 주변 지역과의 고도차, 하천으로부터의 거리 등 도시환경 변수도 영향을 미치는 것으로 확인되었다.

2) 건축물 화재 및 홍수 리스크 분석 모델

서울시 관악구에 대해 개별적으로 도출한 건축물 화재 리스크 분석 결과와 건축물 홍수 리스크 분석 결과를 연계하여 건축물 화재 및 홍수 리스크를 분석하였다. 건축물 단위 화재 및 홍수 리스크의 높고 낮음을 이용하여 화재 리스크와 홍수 리스크가 모두 낮은 건축물, 화재 리스크는 높고 홍수 리스크는 낮은 건축물, 화재 리스크는 낮고 홍수 리스크는 높은 건축물, 그리고 화재 리스크와 홍수 리스크가 모두 높은 건축물의 네 가지 유형으로 건축물을 구분할 수 있었다.

관악구를 대상으로 분석한 결과 홍수 리스크와 화재 리스크가 모두 낮은 건축물이 약 72%로 가장 많았으며, 화재 리스크만 높은 건축물은 약 26%, 홍수 리스크만 높은 건축물은 약 1.24%, 그리고 화재와 홍수 리스크가 모두 높은 건축물은 0.15%로 매우 낮게 나타났다. 다만 이같은 수치는 관악구 내에 홍수 리스크가 높은 건축물의 수가 상대적으로 적기 때문인 것으로 판단되며, 지역에 따라 변동폭이 클 수 있을 것이다.

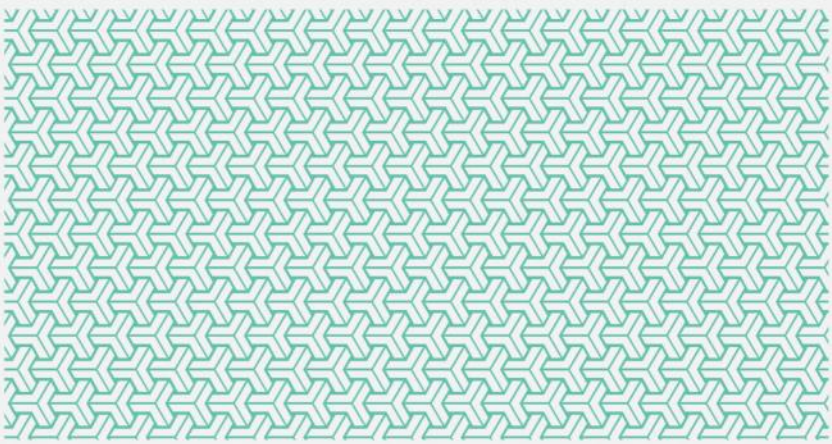
[표 4-5] 건축물 리스크 유형

	화재 리스크 낮음	화재 리스크 높음
홍수 리스크 낮음	L/L	L/H
홍수 리스크 높음	H/L	H/H

출처: 연구진 작성

제5장 결론

Chapter.5



1. 연구의 요약
2. 정책적 시사점
3. 연구의 한계 및 향후 연구 과제

1. 연구의 요약

1) 재난·재해 리스크 분석 동향

재난·재해 관련 선행연구와 모델을 종합적으로 검토한 결과 리스크는 일반적으로 발생확률과 손상함수의 연계를 통해 산정하며, 피해금액을 리스크로 사용하는 것을 확인하였다. 이는 기존 발생확률에 집중하던 방식에서 더 나아가 실질적인 피해를 산정하기 위한 노력으로 판단된다. 기존 건축·도시공간의 화재 및 홍수 분석 연구는 모두 재난·재해 연구이지만 다른 성격으로 분석되어온 것으로 보인다.

화재 예측 모델은 통계 중심이며, 발생 강도와 피해액을 분석하지만 화재 발생 자체 예측은 어려운 것으로 정리할 수 있다. 화재의 특성 자제가 자연발화의 가능성이 낮고, 실화 및 방화 등 인재에 의해 발생되는 경우가 많아 시뮬레이션 기반의 모델링보다는 통계적 분석이 주를 이루는 것으로 판단된다. 리스크 분석 연구의 경우 이 같은 통계적 화재 발생 모델에 기반하며, 추가적으로 손상함수 등 피해액 산정방식과 연계하여 리스크를 도출하고 있다. 건축물에 대한 리스크 분석 또한 통계 기반의 화재 발생 모델과 손상함수를 적용한 피해액 산정방식이 일반적인 추세라고 볼 수 있다.

홍수의 경우 수리·수문 모델링 기반의 홍수 발생과 강도 예측이 활발히 이루어지고 있다. SWMM과 SWAT 등이 대표적인 모델로 확인되며, 해당 모델들은 연구와 지역 홍수 대책 마련 등 광범위하게 적용되고 있다. 홍수의 경우 화재와 달리 강수량 변수를 변화시켜가며 다양한 상황에 대한 홍수 모의가 가능한 상황이다. 이에 시뮬레이션 기반의 홍수 발생확률 및 발생강도 모델링과 홍수 발생 시 예상되는 피해금액을 산정하는 손상함수를 결합하여 홍수 리스크 분석이 이루어지고 있다. Hazus-MH, HEC-FIA 등이 대표적인 홍수 리스크 분석 모델이며, 건축물 단위의 홍수 리스크 분석도 함께 진행되고 있다.

본 연구는 통계기반의 딥러닝 모델을 활용하여 건축물 화재와 홍수 리스크 분석 모델을 개발하였다. 분석에는 딥러닝 모델을 적용함으로써 기존 화재발생 모델의 통계기반 분석방식을 유지할 수 있다, 또한, 본 연구에서 사용된 홍수발생 데이터 중 하수도 역류 피해 데이터와 같이 시뮬레이션 기반 모델에서는 다루기 어려운 데이터를 다루기에도 적합한 것으로 판단하였다. 마지막으로, 본 연구에서 사용한 건축물 화재 및 홍수 피해 데이터에 피해금액 정보가 기입되어있어, 딥러닝 모델이 자연스럽게 손상함수를 내포할 수 있도록 개발할 수 있었다.

2) 건축물 화재 리스크 분석 모델 개발

건축물 화재 리스크 분석 모델 개발에는 2017년부터 2021년까지의 서울시 건축물 화재피해 데이터를 사용하였다. 또한, 리스크 분석을 위해 건축물대장 정보와 추정인구 수, 도로 연결성, 도로폭 등 선행연구를 통해 도출된 독립변수를 연계하여 모델 학습에 활용하였다. 최종적으로 선택된 변수는 건축 및 도시 환경특성 관련 변수, 건축물 자체특성 관련 변수, 인구통계 및 날씨특성 관련 변수로 구분 가능하며, 총 14개 독립변수가 활용되었다.

