

GEE기반 위성 자료를 활용한

경상북도, 대구 지역 특화 산불 위험 예측 모델 개발

과목명: 산업수학

팀장:현대호

팀원:안가은

팀원:장동환

목차

1. 서론.....	4
1.1 연구 배경 및 필요성.....	4
1.2 연구 목적 및 차별성.....	5
2. 연구 범위 및 자료	6
2.1 대상 지역 개요 (경상북도·대구광역시).....	6
2.2 활용 자료	6
3. 데이터 전처리 과정	7
3.1 주요 변수 산출 방법.....	8
3.2 산불 발생·비발생 데이터 구축 과정	8
4. 학습 및 모델링	9
4.1 클러스터링 분석.....	9
4.2 분류 모델	12
4.3 회귀 모델	14
5. 좌표 추출 및 위험 후보 지역 선정	16
5.1 후보 좌표 선정	16
5.2 최종 대표 좌표 결정.....	17
5.3 위험 유형 우선 배정.....	17
6. 좌표 유형 분류	17
7. 시계열 데이터 구축	19
7.1 수집 기간 및 변수	19
7.2 전체 데이터 규모.....	19
8. 계절 구분 및 모델 분리.....	19
9. LSTM 기반 예측 모델 학습	20
9.1 모델 구조 및 1·2차 실험.....	21

9.1.1 1차 모델: FFDRI 직접 예측 시도.....	21
9.1.2 1차 모델의 한계 및 원인 분석.....	21
9.1.3 2차 모델: DWI 예측으로의 전환.....	23
9.1.4 2차 모델의 한계.....	23
9.2 지역특화 FFDRI_NEW 지수 개발.....	24
9.2.1 FFDRI_new 지수 개발 배경.....	24
9.2.2 시행착오를 통한 타깃 변수 검증.....	25
9.2.2 시행착오의 결론: 새로운 접근 필요성.....	25
9.2.3 FFDRI_new의 변수 구성 논리.....	25
10. FFDRI_NEW 기반 LSTM/GRU 시계열 예측 모델 구축.....	26
10.1 모델 구조 및 학습 개요.....	26
10.2 LSTM-GRU 모델 성능 비교	27
10.3 계절 분리 학습의 효과	27
10.4 FFDRI_NEW 분포 기반 위험등급 정의.....	28
11. 실제 사례 기반 예측 성능 검증.....	29
11.1 2025-04-28 대구 함지산 산불 사례 검증.....	29
11.2 2025-12-03 겨울철 검증 (SPRING VS FALL 모델 비교).....	30
12. 경북·대구 주요 산 5곳 대상 FFDRI_NEW 7일 예측 적용	32
12.1 예측 결과 요약.....	32
12.2 지역 위험도 종합 분석	34
12.3 공간 기반 산불위험도 지도 시각화 시스템 구축.....	35
13. 결론 및 향후과제	37
13.1 결론.....	37
13.2 향후 과제.....	38
14. 참고문헌.....	40

1. 서론

1.1 연구 배경 및 필요성

최근 전 세계적으로 기후 변화와 지구 온난화가 진전됨에 따라, 산불 발생이 빈번해지고, 산불 지속이 장기화되는 경향이 나타나고 있다. NASA의 연구에 따르면 전 지구적으로 화재가 발생하기 쉬운 기간이 점점 확대되고 있으며, 극단적 산불 활동은 빈도, 강도, 규모 면에서 두 배 이상 증가한 것으로 조사되었다. 또한, 2001년부터 2023년 사이 산불 관련 탄소 배출량은 전 세계적으로 약 60% 증가한 것으로 나타나, 산불이 기후-환경 변화의 상호작용 속에서 점점 더 심각한 문제로 부상하고 있음을 보여주고 있다.

이러한 전 세계적 현상은 한국에서도 예외가 아니다. 산림청 「산불통계(2024)」에 따르면, 2020년대(2020~2024년 기준)의 연평균 산불 발생 건수는 520건으로 2000년대(523건)와 유사한 수준이지만, 피해 면적은 6,721ha로 2010년대(857ha)보다 약 8배, 2000년대(3,726ha)보다도 2배 가까이 증가하였다.

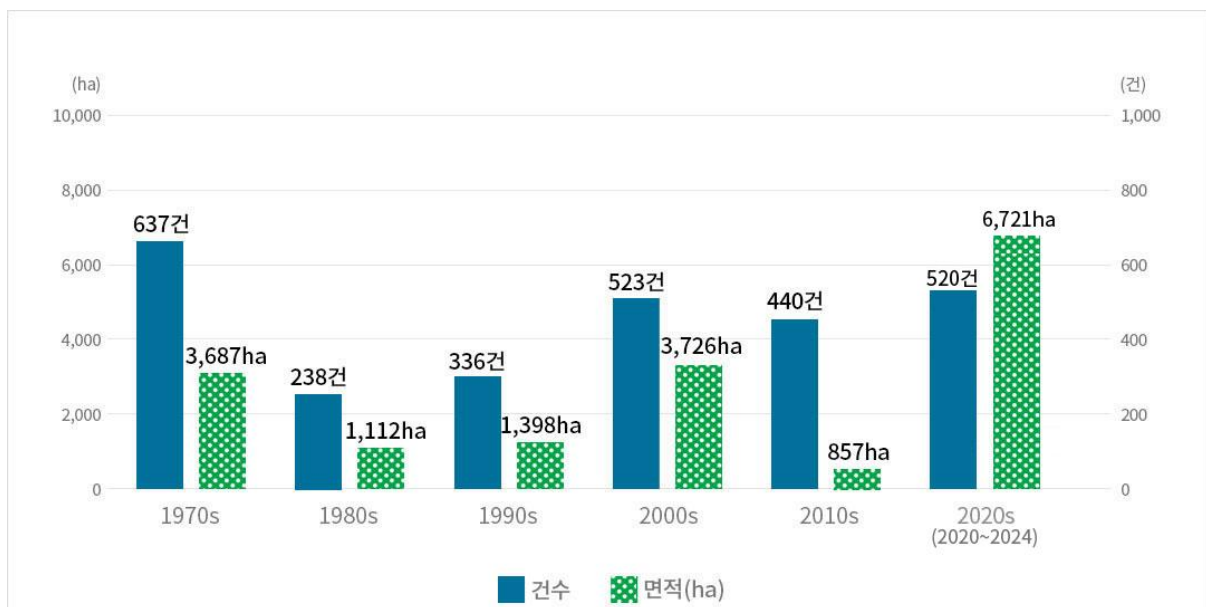


Figure 1. 산림청 10년간 산불 발생 현황

특히 최근 4년치 데이터임에도 불구하고 과거 10년 단위 평균과 비교했을 때 피해 규모가 훨씬 크다는 점은, 국내 산불이 단순한 발생 빈도 증가를 넘어 대형화·장기화되는 추세임을 시사한다.

이러한 통계는 한국 또한 산불 위험으로부터 더 이상 안전지대가 아님을 의미하며, 과거

의 경험적 지수나 일률적 예측 체계로는 변화하는 환경에 대응하기 어렵다는 점을 보여 준다. 따라서 기후·지형·식생 등 지역별 환경 요인을 반영한 정밀한 산불 위험 예측 모델의 구축이 국가 차원의 예방적 대응과 자원 관리 효율화를 위해 필수적이다.

2025년 3월 말 대구 함지산에서 발생한 산불은 재발화로 이어져 2,000여 명의 주민이 대피하는 등, 산불 피해가 더 이상 특정 지역에 국한되지 않음을 보여주었다.

이러한 상황은 한국 또한 기후 변화에 따른 산불 위험으로부터 자유롭지 않으며, 과거의 경험적 지수, 일률적인 예측 체계로는 급변하는 환경에 대응하기 어렵다는 점을 시사한다.

따라서 기후·지형·식생 등 지역별 환경 요인을 반영한 정밀한 산불 위험 예측 체계 구축은 국가 차원의 예방적 대응과 효율적인 자원 배분을 위해 반드시 필요한 과제이다.

