

# 기계학습모형을 활용한 인천광역시 격자 기반 홍수 위험도 예측\*

## Flood Risk Prediction Using the Grid-Based Machine Learning Models in Incheon

정민주\*\* · 김기찬\*\*\* · 이창길\*\*\*\* · 김현우\*\*\*\*\*

Minju Jeong\*\* · Kichan Kim\*\*\* · Chang Kil Lee\*\*\*\* · Hyun Woo Kim\*\*\*\*\*

### Abstract

This study aims to evaluate flood risk in Incheon using a machine learning-based prediction model. The flood risk prediction model was trained using geographic and climatic factors from 2016 to 2021, and its consistency and applicability were validated by predicting flood risk areas for 2021. Model performance evaluation revealed that the Random Forest Regressor showed the best prediction performance, effectively explaining data variability and minimizing errors. Furthermore, maximum precipitation was identified as the most influential variable affecting flood risk, with a sharp increase in flood risk observed when precipitation exceeded 450mm. Based on this, the 2021 Flood Risk Index was predicted, resulting in lower errors compared to the model development phase and high explanatory power ( $R^2 = 0.9619$ ), demonstrating high model reliability. The main flood risk areas identified were agricultural lands near the Ara Waterway in Geyang-gu and Namdong National Industrial Complex in Gojan-dong, Namdong-gu, characterized by gentle slopes and expanded impervious surface areas, respectively. The results suggest that machine learning-based prediction models can effectively assess flood risks caused by urbanization and climate change, providing a basis for data-driven, region-specific flood mitigation strategies in urban planning.

**Keywords:** Climate Change, Urbanization, Random Forest Regressor, Flood Mitigation, Disaster Prevention

## 1. 서 론

### 1.1 연구의 배경 및 목적

2022년 8월 중부권을 중심으로 발생한 집중호우는 일부 지역에서 평년 대비 2배를 넘는 강수량을 기록하는 등 관측 이후 최악의 집중호우로 평가되었다.

이로 인해 총 19명(사망 17, 실종 2)의 인명피해와 3,154억 원의 재산피해, 그리고 409.7ha의 농경지 유실·매몰, 가축 33,910마리 폐사 등 큰 피해가 발생하였다(기상청, 2023). 특히, 도심 지역에서 집중된 강수량으로 인해 투수층이 부족한 시가지 지역이 상대적으로 큰 피해를 입은 것으로 나타났다. 이처럼

\*이 논문은 2022년 대한민국 교육부와 한국연구재단의 일반공동연구지원사업의 지원을 받아 수행된 연구임(NRF-2022S1A5A2A03055420).

\*\*인천대학교 도시계획·정책학과 박사과정(주저자: ming@inu.ac.kr)

\*\*\*인천대학교 도시계획·정책학과 박사과정

\*\*\*\*인천대학교 도시행정학과 교수

\*\*\*\*\*인천대학교 도시행정학과 부교수(교신저자: kimhw@inu.ac.kr)

최근 기후변화로 인해 특정 지역에 단기간동안 많은 강우가 집중되는 국지성 집중호우가 증가하고 있는 현황이나, 집중호우 발생 및 피해 예상 지역 예측과 대비에 어려움을 겪고 있다(이규민 외, 2020; 배재현·김진수, 2022). 특히 도심지역의 경우 대규모 기반시설이 집중되어 있는 등 도시의 성장으로 인한 불투수층의 증가가 침수 피해를 심화시키고 있는 추세이다. 인천시 또한 2022년 집중호우 당시 1시간 최대강수량 84.8mm, 최대 3위를 기록하는 등 홍수와 같은 재해에 취약해지고 있는 실정이다.

인천시를 포함한 지방정부는 「국토의 계획 및 이용에 관한 법률」에 따라, 각 지역의 재해 취약성을 평가하여 도시계획 기초자료로 활용하고 있다. 또한 인천시는 2018년 시민안전본부 주도 하에 재난 Quick Risk Estimation(QRE) 분석을 실시하는 등 도시 계획을 통해 재난에 대한 대응 및 예방을 목적으로 재난 위험을 평가하고 있다(인천연구원, 2021). 그러나, 이러한 평가 방식은 평가 지표 선정의 문제, 가중치 등 지표 산정 단계에서 평가자의 주관이 들어간다는 한계가 존재한다(김지숙 외, 2014). 또한, 행정단위로의 적용을 위해 공간단위를 행정구역으로 한정하여 공간단위에 따른 분석 결과의 신뢰성 확보와 적용이 한정적이다. 따라서, 도시 홍수에 대응하기 위해 홍수 위험을 미시적 단위에서 객관적으로 평가할 수 있는 평가 방법이 필요한 시점이다.

이에 본 연구는 인천광역시를 대상으로 지리적 및 기후적 요인을 고려한 홍수 위험을 평가하고자 하였으며, 2021년에 이를 적용하여 홍수 위험 지역을 예측하고자 한다. 나아가 기후변화로 인한 불규칙한 강우 패턴에 대응하기 위해 본 연구에서는 홍수 위험 평가 체계를 제안하고자 하며, 홍수에 탄력적으로 대응하기 위해 홍수 위험 지역을 도출하여 홍수 예방 정책 우선 시행 지역을 제시하고자 하였다. 이를 통해 본 연구에서는 홍수를 예방하기 위한 홍수 평가 및 장기적인 재난 대응 정책을 위한 중요한 가

이드 라인 및 기초적 자료로 활용 가능할 것이라 기대한다.

## 1.2 연구의 범위 및 방법

본 연구는 기후변화 및 도시화로 인해 홍수 위험에 노출된 인천광역시 육지부를 공간적 범위로 설정하였으며 기상·물리적 요인의 변화를 반영하기 위해 2016년부터 2021년까지를 시간적 범위로 설정하였다. 인천광역시는 UN Office for Disaster Risk Reduction 롤모델 도시 인증을 신청하기 위한 재난 QRE에서 풍수해 발생 가능성 및 심각도가 높다고 평가되는 등 홍수 위험에서 안전하지 못한 지역이다. 이에 본 연구에서는 인천광역시의 물리적 특성 및 홍수 발생 이력을 바탕으로 홍수 위험을 평가하고자 한다. 또한, 미시적 관점에서 홍수 위험 지역을 파악하고, 우선 시행 지역을 설정하여 홍수 대응 정책을 마련하고자 하였다.

## 2. 선행연구 검토

### 2.1 도시 홍수의 정의 및 위험성

홍수는 사전적으로 태풍, 호우 등 강우로 인해 발생하는 재난·피해로 그 중 도시 홍수는 인적, 물적 피해를 유발하는 자연재해이다(조형진, 2017). 도시 홍수의 발생 원인은 일반적으로 Source-Pathway-Receptor 모델 개념인 기상학, 공간, 시설물로 설명된다. 즉, 도시 홍수는 홍수 발생의 시작점인 기상학적 요인에 해당하는 집중호우의 증가 등 기후변화 영향으로 인해 그 규모와 빈도가 증가하는 양상을 보여오고 있다. 특히, 기후변화 영향과 관계가 높아 기후변화로 인해 홍수 위험에 노출된 세계 인구 비율이 2000년 대비 2015년 약 24% 증가하는 등 도시는 홍수의 강도와 빈도의 증가로 인한 홍수 위험에 직면하고 있다(Tellman et al., 2021). 도시에서 발생하는 극심한 홍수의 증가는 도시민의 인명과 재

산에 막대한 피해를 입히는 등 물리적 측면뿐만 아닌 주택과 도시 기반 시설을 파괴로 인한 공동체의 와해 및 일상생활 등 사회적 측면에 직접적인 영향을 미치고 있다(Ermagun et al., 2024; Wing et al., 2022). 나아가 도시 전반에 물리적·사회적 피해를 입히는 홍수는 재난 시점에 발생하는 피해에 국한되지 않으며, 장기적인 측면에서 미래 도시를 개발하는 데 있어 직·간접적인 위험을 가할 가능성 또한 존재한다(Wang et al., 2022). 특히 Pan et al.(2020)는 강수량 증가로 인한 홍수 위험의 증가는 새로운 도시 개발의 위치와 구성을 제한할 수 있음을 강조하며 홍수 위험은 도시 개발 결정에 영향을 미칠 수 있음을 시사하였다.