종속변수의 경우 건축물 화재 피해액을 이용하였으며, 피해금액이 20만원 이상인 건축물의 경우 화재 리스크가 높은 건축물로, 20만원 미만인 경우 화재 리스크가 낮은 건축물로 구분하여 분석을 진행하였다. 학습에는 총 28,266건의 화재 발생 데이터가 사용되었으며, 랜덤포레스트, 로지스틱 회귀 분석, LightGBM, XGBoost 등 다양한 기계학습 알고리즘을 적용하여 가장 높은 정확도를 나타낸 랜덤포레스트 모델을 선택하였다.

모델은 기본적으로 분류 정확도를 평가하였으며, F1점수를 이용하여 정밀도와 재현율을 반영한 평가와 ROC 곡선을 이용한 분류성능 평가를 실시하였다. 결과적으로 건축물 화재 리스크 분석 모델의 최종 정확도는 78% 수준으로 타나났으며, F1점수는 0.75로 준수한 성능을 나타내는 것을 확인하였다. 마지막으로 ROC 곡선의 경우 곡선 아래 부분의 면적을 나타내는 AUC값이 0.89로 우수한 분류 능력이 있음을 확인하였다.

건축물 화재 리스크를 분석하는데 중요하게 사용된 영향요인으로는 도로 연결성, 추정인구 수, 기온, 도로 폭, 주거용도 수, 상업용도 수 등이 도출되었다. 도로 연결성과 도로 폭의 경우 건축물 화재 발생 시 화재 진압을 위한 차량의 접근성과 관련성이 깊은 것으로 판단된다. 추정인구 수의 경우 거주인구가 많은 지역일 수록 재산피해가 발생할 수 있는 인프라, 상품 등의 재산 밀집도가 높기 때문인 것으로 판단된다. 주거용도 수와 상업용도 수의 경우 건축물 화재 피해액 변화에 건축물의 용도가 영향을 많이 미치는 것으로 해석할 수 있다.

3) 건축물 홍수 리스크 분석 모델 개발

건축물 홍수 리스크 분석 모델 개발에는 2016년부터 2022년까지의 서울시 건축물 홍수피해 데이터를 활용하였다. 건축물 홍수피해 데이터는 건축물 화재피해 데이터와는 달리 피해 발생 건수가 매년 고르게 분포하지 않으며, 집계된 데이터 수가 상대적으로 적어 보다 긴 기간의 데이터를 필요로 하였다.

리스크 분석을 위해 건축물대장 정보와 더불어 선행연구를 통해 도출된 지형 및 토양 관련 변수, 인공환경 관련 변수 등 다양한 변수를 추가적으로 구축하였다. 최종적으로 21개 독립변수가 사용되었다. 독립변수 중 홍수 민감 지도의 경우 15개 관련 변수를 이용하여 분석하여 도출하였으며, 이를 건축물 홍수 리스크 분석 모델 학습의 한 변수로 활용하였다.

종속변수의 경우 건축물 홍수 피해액을 이용하였다. 모델의 정확도를 높이기 위해 홍수 피해액을 100만원 단위의 총 11개 클래스로 학습하는 경우, 200만원 단위의 총 6개 클래스로 학습하는 경우, 300만원 단위의 총 4개 클래스로 학습하는 경우, 500만원 단위의 총 3개 클래스로 학습하는 경우, 마지막으로 300만원 이하와 초과의 총 2개 클래스로 학습하는 경우 등 5가지 데이터셋을 구성하여 학습을 수행하였다. 건축물 홍수 리스크 분석 모델은 TabNet classifier를 사용하였으며, 최종적으로 300만원 이하와 초과의 총 2개 클래스로 학습한 모델을 채택하였다. 채택된 모델의 정확도는 88% 수준으로 나타났다.

독립변수 71개 중 건축물 홍수 리스크 분석에 영향력이 높은 것으로 도출된 변수는 건축물의 용도와 구조, 그리고 건축물의 높이와 용적률 등으로 나타났다. 건축물의 용도나 구조의 경우 원-핫 인코딩을 통해 항목별로 변수가 분리되었으며, 단독주택 여부가 가장 높은 중요도인 0.192를 나타내었다. 건축물의 구조 중 기타콘크리트 구조와 벽돌구조는 각각 중요도 0.044와 0.043을 나타내었고, 건축물의 높이와 용적률은 각각 0.064와 0.033을 나타냈다.

4) 건축물 화재 및 홍수 리스크 통합 분석 모델 개발

건축물 화재 리스크 분석 모델과 건축물 홍수 리스크 분석 모델을 연계하여 건축물 화재 및 홍수 리스크 분석 모델을 모의하였다. 이를 위해 서울시 관악구 지역을 대표 지역으로 선정하였으며, 개발된 개별 모델을 활용하여 관악구 내 건축물의 화재 리스크와 홍수 리스크를 분석하였다. 각 모델의 분석 결과가 건축물 단위로 도출되기 때문에 건축물 단위에서의 결과 연계가 용이하였다. 연계된 데이터를 이용하여 개별 건축물의 화재 및 홍수 리스크를 도출하였다.

기존 개별 모델은 화재 리스크 낮음과 높음, 그리고 홍수 리스크 낮음과 높음으로 각각 2가지의 결과를 도출한다. 두 모델을 연계할 경우 건축물의 화재 및 홍수 리스크가 모두 높은 경우, 모두 낮은 경우 혹은 한 가지의 리스크만 높은 경우로 총 4가지 경우의 수가 도출되었다.

2. 정책적 시사점

□ 재난·재해 대응을 위해 제4차 산업혁명 기술 활용성 제고 필요

본 연구는 화재와 홍수를 중심으로 건축물 재난·재해 리스크 분석 모델을 개발하고 서울시 일부 지역에 적용하였다. 이를 위해 최근 다양한 분야에서 문제 해결을 위해 적용되고 있는 머신러닝, 딥러닝 기술을 활용하였으며, 건축물 속성정보와 도로, 인구 등 다양한 공간정보 및 사회환경 정보를 결합하여 활용하였다. 머신러닝, 딥러닝 기술은 제4차 산업혁명을 대표하는 기술 중 하나로 여겨지고 있고, 다양한 공간정보의 연계는 제4차 산업혁명이 추구하는 빅데이터 구축을 표방한다. 즉, 연구를 통해 제4차 산업혁명 기술의 발달에 발맞추어 건축물의 재난·재해 문제 또한 해결할 수 있을 것이라는 방향성을 제시하고 있다. 또한, 연구의 추진과정에서 나타나는 데이터 확보의 어려움, 데이터 간 연계성 부족 등은 건축물 재난·재해 리스크 저감을 위해서도 해결되어야 할 문제로 판단할 수 있다.