1.2 연구 목적 및 차별성

본 연구의 목적은 우리나라가 지역별로 토양, 식생, 기후 등 환경적 요인에서 큰 편차를 보인다는 점에 착안하여, 지역적 특성을 반영한 산불 위험도 예측 모델을 개발하는 데 있다.

현재 산림청에서는 기상·지형·임상 정보를 활용해 전국 단위의 산불위험지수를 산출하고 있으나, 전국에 동일한 알고리즘을 적용하기 때문에 지역별 환경 차이를 충분히 반영하기 어렵다는 한계가 존재한다.

실제로 기상청 통계에 따르면 한반도는 남북으로 뚜렷한 기후대 차이를 보이며, 동서 방향으로 강수량과 풍속 등 주요 기상 조건이 크게 다르다. 또한 국토의 약 70%가 산악 지대로, 고도에 따라 식생이 난대림·온대림·한대림으로 구분되는 등 지형·식생 구조의 다양성도 매우 크다.

이러한 지역적 요인들은 단일한 전국 지수로는 산불 위험을 충분히 설명하기 어려운 복잡성을 지니며, 이는 지역 맞춤형 접근의 필요성을 시사한다. 따라서 본 연구는 지역별 환경 특성을 반영한 정밀 산불위험도 예측 모델을 구축함으로써, 기존 전국 단일지수의 한계를 보완하고 보다 현실적이고 세밀한 산불 위험 예측 체계를 제시하고자 한다.

2. 연구 범위 및 자료

2.1 대상 지역 개요 (경상북도·대구광역시)

본 연구의 대상 지역은 경상북도 및 대구광역시 일원으로, 한반도 남동부에 위치한다.

경상북도는 국토의 약 19%를 차지하는 광역 행정구역으로, 산림 면적 비율이 약 71%에 달하며(경상북도 공식 홈페이지), 동해안 산악지대와 내륙 분지 지역으로 구분된다. 특히 동해안 지역(포항·영덕·울진 등)은 해양성 기후와 강풍의 영향으로 건조하고 바람이 강한 날이 잦아 산불 발생 위험이 높고, 서부 내륙권(안동·예천·의성 등)은 넓은 산림 분포와 낮은 인구 밀도로 인해 산불 대응 여건이 상대적으로 취약하다. 대구광역시는 경북 남부 내륙에 위치하며, 분지형 지형 구조와 높은 평균 기온·낮은 상대습도로 인해 산불 발생 시 확산 위험이 크다.

본 연구에서는 지리적·기후적 연속성 및 유사한 산림 환경 특성을 고려하여, 대구광역시를 경북 서부 내륙권과 동일 권역으로 묶어 분석하였다. 이러한 설정은 지역별 맞춤형 위험 예측 모델을 구축하기 위한 공간적 기준을 제공한다.

2.2 활용 자료

본 연구에서는 경상북도 및 대구 지역의 산불 위험지수(Fire Danger Index) 개발을 위해 Google Earth Engine(GEE) 기반의 다중 환경위상자료(Earth Observation Data)를 핵심 데이터로 활용하였다.

최근 전 세계적으로 위성관측자료의 축적 속도가 급격히 증가함에 따라, 이를 효율적으로 관리하고 분석할 수 있는 클라우드 기반 공간정보 분석 플랫폼의 중요성이 높아지고 있다. 이 중 GEE는 대표적인 지리공간 빅데이터 분석 플랫폼으로, Landsat, Sentinel, MODIS 등 다양한 시계열 위성영상과 ERA5-Land, TerraClimate, SMAP 등 기상·지형 관련 글로벌 데이터셋을 무료로 제공한다. 또한 대규모 병렬 연산이 가능한 클라우드 컴퓨팅 환경을 기반으로, 국가 단위 이상의 광범위한 지역에 대해 고해상도 공간자료를 신속하고 일관되게 산출할 수 있는 장점을 지닌다.

이러한 특성은 산불 예측 연구에 특히 적합하다. 산불 위험은 단일 요인보다는 기상(온

도, 습도, 풍속, 강수), 지형(고도, 경사, 향), 식생(LST, Soil Moisture) 등 다양한 환경요인의 복합적인 상호작용에 의해 결정되므로, 이를 장기간 시계열로 추적할 수 있는 GEE의 데이터 접근성과 처리 성능은 매우 유용하다.

본 프로젝트에서는 GEE를 활용하여 GPS 좌표 기반으로 주요 환경변수들을 열람·추출하였다.

구체적으로는 다음과 같다.

- 기상 요인: 일평균기온(Tmean), 일최고기온(Tmax), 일최저기온(Tmin), 상대습도(RH), 실효습도(EH), 풍속(WSPD), 강수량(TP_mm)
- 지형 요인: 고도(DEM), 경사(Slope)
- 식생 및 토양 요인: 지표면온도(LST), 토양수분(Soil Moisture)

이렇게 구축된 GEE 기반 위상자료는 실제 산불 발생 이력 데이터와 결합되어, FFDRI(산불위험도지수) 산출 및 산불 발생 여부를 분류·예측하는 머신러닝 모델(XGBoost, Random Forest 등)의 입력 변수로 활용되었다.

즉, 본 프로젝트에서의 GEE는 단순한 데이터 수집 도구가 아니라, 다양한 기상·지형·식생 지표를 공간적으로 통합하고 이를 기반으로 산불 위험을 정량화하는 핵심 분석 인프라로 기능하였다. 또한, 산림청에서 제공하는 임상도(임상지수) 데이터와 과거 산불 발생 이력 데이터를 추가로 수집·결합하여, 지역별 산림 특성과 실제 산불 발생 간의 연관성을 종합적으로 분석하였다.

3. 데이터 전처리 과정

본 연구에서는 산불위험 예측모델 구축을 위한 핵심 단계로, GEE(Google Earth Engine)를 활용하여 산불 관련 환경 변수를 일 단위(date), 좌표 단위(lon·lat)로 정제·결합하였다.

특히 기존 연구들이 사용한 격자 평균값(5~10 km)을 배제하고, 실제 산불 발생 위치의 점 단위(point-based) 데이터를 직접 추출함으로써 공간적 해상도를 향상시켰다. 이 과정은 FFDRI 산출 및 분류·회귀 모델 학습의 기반이 되는 고정밀 입력 데이터를 확보하기

위한 전처리 단계이다.

3.1 주요 변수 산출 방법

FFDRI(Forest Fire Danger Rating Index)는 기상지수(DWI), 임상지수(FMI), 지형지수(TMI), 그리고 계절 가중치(day_weight)로 구성된다.

DWI는 기온, 습도, 풍속, 강수량의 조합으로 산출되며, 산불 위험도의 핵심 변수이다.

FMI는 산림청 임상도(2013, 2019, 2024)와 2022년 시군구 임상도를 결합하여 산림의 연료 특성을 반영하였다.

TMI는 SRTM DEM 기반의 고도와 경사도를 이용하여 산불 확산 가능성을 정량화하였다.

day_weight는 날짜별 계절 민감도를 반영하여, 봄철·가을철의 계절적 영향 차이를 보정하였다.

이와 같은 구성요소를 통합함으로써, 단일 기상지표에 의존하지 않는 종합적 산불위험도 산출 체계를 구축하였다.

3.2 산불 발생·비발생 데이터 구축 과정

산불 자료는 산림청에서 제공하는 2015년부터 2024년까지의 산불피해대장을 활용하였다. 산불 발생 위치가 텍스트 형태의 주소로 제공되었기 때문에, 이를 법정동 코드 기반의 행정구역 명칭으로 매칭한 후 카카오 개발자센터의 오픈 API를 이용하여 경·위도 좌표로 변환하였다.

이후 각 산불 이벤트에 대해 날짜(date), 위도(lat), 경도(lon)를 기준으로 ERA5-Land의 기상변수(Tmax, RH, WSPD, TP_mm)를 매칭하였으며, 이를 산불 발생(positive) 데이터로 정의하였다.