홍수의 강도와 빈도를 증가시키는 또 다른 요인 중 하나로는 도시의 불투수 표면적을 증가시키는 도시화가 있다. 특정 지역에 도시화가 진행됨에 따라 녹지와 같이 우수를 흡수할 수 있는 투수 표면적이 감소하여 지면에 우수가 흡수될 수 있는 수문학적 대응 시간이 단축되며 이는 물리적 측면에서의 홍수 위험을 증가시킨다(Booth, 1991; 이정민 외, 2010; Feng et al., 2021). 도시화로 인한 도시의 성장 및 확장은 도시의 회복력과 지속가능성을 변화시키며, 이는 홍수 피해와 그 심각성에 직접적인 영향을 미친다(Huong and Pathirana, 2013). 도시 홍수는 도시화와 기후변화의 상호작용으로 인해 심각한 도시 문제로 대두되고 있으며, 도시에서 발생하는 홍수는 도시의 재난 대응 능력을 초과시켜 도시의 복원력을 저하시키기에 다수의 연구에서 장기적인 관점에서 안전한 도시를 위해서는 도시 홍수를 효율적으로 관리하기 위한 방안이 필요하다 언급하였다(Cai et al., 2021; Rosenzweig et al., 2018; Theodosopoulou et al., 2022). 또한, 도시 홍수를 분석한 Rentschler et al.(2023)의 연구에서는 토지이용 및 지형의 변화를 홍수 위험 요소로 언급하였으며 도시의 변화로 인한 홍수가 인명피해와 경제적 손실을 초래한다고

언급하는 등 물리적 측면에서 홍수 위험을 평가할 필요가 있음을 강조하였다.

도시화와 기후변화로 인한 홍수 위험은 특히 아시아 국가들에서 증가하고 있는데, Global Flood Database에 따르면 2000년부터 2018년 사이에 홍수 노출 위험이 증가한 국가 대부분이 아시아에 집중되어 있다고 분석되었다(Tellman et al., 2021; Global Flood Database, 2024). Rentschler et al. (2023)는 도시화 이래 홍수 위험 지역의 증가는 빠르게 진행되고 있으며, 동아시아에서는 고위험 거주지가 홍수 안전 지역보다 60% 더 빠르게 확장되는 등 국가에 따라 홍수 노출에서 체계적인 차이를 보이고 있음을 언급하였다. 한국 또한 최근 10년(2010~2019) 발생한 재해 피해액 중 홍수가 차지하는 비율이 50%에 육박할 만큼 홍수 위험에 지속적으로 노출되어 있기에 국내 홍수 위험을 평가할 필요성이 대두되는 시점이다(김대호 외, 2020). 특히, 한국은 기후변화에 의한 강우량 증가에 직·간접적 영향권으로 기후변화로 인해 지난 106년 동안 강수량은 10년당 16.3mm씩 증가하였다. 그러나, 강우일수에는 변화가 없었으며 강한 강수의 강도와 일수는 증가하고 있다고 보고되고 있는데, 이는 집중호우 발생이 증가함에 따라 도시 홍수 위험성이 증가하고 있는 것을 의미한다(국립기상과학원, 2018). 2022년 서울을 비롯한 수도권에 100mm가 넘는 강수량이 기록되기도 하는 등 수도권에서의 홍수 위험도가 증가하고 있는 추세이다(김선태 외, 2023). 특히 인천광역시의 경우 2018년 인천광역시 시민안전본부에서 수행한 재난 QRE 분석에서 풍수해 발생 가능성이 높다고 평가되는 등 해안과 인접도가 큰 해안도시로 기후변화에 따른 홍수 취약성이 높은 지역이라 평가되고 있다. 그러나 대부분의 연구에서 인천시 행정구를 대상으로 홍수 위험을 파악하는 등 미시적인 관점에서 홍수 위험성을 평가한 연구는 미흡한 현황이다(조형진·김경배, 2015; 조형진, 2017;

왕원준 외, 2021). 이에 본 연구에서는 보다 작은 단위인 그리드를 활용하여 인천광역시의 지속적인 홍수 위험을 정량적으로 평가한 후 기후변화와 도시화에 따른 홍수 위험에 대응하기 위한 홍수 위험 지역을 분류하여 제시하였다는 점에서 그 차별성이 존재한다.

## 2.2 홍수 위험 지수 관련 선행연구

홍수 위험은 일반적으로 발생 패턴의 불확실성이 존재하지만 예측할 수 없는 존재는 아니며, 그 영향을 완화하고 효율적으로 대응하기 위해 홍수 위험을 파악하고 하는 연구가 다수 진행되어 왔다. 홍수의 위험을 평가한 연구는 초기 재해의 위험 및 취약성이라는 개념적인 부분을 다룬 연구가 주로 이루어졌으나, 최근 도시 홍수의 특징에 따른 위험을 평가하는 실증적인 연구가 활발히 이루어지고 있다. 물리적 측면에서 홍수 위험 지역을 평가한 국외 연구인 Sanders et al.(2020)는 홍수 인식을 반영한 홍수 위험 분류 지도를 제작하기 위해 홍수 취약성이 아닌 홍수 위험의 물리적 측면에 집중하였으며, 지형 정보·하천 유입·강수량 등을 활용하여 도시 홍수 위험 지수를 도출하였다. 또한, 홍수 위험을 평가할 수 있는 데이터가 부족한 국가의 홍수 위험을 평가하기 위해 경사, 강으로부터의 거리, 토지 이용 피복도, 토양 유형, 배수 밀도와 같은 지형 데이터 및 강수량, 5년마다의 강우강도, 지속시간, 빈도 등과 같은 기상 데이터를 활용하여 홍수 위험 지수를 구축한 연구가 진행되기도 하였다(Armenakis et al., 2017; Ramkar and Yadav, 2021; Acosta-Quesada and Quesada-Román, 2025). 홍수 위험 지역을 평가한 국내 연구인 박지수·임철희(2023)의 연구에서는 홍수 위험 지수를 구축하여 지역의 홍수 위험을 파악하기 위해 기후 리스크 개념을 활용하여 서울시 홍수 재해를 대상으로 5일 최대 강수량, 강수강도, 호우 일수를 통해 기후 불평등 및 홍수 리스크를 평