□ 건축물 화재 및 홍수 리스크 저감 방안

건축물 화재 리스크 분석 모델에 높은 영향을 미치는 변수로는 도로 연결성, 추정 인구 수, 기온, 도로 폭, 주거용도 수 등이 도출되었다. 도출 결과를 통해 도로와 같이 교통망과의 접근성 측면이 화재 위험에 큰 영향을 미치는 것으로 나타나, 도로 연결성과 관련된 정책 개발은 화재피해 예방 및 대응에 도움이 될 것으로 보인다. 인구 수와 온도 변수 또한 높은 중요도를 가지고 있다. 따라서 화재로 인한 피해를 예방하는 데 있어서 인구 수 및 온도를 고려할 필요가 있음을 시사한다.

도로 폭과 주거용도의 건물 분포도 또한 중요한 변수로서 해석될 수 있다. 도로 폭이 넓을수록 화재 진압 및 대피가 더 원활해질 수 있으며, 주거용도 건물 수가 많을수록 인적활동이 활발해져 화재 위험이 증가할 수 있다. 하지만, 특정 변수가 화재피해와 어떤 관계인지는 추후 연구를 통해 더 명확하게 규정해야 할 것이다. 본 연구에서는 명확한 관계를 파악하기는 어렵지만 피처 중요도 결과로부터, 건축물 자체 특성뿐만 아니라 건축물 주변 환경 특성이 화재피해에 큰 영향을 미침을 알 수 있다. 따라서 화재 예방 및 대응 정책을 수립할 때 이러한 주변 환경 변수 또한 중점적으로 고려해야 할 것이다. 그리고 이러한 중요한 변수들을 효과적으로 관리하고 개선함으로써 화재 위험을 줄일 수 있는 기회를 제공할 수 있다. 이를 통해 지역별 화재 위험을 평가하고 효과적인 화재 예방 및 대응 전략을 수립하는 데 도움이 될 것이다.

건축물 홍수 리스크 분석 모델에 높은 영향을 미치는 변수는 건축물의 용도, 구조, 높이, 용적률 등이 도출되었다. 건축물은 용도에 따라 건축물의 이용행태가 다르고 또한 내부에서 이용 및 보관하는 자재의 비용이 다를 것이다. 이 같은 이유로 홍수 발생 시 건축물의 용도가 피해금액에 많은 영향을 끼치는 것으로 판단된다. 또한, 건축물의 구조는 홍수에 대응할 수 있는 건축물 자체의 대응능력을 일부 반영하는 변수로 판단되며, 향후 강우강도가 증가하고 홍수 규모가 커지는 상황에 대응하기 위해 홍수 대응능력이 좋은 건축구조를 연구할 필요가 있다.

□ 건축물 단위 재난·재해 통합 대응 필요

기존의 연구들이 건축물의 재난·재해 중 하나의 문제만을 다루는 경우가 많았다면 본 연구는 건축물의 화재와 홍수라는 두 가지 문제를 동시에 다루고자 하였다. 화재와 홍수는 발생 원인이나 대응 방식이 전혀 다를 수 있으나, 건축물 차원에서의 대응 우선순위를 도출한다는 측면에서는 이를 복합적으로 보는 것이 더 효율적일 수 있다고 판단된다. 즉, 본 연구를 통해 확인할 수 있는 건축물 화재 리스크와 홍수 리스크가 모두 높은 건축물을 건축물의 재난·재해 대비 관점에서 우선적으로 고려할 필요가 있다.

성공적인 건축물 재난·재해 예방을 위해 건축물 통합예방안전 시스템 구축이 필요하다. 건축물 단위 개별적인 재난·재해 대응은 국민의 안전 개선 및 국가 차원의 활용 측면에서 효율성이 떨어질 수 있다. 본 연구에서 제안한 건축물 화재 및 홍수 뿐 아니라, 지진, 산사태, 열해 등 다양한 재난안전을 통합적으로 고려할 때 건축물의 안전에서 더 나아가 건축물을 이용하는 국민의 안전을 확보할 수 있을 것이다.

□ 재난·재해 취약 건축물 우선 관리를 위한 제도개선 필요

연구결과 공공데이터와 기계학습을 통하여 화재와 풍수해에 보다 취약한 건축물의 목록을 도출할 수 있었다. 이렇게 도출된 취약건축물의 대부분은 현재 「건축물관리법」, 「소방시설 설치 및 관리에 관한 법률」, 「시설물의 안전 및 유지관리에 관한 특별법」 등에 따른 의무 점검 대상건축물 보다는 규모가 작은 소규모건축물로 나타났다.

사실상 안전에 취약한 소규모 건축물은 정기적인 법적 점검이 이루어지지 못하고 있는 실정이다. 정부도 소규모건축물에 대한 점검 필요성을 공감하여 관계법령에서 이를 시행할 수 있도록 제도를 설계하였다. 다만 막대한 물량의 소규모건축물의 점검에 필요한 예산과 인력 문제로 관련 제도 시행이 제대로 이루어지지

않는 문제가 있다. 본 연구의 결과를 활용하면 위험도가 높은 소규모건축물의 우선 순위를 도출할 수 있으며, 이를 통해 보다 효과적인 건축물 안전관리가 가능할 것으로 기대된다. 따라서 상기한 건축물 통합예방안전 시스템 구축을 선행하고 이에 기반하여 관계법령의 개정을 추진하는 등 취약 건축물 우선 점검 및 관리를 제도화할 필요가 있다.