비발생 데이터는 동일한 시기·유사한 기상 및 지형 조건을 만족하는 산림 지역 내부에서만 추출하였다.

이를 위해 조건부 랜덤 샘플링(Conditional Random Sampling) 기법을 적용하여, 실제 산

불 발생지와 동일한 환경 조건을 갖되 발화 이력이 없는 지역을 비발생(negative) 표본으로 선정하였다.

이 접근은 단순한 무작위 선택(Random Sampling)이 아니라, 환경적 유사성(environmental similarity)을 유지함으로써 Positive-Negative 데이터 간의 편향을 최소화하는 절차이다.

각 데이터는 date-lon-lat을 기본 식별키로 하여 GEE에서 추출된 ERA5-Land, SRTM DEM, GPM, 산림청 임상도 등의 정보를 결합하였고, 총 약 20개의 변수(Tmax, RH, WSPD, TP_mm, DEM, FMI, TMI, day_weight 등)를 포함하였다.

데이터 구축 후 FMI 분포를 검토한 결과, 약 44%의 좌표가 FMI = -1로 나타났다. 이는 실제 산불이 산림 내부에서 발생했음에도 대표 좌표가 도로·마을 등 비산림 지역으로 지정된 오류 때문이었다.

이 문제를 해결하기 위해, FMI 결측 좌표를 ERA5-Land의 시간별 2m 기온 자료를 이용하여 가장 기온이 높았던 인근 픽셀(반경 3~8 km)로 재배치하였다.

그 결과 FMI 결측 비율은 44.1% → 26.0%로 감소하였고, 산림좌표의 유효비율은 74%로 향상되었다. 모든 변수는 평균 0, 표준편차 1로 표준화(Standardization)하였으며, 2일 이하의 짧은 결측은 선형 보간(interpolation)으로 보완하였다.

이와 같은 전처리 과정을 통해 구축된 데이터셋은 기상·지형·임상·식생·계절 변수를 통합한 좌표 단위 기반의 고해상도 FFDRI 학습용 데이터셋으로 완성되었다. 이 데이터는 이후 단계인 클러스터링(4.1), 분류(4.2), 회귀(4.3) 모델의 입력 자료로 사용되며, 시공간적 정합성과 물리적 신뢰성을 동시에 확보한 기반 데이터로 기능한다.

4. 학습 및 모델링

4.1 클러스터링 분석

본 연구에서는 수집된 산불 관련 환경 데이터를 바탕으로, 산불 발생 특성이 계절(봄 / 가을·겨울) 및 지역(동해안 / 내륙)에 따라 차이를 보이는지 검증하기 위해 클러스터링

분석을 수행하였다. 이를 통해 데이터 내 잠재적 패턴을 탐색하고, 이후 분류 및 회귀 모델의 구조를 설계하는 데 참고하고자 하였다.

클러스터링에는 K-Means 알고리즘을 사용하였으며, $k=2$ 및 $k=4$ 의 두 가지 경우를 시도하였다. 먼저, 모든 기상 및 지형 변수(기온, 습도, 강수량, 풍속, 지형고도, FFDRI, FRI_norm 등)를 포함한 상태로 군집화를 진행하였다. 데이터의 불균형(특히 해안권 가을 데이터의 개수가 99개로 매우 적음)을 해소하기 위해, 각 지역-계절 조합의 표본 수를 동일하게 조정한 후 클러스터링을 수행하였다.

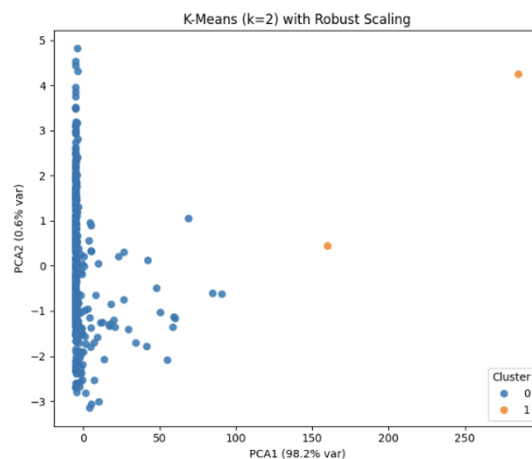


Figure 2. 1차 클러스터링($k=2$)

1차 클러스터링($k=2$) 결과, 일부 클러스터가 봄 데이터에 다소 치우치는 경향은 있었으나, 계절 간의 명확한 구분은 나타나지 않았다. 또한 내륙과 해안권의 분포 차이도 거의 존재하지 않아, 단순한 기상 변수만으로는 지역적 특성을 설명하기 어렵다는 점을 확인하였다.

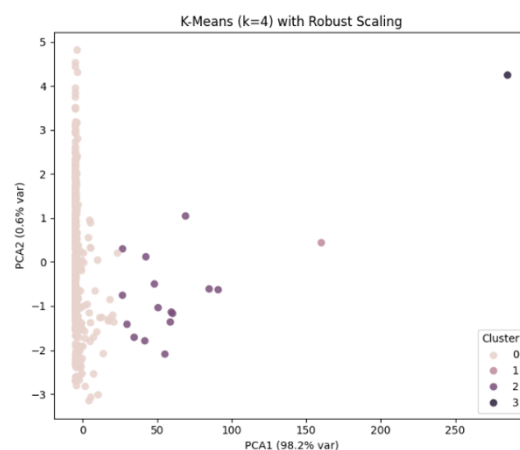


Figure 3. 1차 클러스터링($k=4$)

1차 클러스터링(k=4) 결과, 대부분의 데이터가 단일 클러스터(Cluster 0)에 집중되고 나머지 세 군집은 소수의 데이터만을 포함하는 불균형한 분포가 나타났다. 이로 인해 계절적·지역적 패턴에 대한 구분이 여전히 충분하지 않음을 확인하였다.

이러한 결과를 바탕으로 기존 변수 조합만으로는 충분한 분리도를 확보하기 어렵다고 판단하여, 2차 클러스터링에서는 NDVI와 DSR_total_MJ/m² 변수를 추가로 수집하여 재분석을 수행하였다. NDVI는 식생 상태를, DSR_total은 일조량(이후 직관성을 위해 sunlight로 표기)을 나타내는 지표로, 두 변수 모두 산불 발생 가능성과 밀접한 관련이 있는 것으로 알려져 있다.

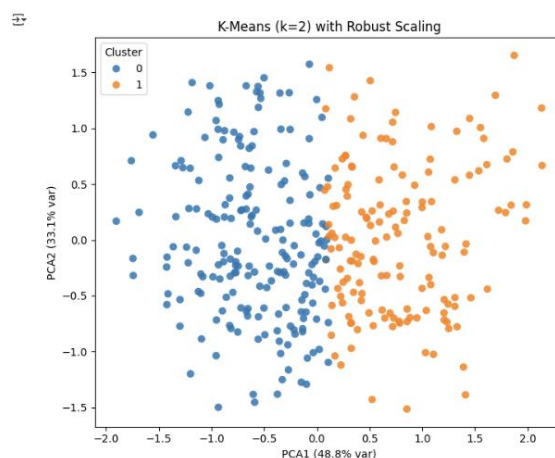


Figure 4. 2차 클러스터링(k=2)

	Season	fall_winter	spring
cluster_k2			
0		0.118	0.882
1		0.724	0.276

Figure 5. 2차 클러스터링(k=2)

2차 클러스터링(k=2) 결과, 계절적 특성에 따른 군집 분리가 부분적으로 확인되었다.

Cluster 0은 봄 데이터가 약 88%, Cluster 1은 가을·겨울 데이터가 약 72%를 차지하며, 이는 NDVI와 일조량이 계절 변화에 민감하게 반응한 결과로 해석된다. 반면, 지역적 분리도는 여전히 뚜렷하지 않았으며, 두 클러스터 모두 내륙과 해안권의 비율이 약 5:5로 유사하게 분포하였다.