가하였다. 김대호 외(2020)는 홍수 위해, 노출, 대응 능력을 바탕으로 전국 단위의 홍수위험도를 평가하기 위해 호우일수, 녹지 면적, 유역 경사 등을 활용하여 전국 11개 중권역의 홍수 위험을 평가하였다. 또한 인천광역시를 대상으로 홍수 위험 지수를 구축한 연구들 또한 존재하였는데, 조형진·김경배(2015)는 인천광역시를 대상으로 물리적 취약성에 해당하는 하천변 1km 내 계획 홍수위 보다 낮은 지역의 면적비율 등을 홍수 위험 지수로 산정하여 홍수 위험의 사회적 취약성에 대해 파악하는 등 물리적인 특성이 홍수 위험을 나타낼 수 있음을 언급하였다. 인천광역시의 홍수위험도 평가를 위해 QRE 위험 평가를 활용한 연구인 왕원준 외(2021)의 연구에서는 확률강수량, 최대강수량, 도로면적 등을 활용하여 시군구별 홍수 위험도를 도출하였다. 나아가 홍수 위험을 예측하기 위해 머신러닝 기반 홍수 위험 예측 모델을 활용한 김동현 외(2022)는 국가하천인 남양주시 유역을 대상으로 심층신경망 모델을 활용하여 홍수 피해 예·경보 기법을 제시하였다.

대부분의 연구에서는 지역의 고도 또는 기울기 등을 활용하여 자연환경으로 인한 홍수 위험을 평가한 연구가 대부분이며 도시지역에서의 홍수 위험을 고려한 연구는 부족하였으며 기상학적 요인 또한 강수량만을 고려한 연구가 대부분이기에 관측 단위까지 시군구로 한정되어 있다는 한계가 존재한다. 또한, 국가적 차원에서 지역별 홍수 위험을 관리하기 위한 전방위 기초자료인 홍수위험지도 제작·활용하고 있으나 활용 번위가 지자체에 한정되어 있어 접근성이 미흡하다는 한계가 존재한다(박문현, 2021). 머신러닝 등 예측모델을 기반으로 홍수 위험도 및 위험 지역을 예측한 기존 연구의 경우 국가 하천을 대상으로 하천홍수의 위험만을 예측하고 있어 도시홍수를 예측한 연구는 미흡한 실정이다. 이에 본 연구에서는 위성 데이터를 기반 데이터를 활용하여 보다 미시적인 단위에서의 홍수 위험을 평가하고자 하였

으며, 정규시가화지수 및 도로까지의 거리 등 도시화를 나타낼 수 있는 변수를 활용하여 도시지역에서의 홍수 위험을 평가하였다는 점에서 그 차이점이 존재하며 홍수 대응을 위해 도시지역에서 홍수 위험 지역을 예측하여 제시하였다는 것에 그 의의가 있다.

### 3. 연구방법론

#### 3.1 연구대상지

본 연구에서는 타 시도에 비해 해안과 인접한 도시인 인천광역시를 대상으로 홍수 위험 지수를 구축하였다. 인천광역시는 산업시설 등이 해안에 밀집되어 있으며, 육지부에 인천신항을 비롯한 북항, 내항, 남항, 배후단지 등이 위치한 지역이다. 특히, 역사적으로 개항 이후 항구를 중심으로 인구와 기반 시설이 밀집되어 개발된 해안지역이기에 물리적으로 홍수 위험에 노출된 지역이다. 이러한 인천시는 특히 도시화와 기후변화로 인해 홍수 위험이 증가할 가능성이 높은 지역인 인천시는 하천 제방을 보장할 수 있는 최고 수위를 의미하는 계획홍수위가 상향 조정되는 등 홍수 위험에 노출되어 있는 지역이다(환경부, 2024). 또한, 인천시는 증가하는 홍수 위험에 대응하기 위해 2024년 부평 굴포천 홍수 방어벽 높이를 당초 계획인 1.2m에서 2.1m로 보다 상향하여 건설하기도 하였다(이재희, 2024).

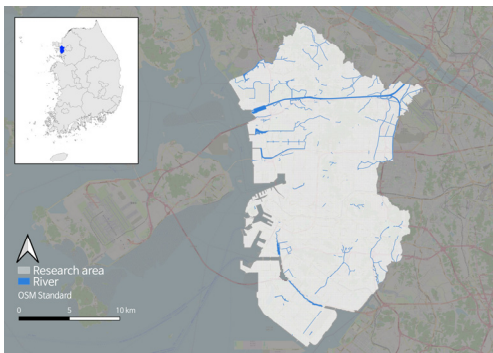


Fig. 1. Study Area

이에 본 연구에서는 홍수 위험성이 증가하고 있는 인천광역시 육지부를 대상으로 기후변화 및 도시화에 따른 홍수 위험을 평가하고 미시적 관점에서 이를 예방하기 위한 도시 정책의 우선순위 지역을 제시하기 위해 100m × 100m 그리드 단위를 분석 단위로 구성하였다(Fig. 1 참고).

#### 3.2 PyCaret

PyCaret은 머신러닝 모델에서 활용되는 데이터 전처리, 기능·모델 선택, 하이파라미터 튜닝 등 모델링 프로세스를 자동화한 Python 환경 기반의 오픈소스 머신러닝 라이브러리이다(Ali, 2020). 이를 통해 자동화 모듈 내에 구성되어 있는 다수의 머신러닝 알고리즘을 효율적으로 학습하고 최적화시킬 수 있는 모델로 MSE(평균 제곱 오차), RMSE(평균 제곱근 오차), MAE(평균 절대 오차) 및  $R^2$  등을 통해 학습 모델 간 예측 성능을 효율적으로 비교할 수 있다는 장점이 존재한다(Kim et al., 2023; Dereci and Tuzkaya, 2024).

본 연구에서는 인천광역시를 대상으로 구축된 100m × 100m 단위의 홍수 위험 지수를 바탕으로 홍수 위험을 예측할 수 있는 모델을 구축하고, 홍수 위험 지수의 변화를 예측하여 홍수 위험 지역을 도출하고자 PyCaret을 활용하였다. 예측 모델을 구축하는 단계에서 PyCaret을 사용하여 최적화된 예측 모델을 훈련하기 위해 학습 모델 간의 성능을 비교 및 평가하여 데이터를 가장 효과적으로 예측할 수 있는 모델을 선택하였다. 이 과정에서 최적화된 모델의 예측 파이프라인을 FRI 예측 요인에 적용함으로써 보다 체계적이고 재현 가능한 분석 체계를 구축하였다.