건축물관리법

제15조(소규모 노후 건축물등 점검의 실시) ① 특별자치시장 · 특별자치도지사 또는 시장 · 군수 · 구청장은 다음 각 호의 어느 하나에 해당하는 건축물 중 안전에 취약하거나 재난의 위험이 있다고 판단되는 건축물을 대상으로 구조안전, 화재안전 및 에너지성능 등을 점검할 수 있다.
1. 사용승인 후 30년 이상 지난 건축물 중 조례로 정하는 규모의 건축물
2. 「건축법」 제2조제2항제11호에 따른 노유자시설
3. 「장애인 · 고령자 등 주거약자 지원에 관한 법률」 제2조제2호에 따른 주거약자용 주택
4. 그 밖에 대통령령으로 정하는 건축물

출처 : 건축물관리법. 법률 제19367호, 2023. 4. 18.. 일부개정

3. 연구의 한계 및 향후 연구 과제

본 연구는 서울시 데이터를 활용하여 수행되었으나, 다른 도시나 지역 데이터를 활용한 연구를 수행할 필요가 있다. 개별 도시 및 지자체는 서로 다른 지역별 특성을 나타내며, 본 연구의 결과가 서울시라는 특수 해에만 적용가능하지 않다는 것을 확인해볼 필요가 있다. 각 지역은 도시 규모, 지리적 특성, 기후, 인구 구조, 경제 활동 등 다양한 변인들이 서로 다르며, 이 같은 요인에 의해 모델 학습이나 분석 결과가 달라질 수 있어 이를 확인하는 과정이 필요하다.

머신러닝 및 딥러닝을 이용한 건축물 재난·재해 분석의 관점에서 본 연구의 모델은 건축물 화재 리스크 분석의 경우 정확도 약 78%, 건축물 홍수 리스크 분석의 경우 정확도 약 88% 수준으로, 기존 조영진 외(2022)의 모델에 비해 비약적으로 정확도를 상승시켰다. 그러나 사전예방적 건축물 관리의 효율성 제고를 위해 분석 정확도를 높일 필요가 있다. 건축물 재난·재해 리스크 분석 또한 더 높은 정확도 확보를 위해 추가적인 데이터 확보 및 딥러닝 기술 적용방식 다각화를 수행해 볼 필요가 있다.

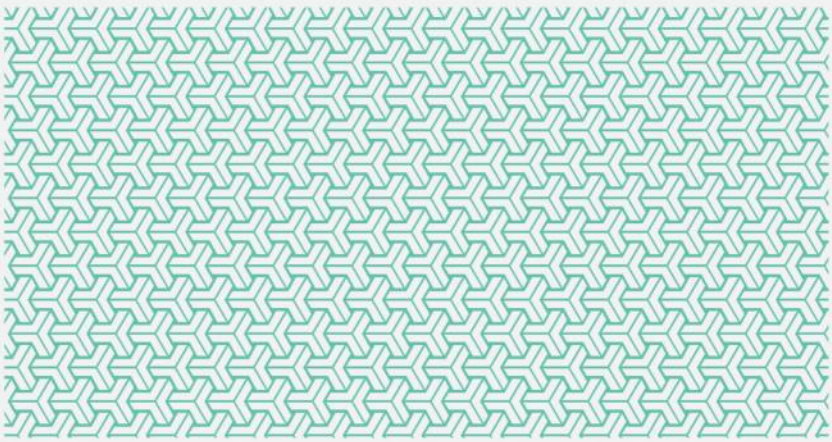
분석 정확도 개선을 위해 모델 뿐 아니라 다양한 데이터의 연계방안을 지속적으로 모색해야 한다. 정부는 공공데이터 플랫폼을 통해 국가가 구축한 다양한 데이터를 유통하고 있으며, 향후 연구는 이 같은 데이터를 연계하여 활용할 수 있는 방안을 마련할 필요가 있다. 딥러닝 알고리즘은 입력된 다양한 데이터 중에서 분석에 필요한 데이터를 자체적으로 선별 및 사용할 수 있으며, 고품질의 입력데이터의 확보가 정확도 개선에 주요한 열쇠가 된다. 즉, 다양한 데이터의 확보방안, 건축물 단위 연계를 위한 데이터 핸들링 방안에 대한 지속적인 연구가 필요하다.

재난·재해에 대응하는 건축물 통합예방안전 확보를 위해 화재 및 홍수 이외의 다른 재난·재해를 모델에 추가할 필요가 있다. 건축물에 피해를 주는 다양한 재난·재해로는 우면산 산사태와 같은 산사태, 온열질환을 일으키는 열해, 태풍과 같은 강풍 등이 있다. 개별 재난·재해에 대한 건축물 단위 재난·재해 리스크 분석 모델을 개발하고, 향후 모델 간 연계를 통해 건축물 통합예방안전 모델을 개발할 수 있을 것으로 기대된다.

본 연구에 사용된 입력변수와 같이 재난·재해에 대한 대응책 마련을 위해서는 건축물 차원의 대응과 지역 및 도시 차원의 대응이 모두 필요한 것으로 판단된다. 이를 위해 재난·재해 발생과 피해액 증감에 대해 건축물 특성이 야기하는 부분과 지역 및 도시 특성이 야기하는 부분을 구분하여 검토할 수 있도록 모델을 개발하는 후속연구 또한 필요할 것으로 판단된다.

참고문헌

References



□ 문헌자료

김재경, 강준석. 2022. SWMM 모델을 활용한 상도동 지역의 우수유출관리 및 저영향개발의 유출저감효과 분석 연구. 한국산학기술학회논문지. 23(8). 한국산학기술학회. pp.337-347

김종계, 박진혁, 윤지현, 정관수. 2008. 남강댐유역을 대상으로 GIS기반의 분포형 모형과 집중형 모형의 유출해석 비교. 한국수자원학회 학술대회논문집. 한국사자원학회. pp.795-799

김효민, 이동근, 박찬. 2013. MaxEnt를 이용한 서울시 도시홍수 적응능력 취약지역 선정에 관한 연구. 국토계획. 대한국토·도시계획학회. 48(4). pp.205-217

노주형, 민세홍, 공민석. 2022. 합성곱 신경망 기반 분류 모델의 화재 예측 성능 분석. 한국화재소방학회 논문지. 36(6). 한국화재소방학회. pp.70-77

조영진, 허한결, 안의순, 류수연, 송유미, 현태환. 2022. 빅데이터 기반 건축물 화재 예측 모델 개발 연구. 건축공간연구원

행정안전부. 2022. 2021 재해연보

허성구, 김남원, 박윤식, 김종건, 김성준, 안재훈, 김기성, 임경재. 2008. SWAT ArcView GIS Extension Patch를 이용한 소유역 분할에 따른 수문 및 유사 거동에 미치는 영향 평가. 한국물환경학회지. 24(2). 한국수질학회. pp.147-155