종합적으로, 클러스터링 결과는 지역 간의 차이는 미미하고, 계절적 차이가 주요 분리 요인으로 작용함을 시사하였다. 따라서 이후 단계의 모델링(분류·회귀)에서는 지역 구분 변

수를 제외하고, 계절적 구분만을 반영한 학습 구조로 수정하였다. 이는 모델의 복잡도를 낮추고, 계절별 특성을 보다 정교하게 반영할 수 있도록 하기 위한 결정이었다.

4.2 분류 모델

클러스터링 분석 결과, 지역적 구분(내륙·해안권)에 따른 차이는 미약하나 계절적 차이가 뚜렷하게 나타났다. 이에 따라 본 연구의 분류 모델은 계절(봄 / 가을·겨울) 특성을 반영한 산불 발생 예측 모델로 설계되었다. 목표는 위도, 경도, 날짜 단위로 수집된 환경 변수들을 기반으로 해당 시점의 산불 발생 여부(1: 발생, 0: 비발생)를 예측하는 것이다.

(1) 모델 개요 및 절차

분류 과정은 다음 네 단계로 구성되었다.

1. 과거 상태를 반영한 파생 변수(feature) 추가
2. 트리 기반 앙상블(Random Forest, Extra Trees) 학습
3. F2-score 기반 임계값(threshold) 최적화
4. 산불 발생 여부(0/1) 최종 예측

(2) 파생 변수 생성

산불은 단일 일자의 기상 조건뿐 아니라 며칠간의 누적 환경 변화에 큰 영향을 받는다. 따라서 본 연구에서는 시간적 연속성을 반영하기 위해 이동평균, 누적강수, 건조 지속일수를 포함한 파생 변수를 추가하였다.

- 이동평균(3일, 7일): 최근 기온, 습도, 풍속, 일교차(DTR)의 추세 반영
- 누적강수량(3일, 7일): 토양 수분 상태 반영
- 연속 무강수일수(DrySpell): 건조 지속성을 정량화

이러한 변수를 통해 모델은 단기적인 날씨 변화뿐 아니라 최근 며칠간의 기상 흐름을 종합적으로 학습하도록 구성되었다. 데이터셋은 랜덤 샘플링 구조로 인해 동일 위치·날짜 간격이 불규칙하므로, 최근 시점의 인접 데이터를 탐색하여 근사적으로 기상 추세를 반

영하였다.

(3) 모델 학습

모델 학습은 2015-2022년 데이터를 훈련용, 2023-2024년 데이터를 검증용으로 사용하여 시계열적 일반화 성능을 확보하였다.

사용된 모델은 Random Forest와 Extra Trees로, 두 모델 모두 트리 기반 앙상블 구조를 가지며 변수 간 비선형 관계와 상호작용을 포착할 수 있다.

- Random Forest: 피쳐 선택은 무작위이나 split point(임계값)는 데이터 기반 계산
- Extra Trees: 피쳐 및 임계값 모두 무작위 선택

따라서 두 모델을 앙상블 평균(ensemble averaging) 하여 예측 안정성과 일반화 성능을 강화하였다.

(4) F2-score 기반 임계값 설정

모델의 출력은 산불 발생 확률(0~1) 형태이므로, 이를 0 또는 1로 변환하기 위한 임계값 설정이 필요하다. 일반적인 0.5 기준은 불균형 데이터에서 비발생을 과도하게 예측하는 문제가 있다.

이에 본 연구는 재현율(Recall)에 2배 가중치를 둔 F2-score를 기준으로 최적의 임계값을 산출하였다.

- Precision: 예측한 '발생' 중 실제 발생 비율
- Recall: 실제 발생 중 놓치지 않은 비율
- F2-score: Recall 비중이 높은 조화평균

이 접근법은 산불 발생을 놓치지 않는 것(False Negative 최소화)을 우선시하는 모델 설계 철학에 부합한다.

(5) 예측 및 성능 평가

최종적으로 각 지점별 예측 확률에 대해 최적 threshold를 적용하여 산불 발생 여부를 분류하였다.

모델 평가는 Accuracy, Precision, Recall, F1-score를 기준으로 수행하였으며,

- 2023년: Accuracy 0.70, Recall 0.50, Precision 0.69, F1-score 0.59
- 2024년: Accuracy 0.66, Recall 0.33, Precision 0.80, F1-score 0.47

이 결과는 전체 정확도는 보통 수준이나, Recall이 낮아 일부 산불 발생을 놓친 사례(FN)가 존재함을 의미한다. 이는 랜덤 샘플링 구조로 인한 시간적·공간적 불연속성의 영향으로 해석된다. 그럼에도 불구하고 Precision이 높게 유지되어, 한 번 '산불 발생'으로 예측했을 때 실제로 맞을 확률이 높았다.

따라서 데이터 셋이 향후 재구성되어 시계열 적합성과 표본 균형이 개선된다면, 본 모델은 더 높은 예측 성능을 보일 잠재력이 크다.

4.3 회귀 모델

회귀 모델은 산불위험지수(FFDRI, Fire Danger Rating Index)의 연속값을 예측하기 위해 구축되었다.

본 모델의 목표는 위도, 경도, 날짜 단위의 환경 변수(기온, 습도, 풍속, 강수량 등)와 과거 FFDRI 값을 입력으로 받아, 향후 L일 뒤(L0~L7)의 FFDRI를 예측하는 것이다.

즉, 단순히 "현재 위험도"를 계산하는 것을 넘어, 1~7일 후 산불 위험 수준을 조기 탐지할 수 있도록 설계되었다.

(1) lag 변수 추가

산불위험지수는 단기적인 변동보다 시간적 연속성(autocorrelation)이 강한 변수이므로, 과거 데이터를 반영하기 위한 lag 변수를 생성하였다.

각 지역(lat, lon) 및 계절(season_cluster) 내에서 날짜(date) 순으로 데이터를 정렬한 뒤, 전날과 이틀 전의 FFDRI 값을 각각 lag1, lag2로 추가하였다.

이러한 과정은 모델이 "이전 위험도가 다음날에도 영향을 미친다"는 시계열적 관계를 학습하도록 돕는다.

(2) 모델 학습 구조

FFDRI_L0~L7 각각에 대해 단일 회귀 모델(single regression model)을 개별 학습시켰으며, 데이터의 양과 복잡도에 따라 세 가지 모델을 자동 선택(adaptive model selection) 하는 구조를 채택하였다.

- ① LightGBM : 데이터가 충분하고 변수 간 상호작용이 복잡한 경우, 비선형 관계를 포착하기 위한 LightGBM을 사용하였다. 이 모델은 빠른 학습 속도와 높은 예측력을 동시에 확보할 수 있다.
- ② ElasticNet : 데이터가 적거나 단순한 경우에는 ElasticNet을 활용하였다. ElasticNet은 L1, L2 규제를 동시에 적용하여 과적합을 방지하고, 적은 데이터에서도 안정적인 회귀 결과를 제공한다.
- ③ Dummy Regressor : 데이터가 극히 적거나 사용 가능한 피처 수가 5 미만일 경우, 평균값으로 예측하는 단순 모델(Dummy Regressor)을 적용하였다. 이는 학습 불안정으로 인한 왜곡을 최소화하기 위한 보완책이다.

이 세 모델은 조건에 따라 자동으로 선택되어 학습되며, 결과적으로 데이터의 양과 복잡도에 따른 최적의 예측 구조를 유연하게 구성하였다.

(3) 성능 평가 및 검증

모델 학습은 2015-2022년 데이터를 학습용, 2023년 데이터를 검증용으로 사용하였다. 각 리드(L0~L7)에 대해 예측 성능을 비교하여 단기 예측(1일 후)과 장기 예측(7일 후)의 정확도를 함께 분석하였다.

평가 지표로는 다음 네 가지를 사용하였다.