#### 3.3 홍수 위험 지수(Flood Risk Index: FRI)

본 연구에서는 해안 도시인 인천광역시를 대상으로 기후변화 및 도시화로 인한 홍수 위험을 평가하고



홍수 위험 지역을 파악하기 위해 기상·지리적 측면에서의 홍수 위험 지수를 구축하였다. 이에 홍수 위험 지수를 홍수 피해 강도를 증가시키는 요인을 중심으로 홍수 발생 시 피해를 심화시킬 수 있는 취약한 지역을 정량적으로 평가할 수 있는 지표라 정의하였다. 홍수 위험 지수에는 선행연구 검토를 통해 홍수 위험에 영향을 미친다고 언급된 요인인 경사, NDVI 등 지리적 요인(Sanders et al., 2020; 박지수·임철희, 2023; 김대호 외, 2020) 및 누적 강수량, 온도 등 기상학적 요인(Armenakis et al., 2017; Acosta-Quesada and Quesada-Román, 2025)에 해당하는 자료를 사용하였다. 또한, 도시 홍수 위험성 분석을 위해 기존 연구에서 고려되지 않은 정규시가화지수 등 도시 지역의 특성을 반영할 수 있는 변수를 추가적으로 활용하였다. 이를 위해 대중적 접근이 가능한 시스템을 이용하여 장기간의 연속적 변화 추적이 가능한 Google Earth Engine(GEE)에서 제공하는 위성 데이터를 활용·구득 하였으며 이를 100m × 100m 단위의 그리드로 가공하였으며 사용한 자료는 아래 Table 1과 같다. 본 연구에 활용된 기후적 요인의 경우 월별 단위로 제공되던 4km × 4km 데이터를 장기적 경향성을 파악하기 위해 연간 월평균 100m × 100m로 정량화 하였으며, 지리적 요인은 분석의 일관성을 위해 30m × 30m 또는 90m × 90m 해상도의 데이터를 100m × 100m 단위로 정량화 하였다.

홍수 위험 지역 예측을 위해 행정안전부에서 제공하는 생활안전지도 중 과거 홍수 발생 지역을 나타낼 수 있는 연도별 침수흔적도·도시침수지도 및 100년 빈도 하천범람도(국가·지방)를 예측 대상으로 활용하여 2016년부터 2021년까지의 홍수 위험 지수를 도출하였다. 홍수 위험 지수는 지방 및 국가 하천의 범람 가능성과 침수 심도에 따라 1~5점 척도로 구분된 도시 침수 지도 및 연도별 침수 흔적도를 각각 0~1 범위로 정규화한 후 합산하여 산출하였다. 나아가 해당 요인들을 활용하여 계산된 FRI를

Pycaret을 통해 구축된 홍수 위험 지수 예측 모델을 활용하여 예측된 FRI 값을 100m × 100m 그리드로 변환하여 홍수 위험을 평가하였다(Fig. 2 참고).

### 3.4 학습데이터 및 데이터 전처리

본 연구에서는 홍수 위험 지수를 예측하기 위해 위성 데이터를 기반으로 학습데이터를 구성하였다. 기상학적 요인의 경우 일관된 분석을 위해 균일한 격자 해상도로 구성된 GEE에서 제공하고 있는 University of California Merced의 TerraClimate dataset을 활용하였다. TerraClimate dataset는 전 세계 지표면의 기후 및 기후수분 균형 정보를 제공하는 데이터셋으로, 기후 보조 조건을 사용하여 4km × 4km (1/24도) 격자로 이루어진 위성 사진 기반 데이터이다(Abatzoglou et al., 2018). 기후적 요인의 장기적인 변동성을 포함하여 홍수 위험 평가의 신뢰성을 확보하기 위해 데이터 수집 기간은 2016년 1월부터 2021년 12월까지로 설정하였으며, 연간 월평균 데이터를 활용하였다. 지리적 요인의 경우 연구대상의 도시화 정도를 나타낼 수 있는 NDVI, NDBI, NDWI는 Sentinel-2에서 촬영된 영상을 기반으로 계산되었으며, 도시 및 자연 환경의 공간적 특성을 보다 명확하게 반영하기 위해 식물의 생장 시기인 5월에 촬영된 영상을 활용하였다. 또한, 90m 단위의 수치표고모델을 활용하여 대상지의 지리적 특성을 반영할 수 있는 변수를 도출하였다. 홍수 영향 요인 및 과거 홍수 발생 자료의 경우 래스터 형식으로 제공되고 있으며, 각 데이터의 해상도가 서로 상이하기에 공간적 일관성을 확보하고 분석의 정확도를 높이기 위해 PySAL 라이브러리에서 제공하는 tobler 면적가중 보간을 활용하여 100m × 100m 해상도로 보간하여 학습 데이터를 구성하였다. 경계부의 경우 면적비를 고려한 보간 방식을 적용하여 데이터 간 왜곡을 최소화하였다. 나아가 결측치는 데이터 분석 및 모델 훈련에 부정적인 영향을 미칠 수 있으므로

**Table 1.** Physical Factors for Constructing Flood Risk Indicators

Classification	DataName	Description	Temporal	Spatial
Climate Factors	ro	Max Maximum runoff derived using a 1D soil moisture balance model	Monthly	4 km×4 km Grid
		Mean Mean runoff derived using a 1D soil moisture balance model		
	vpd	Max Maximum vapor pressure deficit		
		Mean Mean vapor pressure deficit		
	aet	Max Maximum actual evapotranspiration derived using a 1D soil moisture balance model		
		Mean Mean actual evapotranspiration derived using a 1D soil moisture balance model		
	tmmn	Max Maximum average temperature		
		Mean Mean average temperature		
	def	Max Climate water deficit derived using a 1D soil moisture balance model		
		Mean Mean climate water deficit derived using a 1D soil moisture balance model		
	soil	Max Maximum soil moisture derived using a 1D soil moisture balance model		
		Mean Mean soil moisture derived using a 1D soil moisture balance model		
	pet	Max Maximum potential evapotranspiration		
		Mean Mean potential evapotranspiration		
	pr	Max Maximum cumulative precipitation		
		Mean Mean cumulative precipitation		
	pdsi	Max Maximum Palmer Drought Severity Index		
		Mean Mean Palmer Drought Severity Index		
Geographical Factors	tmmx	Max Maximum temperature	Annual	90 m×90 m Grid
		Mean Mean temperature		
	Flow-Accumulation	Flow accumulation		
	S2-NDBI	Normalized Difference Built-up Index from Sentinel-2 satellite		
	S2-NDWI	Normalized Difference Water Index from Sentinel-2 satellite		
	S2-NDVI	Normalized Difference Vegetation Index from Sentinel-2 satellite		
	Flow-Pointer	Flow pointer		
	Flow-Direction	Flow direction		
	Slope	Slope degree		
	Aspect	Direction of the steepest slope		
	Mean-Curvature	Mean curvature		
	Distance to Road	Distance to the nearest road		Line

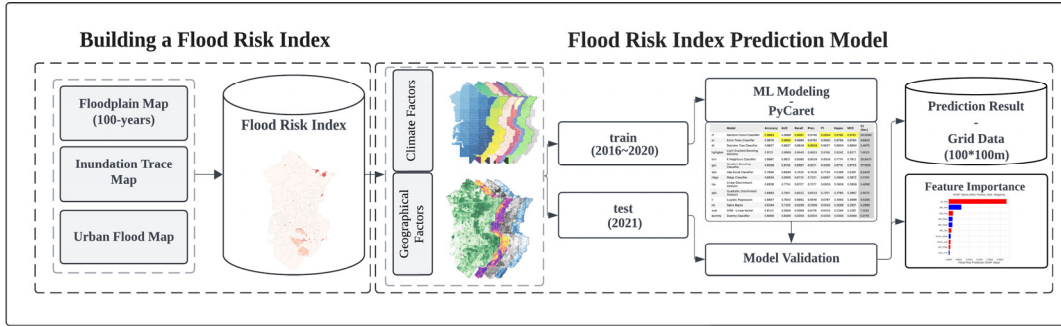


Fig. 2. Flow Chart

Table 2. PyCaret Algorithm Prediction Performance

Model	MAE	MSE	RMSE	R <sup>2</sup>	RMSLE
Random Forest Regressor	0.0024	0.0002	0.0147	0.8845	0.0123
Decision Tree Regressor	0.0009	0.0003	0.0169	0.8453	0.0143
Extra Trees Regressor	0.0058	0.0006	0.0239	0.6947	0.0202
Extreme Gradient Boosting	0.0088	0.0007	0.0262	0.6305	0.0226
Light Gradient Boosting Machine	0.0083	0.0009	0.0298	0.525	0.0255

본 연구에서는 과거 홍수 발생하지 않은 지역이 존재하여 그 값은 0으로 대체하여 진행하였으며 홍수 영향 요인에는 결측치가 무작위로 발생하는 데이터에 효과적으로 적용 가능한 평균 대체법을 활용하여 결측치를 보완하였다.