ACKERE, S. VAN., GLAS, H., BEULLENS, J., DERUYTER, G., WULF, A. DE. and MAE YER, P. DE.. 2016. DEVELOPMENT OF A 3D DYNAMIC FLOOD WEB GIS VIS UALISATION TOOL. *International Journal of Safety and Security Engineering*. 6(3). pp.560-569

Ali, Jehad., Khan, Rehanullah., Ahmad, Nasir. and Maqsood, Imran. 2012. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*. 9(5). pp.272-278

Arik, Sercan Ö. and Pfister, Tomas.. 2020. TabNet: Attentive Interpretable Tabular Learning. arXiv:1908.07442v5 [cs.LG]. https://doi.org/10.48550/arXiv.1908.07442

BIZ-GIS Inc.. nd. Geocoding Tool 매뉴얼. manual ver1. https://www.gislab.co.kr:444/wordpress/geocoding/

BMT. 2018. TUFLOW Classic/HPC User Manual

Bradley, Andrew P.. 1997. The Use of the Area under the ROC curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition*. 30(7). pp.1145-1159

Brody, Samuel D. and Highfield, Wesley E.. 2013. Open space protection and flood mitigation: A national study. *Land Use Policy*. 32. pp.89-95

Brody, Samuel., Blessing, Russell., Sebastian, Antonia. and Bedient, Philip.. 2014. Examining the impact of land use/land cover characteristics on flood losses. *Journal of Environmental Planning and Management*. 57(8). pp.1252-1265

Choo, Yeon-Moon., Kim, Jong-Gu., Park, Shang-Ho., Choo, Tai-Ho. and Choe, Yeon-Woong. 2021. Method for Operating Drainage Pump Stations Considering Downstream Water Level and Reduction in Urban River Flooding. *Water*.13(19). pp.1-15

EPA. 2022. Storm Water Management Model User's Manual Version 5.2

FEMA. 2022. Hazus Flood Technical Manual: Hazus 5.1

FEMA and HAZUS. 2018. Hazus Flood Model: FEMA Standard Operating Procedure for Hazus Flood Level 2 Analysis

Freni, Gabriele. and Liuzzo, Lorena.. 2019. Eectiveness of Rainwater Harvesting Systems for Flood Reduction in Residential Urban Areas. *Water*. 11(7)

Hastie, Trevor., Tibshirani, Robert. and Friedman, Jerome.. 2009. 15 Random forests. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, pp.587-604

IPCC. 2022a. Climate Change 2022: Impacts, Adaptation and Vulnerability. WMO, UNEP

IPCC. 2022b. Climate Change 2022: Mitigation of Climate Change. WMO, UNEP

Iqbal, Asif., Rahman, Md Mizanur. and Beecham, Simon.. 2022. Permeable Pavements for Flood Control in Australia: Spatial Analysis of Pavement Design Considering Rainfall and Soil Data. *Sustainability*. 14(9). pp.1-15

Jang, Jiun-Huei., Hsieh, Chi-Tai. and Chang, Tien-Hao.. 2019. The importance of gullyflow modelling to urbanflood simulation. *URBAN WATER JOURNAL*. 16(5). pp.377-388

Japkowicz, Nathalie.. 2006. Why question machine learning evaluation methods. In AAAI workshop on evaluation methods for machine learning. Technical Report WS-06-06. JULY 16-20. pp.6-11

Karmakar, Subhankar., Simonovic, Slobodan P., Peck, Angela. and Black, Jordan.. 2010. An Information System for Risk-Vulnerability Assessment to Flood. *Journal of Geographic Information System*. 2(3). pp.129-146

Kim, Hyomin., Lee, Dong-Kun. and Sung, Sunyong.. 2016. Effect of Urban Green Spaces and Flooded Area Type on Flooding Probability. *Sustainability*. 8(2). pp.1-17

Li, Ping., Sheng, Mingyang., Yang, Dawen. and Tang, Lihua.. 2019. Evaluating flood regulation ecosystem services under climate, vegetation and reservoir influences. *Ecological Indicators*. 107. pp.1-13

Madaio, Michael A.. 2018. Predictive Modeling of Building Fire Risk: Designing and Evaluating Predictive Models of Fire Risk to Prioritize Property Fire Inspections. Metro21

McGuinness, Robert., Gilbert, Greg., Whittamore, Zakary., Zhong, Leon., Zyl, Ruan Van. and Lee, Jeeyoung.. 2019. A Building Fire Risk Prediction Validation

Project. A Fire Underwriters Survey/Opta Information Intelligence project in partnership with Vancouver Fire Rescue Service and New Westminster Fire Rescue Service.

Menze, Bjoern H., Kelm, B Michael., Masuch, Ralf., Himmelreich, Uwe., Bachert, Peter., Petrich, Wolfgang. and Hamprecht, Fred A.. 2009. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*. 10. pp.1-16

NEITSCH, S.L., ARNOLD, J.G., KINIRY, J.R., WILLIAMS, J.R. and KING, K.W.. 2002. SOIL AND WATER ASSESSMENT TOOL THEORETICAL DOCUMENTATION VERSION 2000. GRASSLAND, SOIL AND WATER RESEARCH LABORATORY and Blackland Research and Extension Center

Novaković, Jasmina Dj., Veljović, Alempije., Ilić, Siniša S., Papić, Željko. and Tomović, Milica.. 2017. Evaluation of Classification Models in Machine Learning. *Theory and Applications of Mathematics & Computer Science*. 7(1). pp.39-46

Panahi, Mahdi., Jaafari, Abolfazl., Shirzadi, Ataollah., Shahabi, Himan., Rahmati, Omid., Omidvar, Ebrahim., Lee, Saro. and Bui, Dieu Tien.. 2021. Deep learning neural networks for spatially explicit prediction of flash flood probability. *Geoscience Frontiers*. 12(3). pp.1-14

Paulik, Ryan., Horspool, Nick., Woods, Richard., Griffiths, Nick., Beale, Tim., Magill, Christina., Wild, Alec., Popovich, Benjamin., Walbran, Glenn. and Garlick, Russel.. 2023. RiskScape: a flexible multi-hazard risk modelling engine. *Natural Hazards*. 119. pp.1073-1090

Pourghasemi, Hamid Reza., Pouyan, Soheila., Bordbar, Mojgan., Golkar, Foroogh. and Clague, John J.. 2023. Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination. *Natural Hazards*. 116. pp.3797-3816