- MAE (Mean Absolute Error): 예측 오차의 절댓값 평균. 작을수록 예측 정확도가 높음을 의미.
- RMSE (Root Mean Squared Error): 큰 오차에 더 민감한 지표로, 예측의 안정성을 평가.
- 피어슨 상관계수 (r): 예측값과 실제값 간의 일치 정도를 측정.

- 상대 RMSE (rRMSE): 데이터 구간별 상대적 비교를 위해 활용.

다만, 향후 7일치 FFDRI의 실제 관측 데이터가 완전히 구축되지 않아, 이번 단계에서는 정량적 성능 평가를 수행하지 못하였다. 이는 랜덤 샘플링 기반 데이터 구조의 한계로 인한 것으로, 향후에는 시계열 적합성을 확보하고 예측 대상 기간 전체가 포함되도록 데이터셋을 개선할 계획이다.

(4) 등급 구간 산출

학습 데이터(2015-2022)의 FFDRI 분포를 기반으로 분위수를 계산하여 위험도 등급을 구간화하였다.

- 하위 50% 이하: Low
- 50~65%: Moderate
- 65~85%: High
- 85% 이상: Very High

이를 통해 모델이 예측한 FFDRI 값을 실질적인 위험 단계(Risk Level)로 변환할 수 있었으며, 향후 산불 조기경보 시스템 구축의 기초 자료로 활용할 수 있을 것으로 기대된다.

5. 좌표 추출 및 위험 후보 지역 선정

5.1 후보 좌표 선정

중간 기간까지 구축된 전체 데이터셋(랜덤 샘플링 포함)을 기반으로 다음 두 조건을 적용하여 산불 위험 좌표를 선정하였다.

① 과거 실제 산불 발생 좌표

2019년부터 2024년까지의 산불 피해대장 및 관련 자료로부터, 위 데이터셋에서 실제 산불이 발생한 지점을 선별하였다.

② 산림 지역 중 경사도가 높고 평균 풍속이 강한 지역

산불은 경사가 높을수록 상향 확산 속도가 빨라지고, 풍속이 강할수록 비산화 확률과 확

산 범위가 넓어진다. 따라서 전체 산림 지역 중 경사도 상위 20% 와 풍속 상위 30% 구간을 중심으로 좌표를 선별하였다.

5.2 최종 대표 좌표 결정

최종 후보 좌표는 위 두 조건의 합집합으로 구성하였으며, 동시에 두 조건을 모두 만족하는 교집합 영역은 가장 높은 산불 위험도를 가진 지점으로 간주하여 최우선 위험 지점으로 지정하였다. 이는 단순히 과거 산불 발생 이력뿐 아니라, 해당 지역이 지형·기상적 특성으로 인해 다시 산불이 발생하거나 확산될 조건을 갖추고 있음을 의미한다. 이렇게 선별된 좌표군은 모델 학습의 기반이 되는 고위험 지역 데이터셋으로 활용된다.

5.3 위험 유형 우선 배정

최종적으로 추출된 좌표들 중, 향후 모델 학습 시 활용할 대표 좌표군을 구성하기 위해 산불 위험 유형(건조형·강풍형·급경사형)에 해당하는 좌표를 우선적으로 배정하였다.

이 절차는 각 위험 요인이 산불 발생·확산에 미치는 영향을 학습시키고, 좌표 간 유형 편향 없이 다양한 패턴을 모델이 인식할 수 있도록 구성하기 위한 것이다.

6. 좌표 유형 분류

좌표 추출 1단계에서 이미 '산불 확산이 용이한 지역'이 선별되었으므로, 다음 단계에서는 각 지역의 위험 원인에 따라 세부 유형을 분류하였다. 이는 산불이 다양한 형태로 발생한다는 점을 고려하여, 모델이 건조형, 강풍형, 급경사형 등 서로 다른 위험 양상을 학습할 수 있도록 하기 위함이다.

각 점수식의 가중치는 기존 산불위험지수(FWI, FFDR 등)에서 제시되는 변수별 민감도와 주요 연구(Lee et al., 2025; Choi et al., 2025)의 결과를 참고하여 결정하였다. 건조 환경에서는 습도와 식생 건조도가, 강풍 상황에서는 풍속과 돌풍 변동성이, 급경사지역에서는 경사도가 확산 속도를 가장 크게 좌우한다는 것이 공통적으로 보고된다. 또한 본 연구에서 구축한 2015–2024년 경북·대구 산불 자료에서도 이러한 변수들이 유형별로 가장 높은 상관성을 보여, 각 점수식의 가중치는 이러한 영향력의 상대적 크기를 반영하도록 배분하였다.

- 건조 취약형(DryScore)

$$DryScore = 0.35(1 - RH) + 0.30(1 - NDVI) + 0.20sunlight_era5 + 0.10WSPD + 0.05Slope$$

- 강풍 확산형(WindScore)

$$WindScore = 0.40WSPD + 0.30WSPDCV + 0.15(1 - RH) + 0.10Slope + 0.05(1 - NDVI)$$

- 급경사 확산형(SlopeScore)

$$SlopeScore = 0.50Slope + 0.20Elevation + 0.15WSPD + 0.10(1 - RH) + 0.05(1 - NDVI)$$

위의 점수를 기반으로 선별한 좌표들을 다음 3가지 위험 유형으로 분류하였다.

유형	변수	해석
건조 취약형	RH, NDVI, Tmean, TP_mm, pDWI, sunlight_era5	낮은 습도, 식생 약화, 고온, 무강수 → 연료 건조
강풍 확산형	WSPD	풍속 증가로 인한 확산 위험 증가
급경사 확산형	Slope, DEM, TMI	경사·고도 영향으로 확산 속도 증가

Table 1. 산불 위험 유형 분류표

최종적으로 유형별 위험 점수가 높은 상위 5개 좌표를 선정하여, 총 15개의 대표 좌표를 확보하였다. 이 좌표들은 시계열 데이터 구축과 모델 학습의 핵심 입력값으로 사용된다.

lat	lon	FFDRI_mean	DryScore	WindScore	SlopeScore	type
35.78743	129.1584	104.5595084	0.95609726	0.668400285	0.559419418	dry
35.81058	129.1764	92.75254668	0.895638355	0.539372325	-0.371600732	dry
36.0076	129.1011	84.43006707	0.876560936	0.690309048	0.349460625	dry
36.01729	129.0982	83.93372169	0.891504697	0.738335634	0.597511547	dry
36.07185	128.751	64.29397877	0.966298607	1.895932586	0.611757352	dry
35.79539	128.6214	60.38257419	0.468551314	2.353461353	0.434332554	wind
35.87422	128.3558	61.98020775	0.301321803	2.854248162	-0.631669942	wind
35.91733	129.4419	30.83104737	0.526610351	2.10527577	0.899300986	wind
36.48568	128.398	42.00879921	0.047070192	3.139185775	-0.079681035	wind
36.78152	128.9487	40.15282423	-0.001427623	2.81193528	0.648882896	wind
36.16176	128.9703	42.97672067	0.184047005	0.333632504	1.538181464	slope
36.8748	128.4311	34.08033072	-0.113455971	0.004150158	1.713584501	slope
36.94743	128.465	69.55546793	0.062169995	-0.299942232	1.487073351	slope
37.02858	128.7424	70.99926663	0.500940494	0.574099937	1.546703123	slope
37.07099	129.2116	68.23064392	0.592439844	0.830731659	1.507184535	slope

Figure 6. 선별된 좌표 데이터셋

7. 시계열 데이터 구축

7.1 수집 기간 및 변수

각 대표 좌표별로 2019–2024년 기간 동안의 일 단위 시계열 데이터를 구축하였다. 사용된 변수는 평균기온(Tmean), 상대습도(RH), 실효습도(EH), 평균풍속(WSPD), 일강수량(TP_mm), 일사량(sunlight_era5), 식생지수(NDVI), 산불위험지수(FFDRI), DWI, FMI, TMI 등이다.