예측 모델 구축에는 2016년부터 2020년까지의 데이터를 활용하였으며 훈련데이터와 검증데이터의 분할은 7:3으로 진행하였다. 데이터 분할의 경우 K-Fold 교차검증을 통해 훈련 및 검증 과정을 거쳤으며, 검증 결과의 신뢰성을 높이기 위해 각 모델별로 10-fold 교차검증을 수행하였다.

## 4. 분석 결과

### 4.1 홍수 위험 지수 평가 모델

본 연구에서는 2016년부터 2020년까지의 지리적 요인과 기상학적 요인 및 과거 홍수 발생 데이터를 학습 데이터로 활용하였다. 이를 기반으로 홍수

위험 지수 예측 모델을 구축하기 위해 PyCaret을 통한 머신러닝 모델의 예측 성능을 평가하였으며, 평균 절대 오차(MAE), 평균 제곱 오차(MSE), 평균 제곱근 오차(RMSE), 결정 계수(R<sup>2</sup>), 로그 평균 제곱근 오차(RMSLE), 평균 절대 백분율 오차(MAPE)를 사용하였다. 모델 성능 비교를 통해 최적 모델을 도출하고자 하였으며, 모델간 성능을 비교한 결과는 Table 2와 같다.

종합적인 측면에서 모델 성능을 비교한 결과, Random Forest Regressor의 R<sup>2</sup>가 0.8845로 도출되었으며 이는 해당 모델이 데이터의 변동성을 비교적 가장 우수하게 설명하는 모델이라는 것을 의미한다. MSE, RMSE, RMSLE 또한 가장 낮은 값이 도출되어 해당 모델이 예측값과 실제값 간의 차이가 최소화되었으며 데이터 과소·과대 추정의 가능성이 매우 낮다는 것을 확인하였다. 반면, 모델의 예측값과 실제값 간의 차이의 절댓값 평균을 의미하는 MAE는 Decision Tree Regressor보다 다소 높게



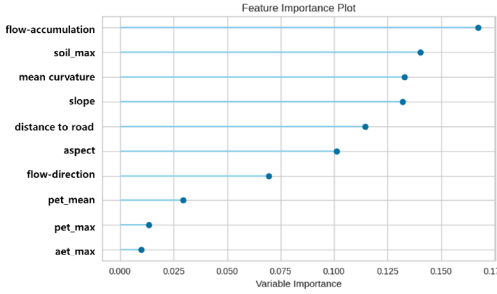


Fig. 3. Feature Importance

도출되었으나, 모델의 종합적인 예측의 정확성과 신뢰도를 고려하였을 때 다른 지표들의 성능이 전반적으로 우수하기에 Random Forest Regressor를 사용하는 것이 적절하다고 나타났다. 이러한 결과를 바탕으로, FRI를 예측하기 위한 모델 중 Random Forest Regressor가 대부분의 성능 지표에서 가장 우수한 결과가 도출되었기에 본 연구에 가장 적합한 모델로 판단하였다.

모델 훈련 과정에서 예측 모델 성능에 가장 큰 영향을 미친 변수는 누적 유량, 토양 수분, 평균곡률, 기울기로 예측 모델에서 자주 분할에 사용되어, 예측을 구분하는 데 중요한 역할을 했다는 것을 의미한다(Fig. 3 참고). 그러나, 이는 모델 훈련 과정에서 자주 사용된 특성들을 기반으로 중요도를 단순히 계산한 것으로 홍수 영향 요인이 홍수 위험에 미치는 영향을 의미하는 것이 아니기에 Shap Value를 활용하여 홍수 요인이 예측 결과에 미친 영향을 파악하고자 하였다.

홍수 위험 예측 모델에서 홍수 영향 요인이 FRI에 미치는 영향을 Global Shap Value를 활용하여 정량적으로 파악하였다. 분석 결과 FRI 영향 상위 10개 요인의 대부분을 차지하는 기상학적 요인이 주요 홍수 위험 요인으로 확인되었다. 특히, 기상학적 요인 중 최대 강수량을 의미하는 pr\_max가 홍수 위험 예측에 가장 큰 영향을 미쳤다고 나타났으며, 홍수 위험을 증가시키는 주요 요인으로 도출되었다(Fig. 4

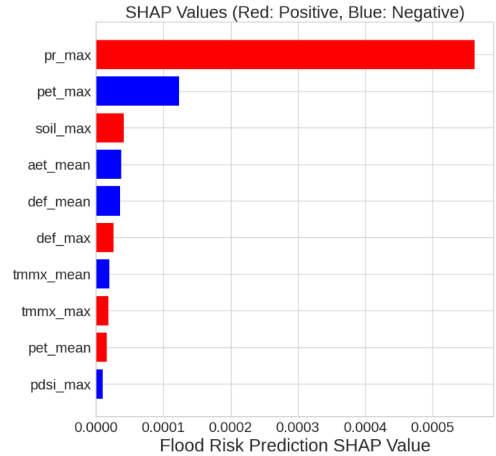


Fig. 4. Global Shap Value

참고). 즉, 1년 중 특정 시기의 월평균 강수량이 높을수록 FRI가 증가한다는 것을 의미하여 강수량이 증가할수록 도시 홍수 피해에 미치는 영향이 가장 크다는 하경준·정주철(2017)의 연구와 동일한 결과도 도출되었다. 반면 최대 기준 증발산량인 pet\_max는 FRI를 감소시키는 요인으로 토양 내 수분 증발 및 식물 증산 작용이 활발할수록 홍수 위험은 낮아진다는 것을 의미한다. 증발산량이 증가할수록 토양 수분량은 줄어들기에 강우를 저류하여 홍수 피해를 저감할 수 있다는 기존 연구 결과와 유사한 결과로도 도시 내 홍수 저감 방안으로 투수층 오픈스페이스를 고려할 필요가 있음을 시사하는 결과이다(조봉준 외, 2020).

나아가 모델 예측 성능에 가장 큰 영향을 미친 변수와 예측 성능 간의 관계를 파악하기 위해 Partial Dependence Plot (PDP)을 활용하여 그 관계를 시각화하였다. PDP는 하나의 변수가 다른 변수와 상호작용하지 않는 가정하에 해당 변수 범위를 변화시켰을 때 모델 예측값이 어떻게 변화하는지 볼 수 있는 방법으로 정확한 모델 해석이 가능하다는 장점이 존재한다(Greenwell, 2017).