Shafizadeh-Moghadam, Hossein., Valavi, Roozbeh., Shahabi, Himan., Chapi, Kamran. and Shirzadi, Ataollah.. 2018. *Journal of Environmental Management*. 217. pp.1-11

Siahkamari, Safura., Haghizadeh, Ali., Zeinivand, Hossein., Tahmasebipour, Naser. and Rahmati, Omid.. 2018. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto International*. 33(9). pp.927-941

Simon, Michelle.. 2021. Getting Started with EPA's Storm Water Management Model (SWMM). University of Cincinnati Hydrology and Hydraulic graduate and undergraduate class. Cincinnati, OH. October 04-06. https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=353011&Lab=CESER

U.S. Army Corps of Engineers Hydrologic Engineering Center. 2018. HEC-FIA Flood Impact Analysis: Technical Reference Manual

Wang, Qianru., Zhang, Junbo., Guo, Bin., Hao, Zexia., Zhou, Yifang., Sun, Junkai., Yu, Zhiwen. and Zheng, Yu.. 2019. CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 3(4) Article 156. pp.1-21.

Zuo, Daxing., Wu, Chunyi., Zheng, Yanhui., Chen, Xiaohong and Wang, Lina.. 2023. Climate change and human activity impacts on future flood risk in the Pearl River Delta based on the MaxEnt model. *Frontiers in Earth Science*. 10. pp.1-18

□ 홈페이지

공공데이터포털. https://www.data.go.kr/

서울 열린데이터 광장. https://data.seoul.go.kr/

서정익. 전국매일신문. https://www.jeonmae.co.kr/news/articleView.html?idxno=896782

우림텍. http://www.woorimtech.com/

정세진. 머니투데이. https://news.mt.co.kr/mtview.php?no=2022080917011789586

환경 빅데이터 플랫폼. https://www.bigdata-environment.kr/

BURGESS&NIPLE. https://www.burgessniple.com/

DreamQuark. https://github.com/dreamquark-ai/

EPA. https://www.epa.gov/

esri. ArcGIS Pro. https://pro.arcgis.com/

FEMA. https://www.fema.gov/

Hoffman, Ken.. 2020. Random Forest Hyperparameters Explained. Medium. https://ken-hoffman.medium.com/random-forest-hyperparameters-explained-8081a93ce23d

Maklin, Cory.. 2022. Synthetic Minority Over-sampling TEchnique (SMOTE). Medium. https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c

Metro21: Smart Cities Institute. https://www.cmu.edu/metro21/index.html

Natural Resources Canada. https://cwfis.cfs.nrcan.gc.ca/home

Riskscape. https://riskscape.org.nz/

Scikit-learn. https://scikit-learn.org/

TUFLOW. https://www.tuflow.com/

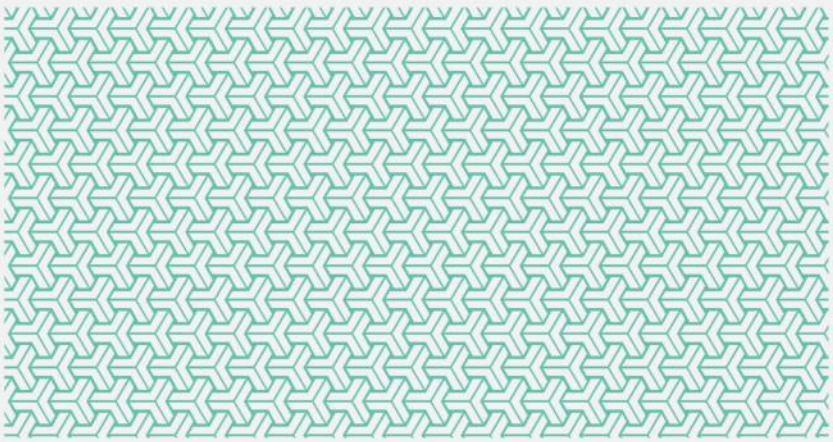
U.S. Army Corps of Engineers Hydrologic Engineering Center. https://www.hec.usace.army.mil/

□ 법령

건축물관리법, 법률 제19367호, 2023. 4. 18., 일부개정
건축물관리법 시행령. 대통령령 제33632호, 2023. 7. 11., 일부개정
소방시설 설치 및 관리에 관한 법률. 법률 제19160호. 2023. 1. 3.. 일부개정
시설물의 안전 및 유지관리에 관한 특별법. 법률 제17946호. 2021. 3. 16.. 일부개정

Summary

A Study of Developing a Big-Data-Based Building Fire and Flood Risk Analysis Model



Cho, Young Jin
Heo, Han Kyul
Song, Yu Mi
Hyeon, Tae Hwan

1. Introduction

Building fires and floods are threats to public safety, and they are projected to worsen due to climate change. Building fires and floods are the most costly of all disasters and accidents that damage buildings and are a major threat to public safety. In particular, the fire at the Coupang distribution center in Icheon in 2021 and the flooding of the Gangnam Station area and the deaths of residents in the semiunderground dwellings in the summer of 2022 have raised the public's demand for secure buildings that are safe from fire and flood.

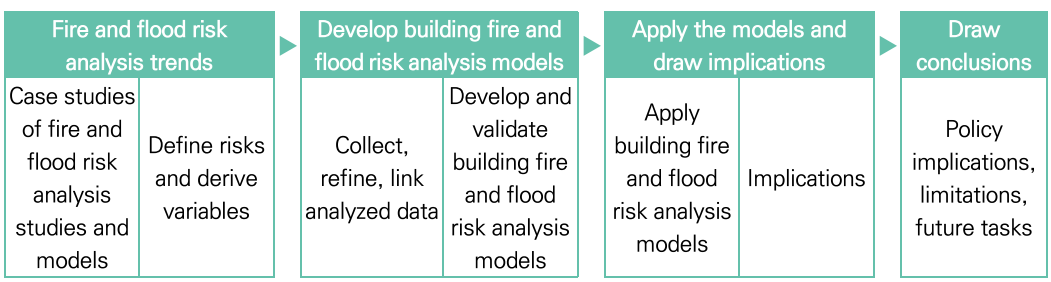
Analyzing building fire and flood risk is important to ensure building safety. Building characteristics and location influence the extent of damage during a fire or flood event, and insights obtained from them can be used to analyze the building's fire and flood risk. For example, buildings with a large floor area are relatively more prone to fire, and buildings located in low-lying areas are more likely to suffer flood damage. Researchers have been analyzing fire and flood risk in buildings based on the characteristics of the disasters damaging buildings, and the latest technological advancement is a turning point for this research.