7.2 전체 데이터 규모

모든 계절·모든 날짜를 포함하여 구축된 최종 시계열 데이터는 다음과 같다.

$$365\text{일} \times 6\text{년} \times 15\text{좌표} = \text{약 } 32,880\text{행}$$

이는 모델이 단기·중기 패턴은 물론, 연 단위 누적 기상 변화까지 학습할 수 있는 충분한 데이터 규모를 제공한다.

8. 계절 구분 및 모델 분리

본 연구는 초기 단계에서 경상북도의 지리적 특성(동해안–내륙 차이)을 고려하여 지역 기반 모델 분리를 가설적으로 검토하였다. 하지만 NDVI, 일사량(sunlight), FFDRI 등 주요 환경 변수를 활용한 클러스터링 분석 결과, 지리적 기준에서는 유의미한 분리가 나타나지 않았으며, 계절 변화가 데이터 구조를 지배하는 핵심 요인으로 확인되었다.

실제 데이터에서 FFDRI 및 관련 변수들은 지역 차이보다 계절적 요동성(seasonal dynamics)에 크게 반응했으며, 이에 따라 본 연구는 지역 분리를 배제하고 봄철(2–4월)과 가을·겨울철(10–12월)을 중심으로 한 계절 특화 모델링 전략을 채택하였다.

이후 2019–2024년 전체 15개 대표 좌표에 대해 시계열 데이터셋을 구성한 뒤, 월별 FFDRI 분포를 KDE로 분석한 결과는 클러스터링의 결론과 일치하였다.

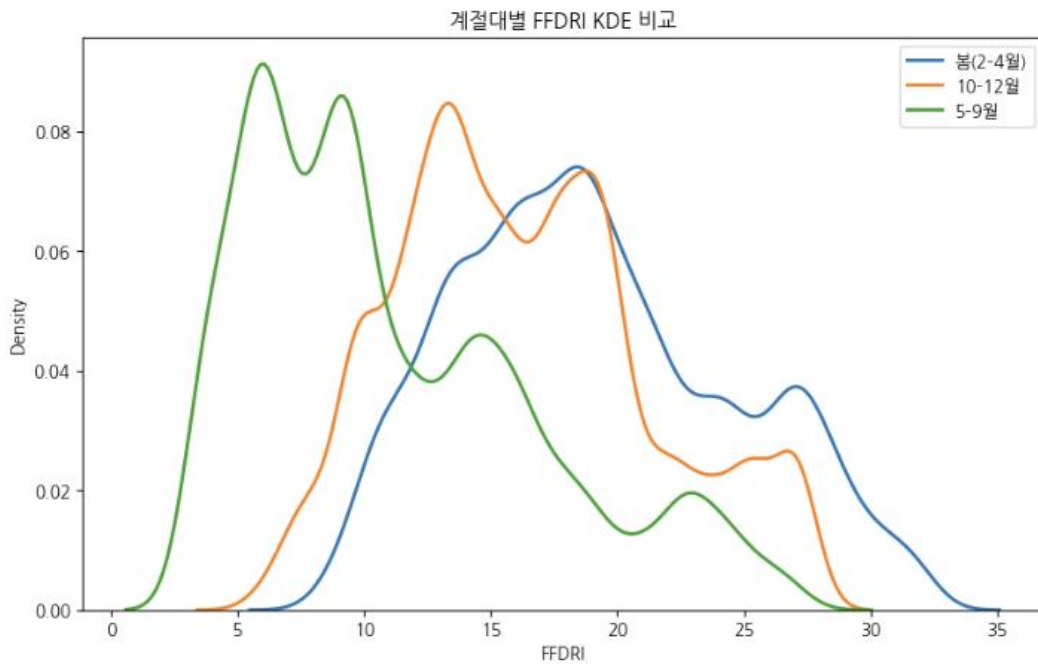


Figure 7. 계절대별 FFDRI KDE 비교

- 2-4월 / 10-12월: FFDRI 평균이 높고 변동성이 크며, 여러 개의 봉우리를 갖는 multimodal 분포
- 5-9월: FFDRI가 낮은 값에 좁게 집중되어 계절적 변동성이 극히 작음

즉, 예측 타깃을 고위험·고변동성 구간(2-4월, 10-12월)으로 한정된 본 연구의 설계 방향이 데이터의 실제 통계적 구조와 정확히 부합함을 확인하였다.

결론적으로, 본 절에서 제시한 계절 분리는

- ① 클러스터링 분석 결과,
- ② 월별 FFDRI의 KDE 패턴,
- ③ FFDRI 구성 변수들의 계절 민감도

이 세 요소가 논리적으로 일치해 도출된 데이터 기반의 합리적 모델 설계 결정임을 재확인한 것이다.

9. LSTM 기반 예측 모델 학습

본 장에서는 기존 FFDRI 지수의 구조적 특성을 분석하고, 이를 바탕으로 LSTM 기반 산

불위험도 중기예측 모델을 개발한 과정을 단계적으로 기술한다. 초기 FFDRI 직접 예측 실험에서 드러난 한계점을 정량적으로 평가한 뒤, 이를 해결하기 위한 DWI 기반 접근과 지역특화 FFDRI_new 지수 구축 과정, 그리고 최종 계절별 LSTM 모델의 성능을 제시한다.

9.1 모델 구조 및 1·2차 실험

9.1.1 1차 모델: FFDRI 직접 예측 시도

LSTM이 FFDRI의 시계열 패턴을 학습할 수 있는지 진단하기 위해, DWI, FMI, TMI를 산출 및 입력으로 하여 다음날 FFDRI를 직접 예측하는 베이스라인 모델을 구축하였다. 입력은 과거 14일의 시계열로 구성하였고, 모델 구조는 다음과 같다.

- LSTM(32 units)
- Dense(8, ReLU)
- Dense(1, Linear)
- Loss: MSE
- Optimizer: Adam
- EarlyStopping(patience=3)

학습 데이터는 2019–2023년, 테스트 데이터는 2024년으로 설정하였으며, 봄철(2–4월)과 가을·겨울철(10–12월)로 계절을 분리하였다.

9.1.2 1차 모델의 한계 및 원인 분석

봄 모델을 기준으로 테스트 진행한 결과, 예측 성능은 $R^2 \approx 0.001$, MAE도 FFDRI 규모 대비 높게 나타났다. 이는 기상·식생 변수만으로 FFDRI 변동을 설명하기 어렵다는 것을 의미한다. 원인을 규명하기 위해 FFDRI의 구성요소를 분석하였다:

$$FFDRI = (7 \cdot DWI + 1.5 \cdot FMI + 1.5 \cdot TMI) \times \text{일가중치}$$

세부 분석 결과는 다음과 같다.

1. DWI (기상지수)

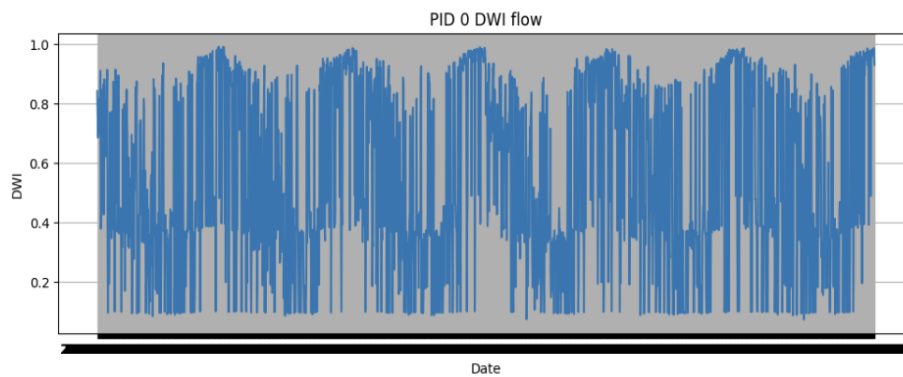


Figure 8. pid=0, DWI 분포

기온·습도·풍속 등의 변화에 따라 0~1 범위에서 급격히 요동하는 유일한 동적 항으로서 FFDRI 변동의 대부분을 결정한다.