모델 성능에 가장 큰 영향을 미친 변수인 최대 강

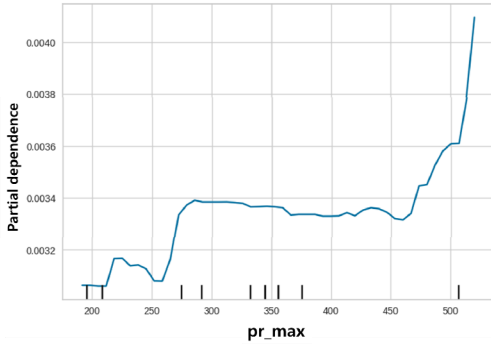


Fig. 5. PDP Analysis Results

수량과 예측값인 FRI의 관계를 분석한 결과는 Fig. 5와 같이 나타났다. 분석 결과, 최대 강수량이 증가함에 따라 홍수 위험은 점진적으로 증가하며, 최대 강수량이 250mm를 기점으로 홍수 위험 가능성이 증가한다. 특히, 최대 강수량이 450mm를 초과하면 홍수 위험이 급격히 증가하였는데 이는 강수량이 임계치를 초과하였을 때 홍수 위험이 급증할 수 있다는 것을 의미한다. 이를 통해 도시 홍수 대응을 위해 최대 강수량 450mm를 주요 경계값으로 고려할 필요가 있음을 확인하였다.

#### 4.2 홍수 위험 지역 예측

앞서 구축된 예측 모델의 검증 및 적용 가능성을 실증하기 위해 2021년 홍수 영향 요인을 기반으로 FRI 예측을 수행하였으며 예측 모델 적용하였을 때의 모델 성능은 Table 3과 같이 나타났다. 모델 구축 시 평가된 모델 성능과 2021년 FRI 예측 성능을 비교한 결과, 2021년 FRI 예측의 모든 오차 지표가 더 낮게 나타났으며  $R^2$  또한 0.8845에서 0.9619로 증가한 것을 확인할 수 있다. 이는 예측 모델이 구축 단계뿐만 아니라 예측 데이터에서도 높은 예측 성능을 유지함을 의미하는 것으로 모델의 일관된 성능과 적용 가능성을 입증하는 결과이다.

2021년 그리드별 FRI를 예측한 결과는 아래 Fig.

Table 3. 2021 FRI Prediction Performance

Model	MAE	MSE	RMSE	$R^2$	RMSLE
Random					
Forest	.0011	.0001	.0084	.9619	.0070
Regressor					

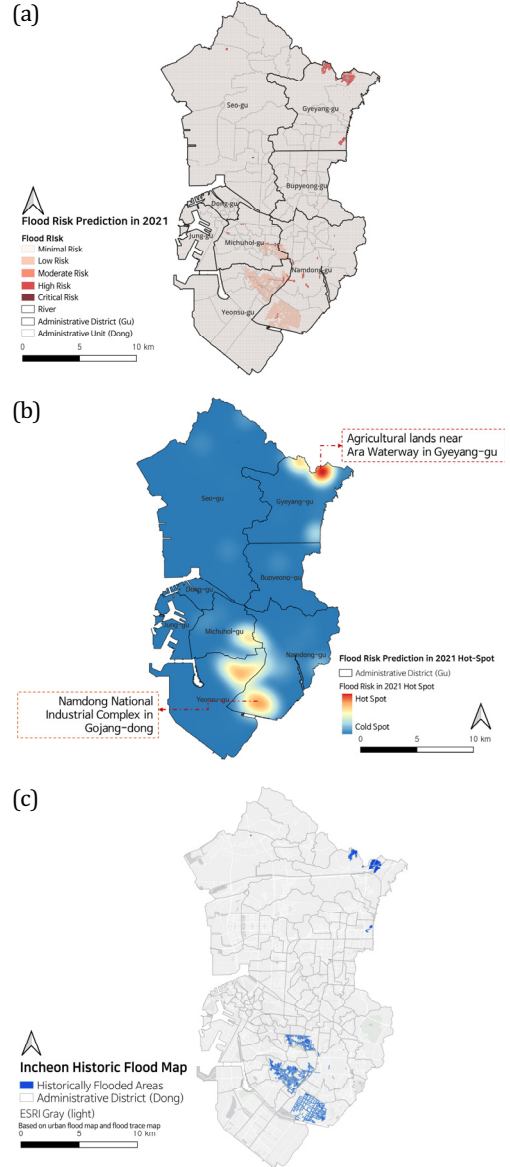


Fig. 6. (a) 2021 Flood Risk Area Prediction Results, (b) Hot-spot Analysis Result and (c) Historic Flood Map

6(a)과 같이 도출되었다. 홍수 위험이 가장 높으며 Fig. 6(b)와 같이 FRI 핫스팟으로 도출된 지역은 계양구 노오지동 아라뱃길 인근 농경지 지역으로 하천 인근에 위치한 기울기가 낮은 완경사 및 평지 지형이다. 이는 토지이용 특성이 농경지이며, 완경사일 수록 침수 피해가 일어날 가능성이 높다고 언급한 장미홍·박창열(2019)의 연구 결과와 유사한 결과로 하천 범람 시 외부로의 우수 유출이 원활하지 않아 이러한 결과가 나타났다고 판단하였다. 남동구 고잔동 인근에 위치한 남동국가산업단지 또한 홍수 위험이 높은 지역으로 예측되었다. 해당 지역은 상습적으로 도시 침수가 발생하는 지역으로 녹지가 부족하며 대부분의 지역이 불투수층으로 개발되어 있어 빗물의 자연 유입 및 배수가 원활하지 않은 지역이다(Fig. 6(c) 참고). 이는 불투수 면적이 넓은 공업 지역의 면적이 클수록 자연재해에 취약하다는 기존 연구의 결과와 유사한 결과이다(심재현·김자은, 2012). 다만, 본 연구의 결과는 기상학적 요인이 주요한 영향을 미치고 있기에 분석 결과를 해석할 때, 이러한 점을 함께 고려할 필요가 있다.

## 5. 결 론

본 연구는 인천광역시를 대상으로 물리적 측면에서 홍수 위험 예측 모델을 통한 도시 홍수 위험도 평가 방법을 제시하였다. 홍수 위험 예측을 위해 선행 연구를 기반으로 2016년~2021년까지의 지리적 요인과 기상학적 요인 및 침수흔적도·하천범람도·도시침수지도를 활용하였다. 2016년부터 2020년까지의 데이터를 활용하여 홍수 위험을 예측하기 위한 모델을 학습시켰으며, 이를 활용하여 2021년 홍수 위험 지역을 예측하여 홍수 예측 모델의 일관성을 확인하는 등 모델 적용가능성을 검증하였다. 이에 본 연구는 예측 모델 활용을 통해 일관적이고 적용 가능한 객관적인 홍수 위험 평가 방안에 대해 제안

하였다는 것에 그 의의가 있다. 나아가 예측 결과를 활용하여 홍수 위험 지역을 파악하기 위한 핫스팟을 도출하여 미시적 관점에서 홍수 대응 정책의 제시를 위한 우선 시행 지역을 제시하였다.