The accuracy of risk analysis is improving with the development of the Fourth Industrial Revolution technologies such as machine learning, deep learning, IoT, and ICT. Machine learning algorithms such as decision trees, random forests, and support vector machines have improved risk analysis, and deep learning technologies such as the RNN (recurrent neural network), LSTM (long short-term memory), and CNN (convolutional neural network) have dramatically increased analysis accuracy. Advancements in IoT and ICT have contributed to obtaining data for analysis and are leading to problem solving using big data.

To respond to building fire and flood damage, it is necessary to identify buildings at high risk of fire and flood. It is also necessary to determine the factors that contribute to elevated risk and to practice building safety management with a particular focus on them. Building inspections are already required under the Building Management Act, the Act on the Installation and Management of Firefighting Systems, and the Special Act on the Safety and Maintenance of Facilities. The results of this study will contribute to developing a more reasonable framework for regular building inspections.

Against this backdrop, this study has four main objectives: 1) develop a building-level flood and fire risk analysis model, 2) demonstrate data linkage methods and establish a foundation for future data linkage, 3) identify buildings that require flood and fire preparedness and verify the accuracy of the model, and 4) propose a building management plan to improve building fire and flood safety.

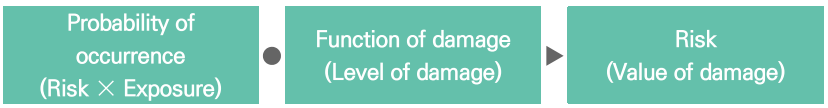
To achieve these objectives, a four-step research process was established. 1) We surveyed fire and flood risk studies and analysis models to define the building fire and flood risks to be addressed in this study and derived the variables to be used in the analysis. 2) We developed and validated building fire and flood risk analysis models, respectively. Machine learning (deep learning) algorithms were applied to the models, which involved the process of collecting, refining, and linking analyzed data. 3) We applied the building fire and flood risk analysis models to Gwanak District and put them together to analyze building fire and flood risk. 4) Based on the overall findings of the study, we drew policy implications, identified limitations of the study, and proposed future research tasks.



Flow of the study

2. Fire and flood risk analysis trends

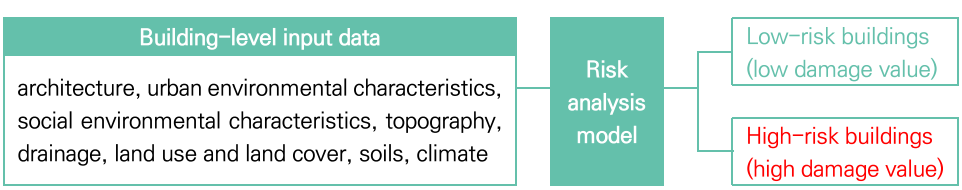
To analyze building fire and flood risks, we surveyed and analyzed previous studies and analytical models on fire and flood risk analysis. In general, risk analysis is based on the probability of occurrence and functions of damage. We found that the most common way of analyzing the final risk is to calculate the value of the damage. This suggests that while previous studies focused on the occurrence of fires and floods, risk analysis in practice aims to estimate the value of the damage to understand the actual damage better.



Concept of risk estimation

Studies of fire prediction models have been primarily based on statistical methodologies. They allow for analyzing the intensity and value of damage but are considered to have challenges in predicting the occurrence of fire itself. This is because a fire seldom starts spontaneously and is often caused by human error or arson; hence, environmental characteristics and location have a huge impact on fire occurrence. Risk analysis studies use a statistical fire occurrence model and combine it with a function of damage to estimate risks. The risk analysis of buildings seems to follow a similar methodology, which applies to both area- and building-level analyses.

Flood-related studies and analytical models are primarily based on mathematical and hydrological models to predict the probability and intensity of floods. SWMM and SWAT are some of the most commonly used models, which are widely applied to develop local flood countermeasures. Flood analysis models simulate the occurrence of floods in various situations in consideration of changes in precipitation and terrain characteristics, enabling the prediction of the probability and intensity of floods. In addition, models have been developed to perform risk analysis of flood events by combining them with functions of damage to estimate the value of damage. Hazus-MH, HEC-FIA, etc., are used as typical damage-based flood risk analysis models.



Concept of the risk analysis model to be developed

Based on the previous studies and model analysis, we established the framework of the building fire and flood risk analysis models. We designed both fire and flood risk analysis models to estimate the value of damage and categorize buildings into low-risk buildings with low damage value and high-risk

buildings with high damage value. For the probability of occurrence and functions of damage, we used a machine learning (deep learning) algorithm based on statistical models. In addition, we selected the variables for fire and flood risk analysis based on the findings from previous studies and model analysis.

Clean up catastrophe-catastrophe analysis variables

Risk	Input Variables
Fire	Variables related to urban environmental characteristics, variables related to social environmental characteristics, weather characteristics, and architecture
Flood	Topography-related variables, drainage-related variables, land use and land cover-related variables, Saturn-related variables, weather-related variables, and building-related variables.

3. Developing the building fire and flood risk analysis models

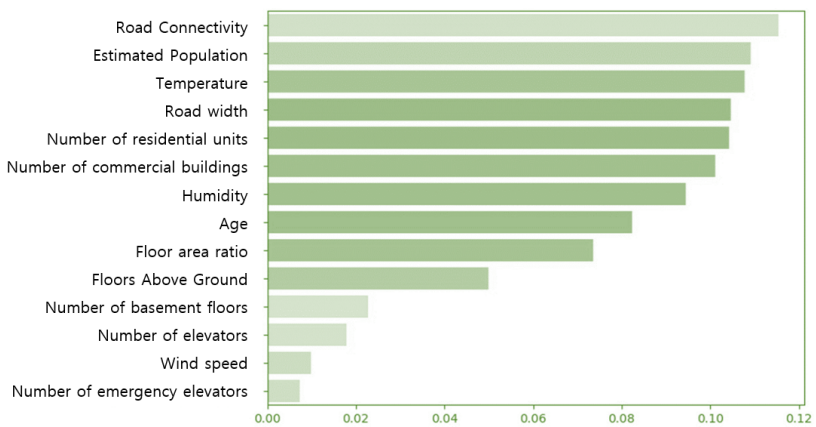
The development of the building fire and flood risk analysis models was divided into 1) data construction and 2) model development and validation phases. 1) The data linkage process involved linking building fire and flood damage occurrence and damage cost data with building register data, and producing and linking additional independent variables. 2) The model development and validation process involved selecting an algorithm for the learning of the analyzed dataset, training the models based on the selected algorithm, and validating and testing the models through a feedback process. The building fire and flood risk analysis models were developed independently.