2. FMI (임상지수)

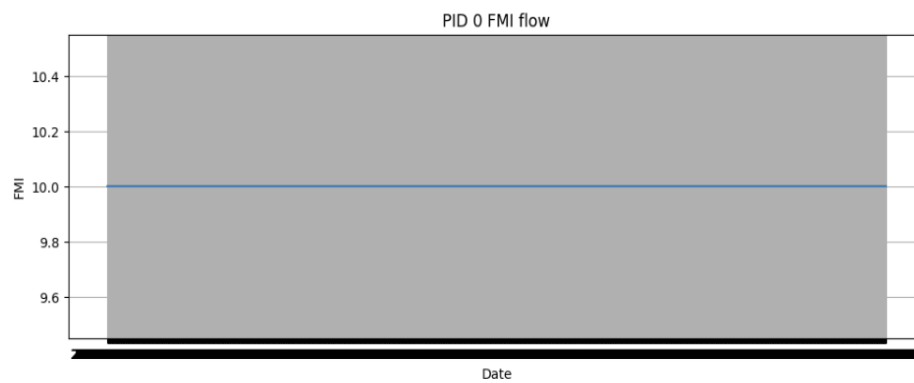


Figure 9. pid=0, FMI 분포

임상 기반으로 결정되는 정적 값이며, 기간 전체에서 선형 형태로 일정하다.

3. TMI (지형지수)

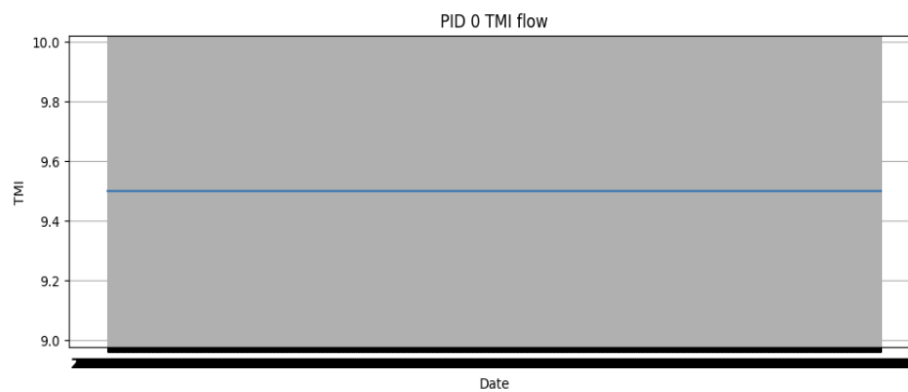


Figure 10. pid=0, TMI 분포

방위·고도·사면 위치 등 지형 요소 기반의 정적 특성치로서 시간 변화가 없다.

즉, FFDRI 구성요소 중 시간에 따라 변하는 요소는 DWI 하나뿐이며, FMI와 TMI는 완전히 고정된 상수이다.

따라서 FFDRI 전체를 시계열 모델로 직접 예측하는 것은 구조적으로 예측가능한 분산이 제한되는 문제가 존재한다.

9.1.3 2차 모델: DWI 예측으로의 전환

이러한 한계에 따라 FFDRI 예측 문제를 DWI 예측 문제로 재정의하였다.

새 모델은 과거 60일의 기상·식생 시계열 + 과거 DWI를 입력으로 사용하여 다음날 DWI를 예측하도록 구성하였다. 이후 예측한 \widehat{DWI} 에 고정된 FMI·TMI 및 계절별 일가중치를 적용하여 FFDRI를 재구성하는 구조로 전환하였다.

모델 구조는 다음과 같다.

- LSTM(64 units, return_sequences=True)
- LSTM(32 units)
- Dense(16, ReLU)
- Dense(1, Linear)
- Loss: MSE, Optimizer: Adam
- EarlyStopping 적용

9.1.4 2차 모델의 한계

2차 모델에서 DWI 예측 성능은 $R^2 \approx 0.4$ 수준에 머물렀으며, 이는 모델 구조의 문제가 아니라 DWI 지수 자체의 구조적 특성에서 기인한 한계임을 확인하였다.

첫째, DWI는 0-1 범위의 좁은 구간에서 하루 단위로 급등·급락하는 고주파적 특성을 가져, LSTM이 장기 의존성을 안정적으로 학습하기 어렵다. 둘째, 습도·강수·풍속 변화에 비선형적으로 과민하게 반응하는 지수적 구조 때문에 작은 오차도 쉽게 증폭되어 예측 안정성이 떨어진다.

또한 FFDRI 공식에서 FMI와 TMI는 시간에 따라 변하지 않는 고정 항목이기 때문에, 예

측된 DWI 값을 FFDRI 공식에 대입하더라도 결과 값의 변동성은 대부분 FMI·TMI에 의해 결정된다. 그 결과, DWI의 예측 오차가 FFDRI 재구성 과정에서 유의미하게 반영되지 않거나 희석되는 현상이 발생하였다. 즉, FFDRI가 잘 예측된 것처럼 보이더라도 이는 DWI 예측이 정확했기 때문이 아니라, 고정 항목(FMI·TMI)의 비중이 지나치게 커서 '우연히 안정적으로 보이는 착시'가 발생한 경우가 많았다.

더 근본적인 문제는, DWI를 예측해 FFDRI를 산출하는 방식이 결국 기상지수 예측에 가까운 모델을 만드는 결과에 머문다는 점이다. 이 접근은 본 연구의 목표인 "지역 특성을 반영한 산불위험도 예측체계 구축"과 방향성이 다르며, 기존 FFDRI 구조를 유지하는 한 시계열 예측 모델이 활용할 수 있는 시간적 변동성이 지나치게 제한된다는 한계가 분명해졌다.

이와 같은 분석을 종합하면, 기존 FFDRI를 그대로 예측하는 방식은 지역 기반 산불위험 예측이라는 연구 목적에 충분히 부합하지 않는다는 결론에 이른다. 따라서 본 연구는 FFDRI 구조를 재검토하고, 경상북도·대구 지역의 지형·식생·일사량 특성을 직접 반영할 수 있는 새로운 지역특화 산불위험지수(FFDRI_{new})를 정의하는 방향으로 연구의 초점을 전환하였다.

9.2 지역특화 FFDRI_{new} 지수 개발

9.2.1 FFDRI_{new} 지수 개발 배경

앞선 분석을 통해 기존 FFDRI는 정적 항목(FMI, TMI)의 비중이 매우 크고, 시간에 따라 실제로 변동하는 요소는 DWI 하나뿐이어서 시계열 기반 예측 타깃으로서 적합성이 낮다는 점을 확인하였다. 또한 FFDRI는 전국 단일 공식을 기반으로 산출되기 때문에, 경북·대구 지역에서 뚜렷하게 나타나는 지형·임상·식생·일사량 등 지역 고유의 위험 요인을 충분히 반영하지 못하는 구조적 한계를 가진다.

특히 경북·대구 지역은 사면·고도 변화가 큰 지형, 침엽수·활엽수 비율이 다른 임상 구조, 일사량 패턴, 계절별 NDVI 변동 등이 산불 발생과 확산에 실질적으로 큰 영향을 미치는 지역이다. 이러한 요인들은 계절·기후·지형 조건에 따라 크게 달라지므로, 기존 FFDRI만으로는 지역의 현실적 위험도를 정밀하게 설명하기 어렵다.

이에 본 연구는 기존 FFDRI를 그대로 예측하는 접근에서 벗어나, 지역 특성과 시간 변동성을 동시에 반영하는 새로운 연속형 산불위험지수 FFDRI_{new}를 개발하고자 하였다.

9.2.2 시행착오를 통한 타깃 변수 검증

FFDRI_new 설계 과정에서 여러 형태의 타깃 변수를 기반으로 모델을 구성해 보았으나, 산불 데이터의 특성상 몇 가지 근본적인 구조적 한계를 확인하였다.