본 연구의 결론 및 의의는 다음과 같다. 첫째, 머신러닝 기반의 FRI 예측 모델을 구축하기 위해 데이터 변동성을 효과적으로 설명하고 오차를 최소화하는데 강점을 가지는 등 가장 우수한 예측 성능을 보인 Random Forest Regressor를 활용하였다. 모델 구축 결과, 최대 강수량이 FRI에 가장 큰 영향을 미치는 변수로 도출되었으며, 최대 강수량이 450mm를 초과할 경우 홍수 위험이 급격히 증가하는 경향을 보이는 것을 확인하였다. 즉, 본 연구에서 홍수 임계치로 도출된 450mm를 고려하여 홍수로 인한 피해를 최소화하기 위한 대책에 이를 고려할 필요가 있다는 것을 시사하는 결과이다. 이를 통해 인천시 도시기본계획 방재 및 안전 재해유형별 저감 대책 내에 이를 고려한다면 자연재해를 사전 예방할 수 있는 저감 대책의 마련이 가능할 것이라 판단된다.

둘째, 모델 검증을 위해 2021년 데이터를 기반으로 FRI를 예측하였으며 기존 모델 구축 시보다 더 낮은 오차 지표와 높은 결정 계수( $R^2 = 0.9619$ )를 보이는 등 예측 모델의 일관성과 적용 가능성을 입증하였다. 현재 홍수 위험 및 홍수 취약 지역은 평가자의 주관에 따라 평가되고 있으며 공간적 한계가 존재하는 실정이다. 이에 본 연구에서 제시한 홍수 위험 평가 및 예측 모델을 활용한다면 객관적이고 미시적인 단위의 홍수 위험 평가 체계를 수립할 수 있을 것이라 사료된다.

셋째, 예측 결과를 통해 도출된 주요 홍수 위험 지역으로는 계양구 노오지동 아라뱃길 인근 농경지와 남동구 고잔동 남동국가산업단지가 도출되었다. 해당 지역의 경우 각각 완경사 지형과 불투수층 면적 확대에 의해 홍수 위험이 증가하는 특성 가진 지역임을 확인하였다. 이러한 결과는 머신러닝 기반 예

측 모델이 도시화 및 기후변화로 인한 홍수 위험 평가에 효과적으로 활용될 수 있음을 시사하며, 도시 기본계획 내 재난 계획에 홍수 저감을 위한 대책 수립 시 물리적 특성 기반의 위험 지역을 우선적으로 고려하는 등 데이터 기반의 지역 맞춤형 홍수 저감 대책 수립을 위한 기초자료를 제공할 수 있을 것이라 사료된다.

그러나 본 연구에는 다음과 같은 한계가 존재한다. 첫째, 홍수 위험을 평가하기 위한 지수 구축 시 물리적 및 기후학적 요인만을 고려하였기에 사회적 측면에서의 홍수 취약성, 민감성 등을 고려하지 못하였다는 한계가 존재하기에 이를 반영하여 홍수 취약지역에 대한 연구를 추가적으로 진행하고자 한다. 둘째, 분석 시 데이터 가용 기간의 한계로 2016년~2021년까지의 연간 데이터를 활용하였기에 FRI의 시계열적인 변화를 반영하지 못하였으며 최신 데이터를 충분히 반영하지 못한 한계가 존재한다. 추후 연구 진행 시 이를 반영하여 최신 데이터 확보하여 일평균 또는 월평균 단위의 데이터를 활용하여 월별 홍수 위험을 평가하고자 한다. 셋째, 본 연구의 대상지인 인천광역시외의 경우 시가지 면적이 넓은 지역으로, 도시 홍수 위험을 평가함에 있어 자연 홍수 요인을 지표로 사용하였다는 한계가 존재한다. 따라서, 추후 연구에서는 도시홍수의 특성을 보다 세밀하게 반영할 수 있는 추가적인 지표 선정이 이루어져야 할 필요가 있다. 넷째, 본 연구에서 개발한 홍수 위험 예측 모델은 기상 관련 변수, 특히 강수량에 크게 의존하는 경향을 보여 단순히 강수량이 많은 지역을 고위험 지역으로 분류할 가능성이 존재한다. 이는 지리적 변수(예: 토지 이용, 배수 인프라 등)를 일정 부분 고려하였음에도 불구하고, 이들이 모델의 예측 결과에 충분한 영향을 미치지 못한 것으로 향후 연구에서는 지리적 요인의 반영 방식을 개선하거나, 토지이용 등 지역적 특성을 보다 정교하게 반영할 수 있는 지리적 변수를 추가적으로 고려할 필요가 있다.

## 참고문헌

1. 국립기상과학원(2018), 「한반도 100년의 기후변화 보고서」, 제주특별자치도.
2. 기상청(2023), 「2022년 이상기후 보고서」, 대전광역시.
3. 김대호·김영오·지희원·강태호(2020), “전국 단위 홍수위험도 평가를 위한 지수 개발과 미래 전망”, 「한국수자원학회 논문집」, 53(5): 323~336.
4. 김동현·이기성·황보종구·김형수·김수전(2022), “AI 기반 모형을 이용한 홍수위 예측 및 홍수피해 예·경보 기법 개발”, 「한국방재학회논문집」, 22(4): 145~156.
5. 김선태·이우섭·정일원·한정민·변영화·김진욱(2023), “미래 기후변화 시나리오에 따른 한반도 유역별 극한 강수 변화 전망”, 「한국기후변화학회지」, 14(2): 83~93.
6. 김지숙·김호용·이성호(2014), “도시 기후변화 재해 취약성분석 방법의 개선방안 검토: 해수면상승 재해를 중심으로”, 「한국지리정보학회지」, 17(1): 50~60.
7. 박문현(2021), “우리나라 홍수위험지도 작성 현황”, 「물과 미래」, 54(8): 8~14.
8. 박지수·임철희(2023), “기후불평등의 공간적 평가: 서울시 홍수재해를 대상으로”, 「한국기후변화학회지」, 14(4): 491~500.
9. 배재현·김진수(2022), “도심 집중호우 피해예방 및 대응방안”, 「Nars 현안분석」, 261
10. 심재현·김자은(2012), “토지이용특성과 자연재해 피해액의 상관성 분석”, 「한국산학기술학회 논문지」, 13(9): 4319~4325.
11. 왕원준·김동현·유영훈·이준형·김경탁·김형수(2021), “홍수위험도 평가를 활용한 인천광역시 복원탄력성 의사결정 지원”, 「한국방재학회논문집」, 21(4): 197~210.
12. 이규민·최진원·전경수(2020), “사회적 평가 지표를 반영한 도시 홍수취약성 평가”, 「LHI Journal」, 11(1): 109~116.
13. 이재희(2024.7.25), “인천 부평 굴포천 홍수방어벽 높이 2.1m로 상향... 계획홍수위 조정, 인천투데이.
14. 이정민·윤정란·김영진·진규남·한형근(2010), “사전재해영향성검토 시 합리적인 홍수유출 모의모형 선정에 관한 연구”, 「LHI Journal」, 1(1): 75~82.