To develop the building fire risk analysis model, we used a total of 28,266 cases of building fire damage data in Seoul from 2017 to 2021. Buildings with more than KRW 200,000 worth of damage were classified as high-risk buildings, and those with less than KRW 200,000 were classified as low-risk buildings. We applied various machine learning algorithms including the random forest, logistic regression, LightGBM, and XGBoost, and we selected the random forest that showed the highest accuracy.

Data building		Model development and validation	
Link building fire and flood damage occurrence data with building register data	Produce input variables for analysis and link with the building register	Select a machine learning (deep learning) algorithm and conduct the leaning of the database	Validate and test the models and develop the final risk analysis models through feedback

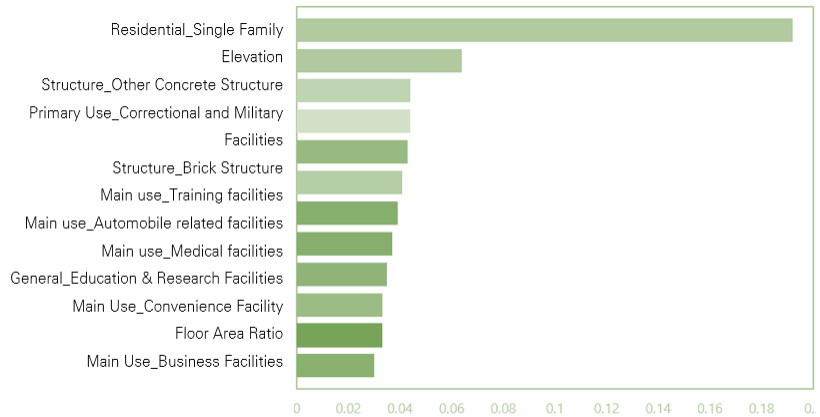
Model development process

The key influencing factors of the building fire risk analysis included road connectivity, estimated population, temperature, road width, number of residential buildings, and number of commercial buildings. The final accuracy of the building fire risk analysis model was 78%. The F1 score was 0.75, and the AUC value, which represents the area under the ROC curve, was 0.89, indicating a quality risk analysis model.



Importance of variables of the fire risk analysis model

To develop the building flood risk analysis model, we used a total of 27,438 cases of building flood damage data in Seoul from 2016 to 2022. Buildings with more than KRW 3,000,000 worth of damage were classified as high-risk buildings, and those with less than KRW 3,000,000 were classified as low-risk buildings. For the building flood risk analysis model, we used the TabNet classifier. The key influencing factors of the building flood risk analysis included the use and structure of buildings, height, and floor-area ratio of buildings. The final accuracy of the building flood risk analysis model was very high at 88%.



Importance of variables of the flood risk analysis model

4. Application of the building fire and flood risk analysis models

We applied the building fire and flood risk analysis models to Gwanak District, Seoul. To develop the model, we used buildings that had already experienced fire or flood damage, and we applied the developed models to all buildings in Gwanak District to analyze their fire and flood risks. There are a total of 32,079 buildings in Gwanak District, and we developed input variables for the analysis and linked the data with all building points.

[Map of buildings with high fire risk]
red: high risk of fire

[Map of buildings with high flood risk]
blue: high risk of flood

Maps of building flood and fire risk (results of individual analyses)

Feeding the data into the models resulted in building fire and flood risk maps. The analysis showed that there are 8,405 buildings (~31%) at high risk of fire damage, and 446 buildings (~1.5%) at high risk of flood damage.

Yellow: low risk of fire and flood, asterisk: high risk of fire and flood,
red: high risk of fire, blue: high risk of flood

Map of building flood and fire risk (results of integrated analysis)

To obtain comprehensive insight into building fire and flood risk, we superimposed the individual predictions to derive four types of buildings depending on their fire and flood risk levels. Of the 32,079 buildings in Gwanak District, 23,276 buildings were identified at low fire risk and low flood risk, 8,357 buildings at high fire risk, 398 buildings at high flood risk, and 48 buildings at high fire risk and high flood risk.

Categorization of building fire and flood risk

	Low risk of fire	High risk of fire
Low risk of flood	23,276 buildings (72.56%)	8,357 buildings (26.05%)
High risk of flood	398 buildings (1.24%)	48 buildings (0.15%)

5. Conclusions

This study has four policy implications. First, we revealed the need to take advantage of the Fourth Industrial Revolution technologies to respond to disasters and accidents that affect buildings. Using these technologies, for example, machine learning, deep learning, and big data, enables one to analyze fire and flood risk in buildings and ensure higher accuracy than the existing methods. It is believed that embracing these technologies and actively using them will help protect buildings from disasters and accidents.

Second, we proposed measures to reduce building fire and flood risk. Based on the development of the building fire and flood risk analysis models we developed, we found factors that lead to increased risk. For building fire risk, key factors include road connectivity, estimated population, temperature, road width, and number of residential buildings, and for building flood risk, key factors include use and structure of buildings, height, and floor-area ratio. It is expected that improving building management policies with a focus on these factors will help reduce the damage caused by fire and flood.

Third, we proposed the need for integrated disaster response at the building level. Fire and flood are disasters that cause buildings to suffer the most severe damage. To respond to these issues, the existing models and studies have performed risk analysis for individual disasters and hazards. However, in this study, we developed models that enable an integrated response to both fire and flood. While fires and floods may have totally different causes and require totally different responses, it may be more efficient to view them together in terms of identifying response priorities at the building level. Successful building disaster and accident prevention requires developing an integrated building safety system.

Fourth, we identified the need for institutional improvement to prioritize the management of buildings prone to disasters and accidents. We found that smaller buildings are more prone to fire and flood risk compared to larger buildings that are subject to statutory inspections under relevant laws. In other words, the law does not provide for inspections for buildings that are prone to fire and flood; hence, there is a need to improve the system. Currently, Article 15 of the Building Management Act (Inspection of Small and Old Buildings) stipulates the inspection of small and old buildings, but considerations should be given to tightening the regulation. In addition, further research is warranted to utilize public data better and apply new algorithms to improve the accuracy of risk analysis and expand the scope to disasters and accidents other than fire and flood.

Keywords :
Big-Data, Building Fire risk, Building Flood risk, Machine Learning, Deep Learning