15. 인천연구원(2021), 「팬데믹과 도시정책: 감염병 위기와 도시회복력」, 인천광역시
16. 장미홍·박창열(2019), “침수피해지역 토지이용 특성과 방재대책 연구”, 「한국방재학회 논문집」, 19(7): 135~144.
17. 조봉준·최완민·김영대·김기성·김종건(2020), “머신러닝 기법을 활용한 토양수분 예측 가능성 연구”, 「한국수자원학회 2020년도 학술발표회」, 248.
18. 조형진(2017), “기후변화에 따른 도시홍수 취약성 평가모델 연구”, 박사학위논문, 인하대학교.
19. 조형진·김경배(2015), “기후변화로 인한 홍수위험의 사회적 취약성 평가 연구: 인천시를 사례로”, 「기후연구」, 10(4): 341~354.
20. 하경준·정주철(2017), “도시화와 강우량이 도시홍수 피해에 미치는 영향에 관한 연구”, 「국토계획」, 52(4): 237~252.
21. 환경부(2024), 「굴포천 하천기본계획(변경) 보고서」, 경기.
22. Abatzoglou, J. T., S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch (2018), “TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958-2015”, *Scientific Data*, 5(1): 170191.
23. Acosta-Quesada, M. and A. Quesada-Román (2025), “Landslide and Flood Risk Assessment in a Rapidly Urbanizing Municipality of Costa Rica”, *Journal of South American Earth Sciences*, 152: 105330.
24. Ali, M. (2020), “PyCaret: An Open Source, Low-Code Machine Learning Library in Python”, PyCaret Version, 2
25. Armenakis, C., E. X. Du, S. Natesan, R. A. Persad, and Y. Zhang (2017), “Flood Risk Assessment in Urban Areas Based on Spatial Analytics and Social Factors”, *Geosciences*, 7(4): 123.
26. Booth, D. B. (1991), “Urbanization and the Natural Drainage System: Impacts, Solutions, and Prognoses”, *The Northwest Environmental Journal*, 7: 93~118.
27. Cai, J., X. Li, L. Liu, Y. Chen, X. Wang, and S. Lu (2021), “Coupling and Coordinated Development of New Urbanization and Agro-Ecological Environment in China”, *Science of the Total Environment*, 776: 145837.
28. Dereci, U. and G. Tuzkaya (2024), “An Explainable Artificial Intelligence Model for Predictive Maintenance and Spare Parts Optimization”, *Supply Chain Analytics*, 8: 100078.
29. Ermagun, A., V. Smith, and F. Janatabadi (2024), “High Urban Flood Risk and no Shelter Access Disproportionally Impacts Vulnerable Communities in the USA”, *Communications Earth & Environment*, 5(1): 2.
30. Feng, B., Y. Zhang, and R. Bourke (2021), “Urbanization Impacts on Flood Risks Based on Urban Growth Data and Coupled Flood Models”, *Natural Hazards*, 106(1): 613~627.
31. Greenwell, B. M. (2017), “Pdp: An R Package for Constructing Partial Dependence Plots”, *R Journal*, 9: 421~436.
32. Huong, H. T. L. and A. Pathirana (2013), “Urbanization and Climate Change Impacts on Future Urban Flooding in can Tho City, Vietnam”, *Hydrology and Earth System Sciences*, 17(1): 379~394.
33. Kim, H. W., D. McCarty, and M. Jeong (2023), “Examining Commercial Crime Call Determinants in Alley Commercial Districts before and after COVID-19: A Machine Learning-Based SHAP Approach”, *Applied Sciences*, 13(21): 11714.
34. Pan, H., J. Page, L. Zhang, C. Cong, C. Ferreira, E. Jonsson, and Z. Kalantari (2020), “Understanding Interactions between Urban Development Policies and GHG Emissions: A Case Study in Stockholm Region”, *Ambio*, 49(7): 1313~1327.
35. Ramkar, P. and S. M. Yadav (2021), “Flood Risk Index in Data-Scarce River Basins Using the AHP and GIS Approach”, *Natural Hazards*, 109(1): 1119~1140.
36. Rentschler, J., P. Avner, M. Marconcini, R. Su, E. Strano, M. Vousdoukas, and S. Hallegatte (2023), “Global Evidence of Rapid Urban Growth in Flood Zones since 1985”, *Nature*, 622(7981): 87~92.
37. Rosenzweig, B. R., L. McPhillips, H. Chang, C. Cheng, C. Welty, M. Matsler, and C. I. Davidson (2018), “Pluvial Flood Risk and Opportunities for



- Resilience”, *Wiley Interdisciplinary Reviews: Water*, 5(6): e1302.
38. Sanders, B. F., J. E. Schubert, K. A. Goodrich, D. Houston, D. L. Feldman, V. Basolo, and R. A. Matthew (2020), “Collaborative Modeling with Fine-Resolution Data Enhances Flood Awareness, Minimizes Differences in Flood Perception, and Produces Actionable Flood Maps”, *Earth’s Future*, 8(1): e2019EF001391.
  39. Tellman, B., J. A. Sullivan, C. Kuhn, A. J. Kettner, C. S. Doyle, G. R. Brakenridge, T. A. Erickson and D. A. Slayback (2021), “Satellite Imaging Reveals Increased Proportion of Population Exposed to Floods”, *Nature*, 596(7870): 80~86.
  40. Theodosopoulou, Z., I. M. Kourtis, V. Bellos, K. Apostolopoulos, C. Potsiou, and V. A. Tsihrintzis (2022), “A Fast Data-Driven Tool for Flood Risk Assessment in Urban Areas”, *Hydrology*, 9(8): 147.
  41. Wang, J., W. Y. Chuck, and C. Shi-Jie (2022), “Urban Development in the Context of Extreme Flooding Events”, *Indoor and Built Environment*, 31(1): 3~6.
  42. Wing, O. E., W. Lehman, P. D. Bates, C. C. Sampson, N. Quinn, A. M. Smith, and C. Kousky (2022), “Inequitable Patterns of US Flood Risk in the Anthropocene”, *Nature Climate Change*, 12(2): 156~162.
  43. Global Flood Database, “Findings from Our Research”, Accessed 12 January, 2025. <https://global-flood-database.cloudtostreet.ai/>

## 요 약

본 연구는 기계 학습 기반 예측 모델을 활용하여 인천광역시의 홍수 위험을 평가하는 데에 그 목적이 있다. 2016년부터 2021년까지의 지리적 및 기후적 요인을 활용하여 홍수 위험 예측 모델을 학습하였으며, 2021년의 홍수 위험 지역을 예측하여 모델의 일관성과 적용 가능성을 검증하였다. 모델 성능 평가 결과, Random Forest Regressor 모델이 가장 우수한 예측 성능을 보였으며 데이터의 변동성을 효과적으로 설명하고 예측 오류를 최소화하는 것으로 나타났다. 나아가, 최대 강수량이 홍수 위험에 가장 큰 영향을 미치는 변수로 확인되었으며, 450mm를 초과할 경우 홍수 위험이 급격히 증가하는 경향을 보였다. 이를 기반으로 2021년 홍수 위험 지수(Flood Risk Index)를 예측한 결과, 모델 구축 단계보다 오차가 작고 높은 설명력( $R^2 = 0.9619$ )이 도출되는 등 높은 모델 신뢰도를 보였다. 주요 홍수 위험 지역으로는 계양구 아라뱃길 인근 농경지와 남동구 고잔동 남동국가산업단지로 나타났으며, 각각 완만한 경사와 불투수층 비율이 높다는 특징을 가진 지역으로 분석되었다. 본 연구 결과는 기계 학습 기반 예측 모델이 도시화 및 기후변화로 인한 홍수 위험을 효과적으로 평가할 수 있으며, 도시 계획에서 데이터 기반의 지역 맞춤형 홍수 저감 대책을 마련하는 데 기여할 수 있음을 시사하는 결과이다.

**주제어:** 기후변화, 도시화, 랜덤포레스트, 홍수 저감, 재난 대응