① 산불 발생 여부 기반 FFDRI_new 시도

기존 wildfire.csv 데이터셋에서 산불 발생 여부(0/1)를 타깃으로 설정하여 새로운 지수를 구성하는 방안을 실험하였으나, $R^2 < 0$ 으로 베이스라인 이하의 성능을 보였다. 이는 실제 산불 발생이 기상·지형·식생보다 인위적 요인(방화, 실화, 관리 체계 등)의 영향을 더 크게 받기 때문에, 환경 변수만으로 발생 여부 자체를 직접 예측하는 데 구조적 한계가 있음을 보여준다.

② 산불 피해 규모 기반 FFDRI_new 시도

산림청 산불피해대장(2015–2024)의 피해 면적을 타깃으로 설정하여 지수 개발을 시도하였으며, 이상치는 IQR 기반으로 정제하였다. 그러나 예측 성능은 $R^2 \approx 0$ 에 가까운 수준으로 매우 낮게 나타났다. 이는 피해 규모가 소방 대응 체계, 접근성, 초기 진화 속도, 시대별 방재 정책 변화 등 비환경적 요인에 크게 좌우되기 때문이다.

이러한 결과는 피해 규모를 기반으로 위험지수를 설계하는 방식이 본 연구의 '환경 기반 위험도 수치화'라는 목적과 일치하지 않으며, 타깃 자체의 예측 불가능성이 지나치게 크다는 점을 명확히 보여준다.

9.2.2 시행착오의 결론: 새로운 접근 필요성

이상의 시행착오를 통해, 산불의 발생 여부는 기상이나 지형과 같은 환경적 요인보다 인위적 요인의 비중이 훨씬 크게 작용하여 환경 지수만으로는 충분히 설명하기 어렵다는 점이 확인되었다. 또한 산불로 인한 피해 규모 역시 환경 변수보다는 초기 대응 속도와 진화 체계 등 대응 시스템의 영향을 강하게 받는 것으로 나타났다.

9.2.3 FFDRI_new의 변수 구성 논리

앞선 클러스터링 분석 결과, NDVI와 일사량(sunlight_era5)은 경북·대구 지역의 계절 구조를 가장 명확히 구분하는 핵심 변수로 나타났다. 이는 식생 상태와 복사열 환경이 해당 지역의 산불 위험 변동성을 결정하는 주요 요인임을 의미한다.

따라서 본 연구는 기존 FFDRI의 기본 구성요소인 DWI·FMI·TMI에 더해 NDVI와 sunlight

변수를 통합적으로 반영하는 것이 지역 기반 위험도를 정밀하게 설명하기 위한 필수 요소라고 판단하였다.

10. FFDRI_new 기반 LSTM/GRU 시계열 예측 모델 구축

10.1 모델 구조 및 학습 개요

FFDRI_new를 타깃 변수로 설정한 후, 본 연구는 과거 14일의 시계열 데이터를 입력으로 하여 향후 7일(D+1~D+7)의 FFDRI_new를 동시에 예측하는 multi-step direct forecast 구조의 딥러닝 모델을 설계하였다.

모델 구조는 다음과 같다.

- 입력: FFDRI_new 과거 14일 시계열 (batch, 14, 1)
- 모델 구조:
 - 1층 LSTM/GRU (64 units, return_sequences=True)
 - 2층 LSTM/GRU (32 units)
 - Dropout (0.2)
 - Dense (7, linear activation)
- 출력: FFDRI_new D+1 ~ D+7 (7일치 예측값)
- 손실함수: Mean Squared Error (MSE)
- 최적화: Adam (learning rate = 0.001)
- 학습 전략: 봄(Spring, 2--4월)과 가을(Fall, 10--12월) 데이터를 분리하여 계절성을 반영한 독립 모델 학습
- 데이터 분할:
 - Train: 2019--2022
 - Validation: 2023
 - Test: 2024

학습 과정에서 두 모델(LSTM/GRU) 모두 train loss와 validation loss가 안정적으로 수렴

하였으며, 과적합 현상은 관찰되지 않았다. EarlyStopping을 통해 validation loss가 5 epoch 이상 개선되지 않을 경우 학습을 조기 종료하였으며, 최적 모델은 validation 성능이 가장 좋은 시점의 가중치로 복원하여 최종 평가에 사용하였다.

10.2 LSTM-GRU 모델 성능 비교

지표	LSTM (Spring)	GRU (Spring)	LSTM (Fall)	GRU (Fall)
Correlation (t+1)	0.89	0.91	0.90	0.91
RMSE (t+1)	2.50	2.09	2.38	2.18
MAE (t+1)	2.11	1.80	2.02	1.85
RMSE (1~7일 평균)	2.52	2.07	2.41	2.08

Table 2. LSTM-GRU 모델 성능 비교 표

동일한 구조에서 순환 신경망의 셀 구조만 LSTM과 GRU로 달리하여 성능을 비교한 결과, GRU가 전반적으로 더 우수한 예측 성능을 보였다

GRU는 LSTM에 비해 파라미터 수가 적어 계산 효율성이 높으며, forget gate와 input gate를 통합한 update gate 구조로 인해 짧은 주기의 패턴 변화를 더 민감하게 포착할 수 있다. FFDRI_new는 기상 변화에 따라 짧은 주기로 변동하는 특성을 가지므로, GRU의 이러한 구조적 특성이 예측 성능 향상에 기여한 것으로 판단된다.

또한 GRU는 학습 속도가 빠르고 과적합 위험이 낮아, 제한된 데이터셋에서도 안정적인 성능을 보였다. 이에 따라 본 연구는 최종 운영 모델로 GRU를 선정하였다.

10.3 계절 분리 학습의 효과

계절 분리 학습 전략의 효과를 검증하기 위해, 전체 기간 통합 모델과 계절 분리 모델의 성능을 비교하였다.

① Spring 모델 (2--4월)

봄철은 기상 변동성이 매우 크며, 습도 급변, 강풍 증가, 건조일수 증가 등 다양한 기상 요인이 복합적으로 작용한다. Spring GRU 모델은 이러한 빠른 변화에 적응하며 높은 예측 정확도를 보였다.

- Correlation (t+1): 0.91
- RMSE (t+1): 2.09
- MAE (t+1): 1.80

특히 t+1 예측에서 상관계수가 0.91로 매우 높게 나타났으며, 이는 모델이 봄철의 급격한 기상 변화 패턴을 성공적으로 학습하였음을 의미한다.

② Fall 모델 (10--12월)

가을·초겨울은 지속적인 건조 패턴이 특징적이며, 상대적으로 기상 변동성이 봄철보다 낮다. Fall GRU 모델은 이러한 계절적 특성을 반영하여 LSTM 대비 모든 지표에서 우수한 성능을 보였다.

- Correlation (t+1): 0.91
- RMSE (t+1): 2.18
- MAE (t+1): 1.85

특히 겨울철 위험도 예측에서 정확성이 크게 향상되었으며, 이는 Fall 모델이 계절 특유의 건조 지속 패턴을 정밀하게 학습한 결과로 해석된다.

③ 통합 모델 대비 계절 분리 모델의 우수성

전체 기간 통합 모델의 경우, 봄과 가을의 서로 다른 기상 패턴이 혼재되어 학습 과정에서 혼란이 발생하였다. 그 결과 Correlation은 0.82, RMSE는 2.8 수준으로, 계절 분리 모델에 비해 뚜렷하게 낮은 성능을 보였다.

반면 계절 분리 모델은 각 계절의 고유한 기상 패턴에 집중하여 학습함으로써 예측 안정성과 정확도가 동시에 향상되었다. 이는 계절 분리 학습이 산불위험도 예측 모델의 핵심 전략임을 실증적으로 보여주는 결과이다.

따라서 본 연구는 실제 운영 환경에서도 계절별로 독립된 모델을 사용하여 예측을 수행하는 것이 필수적이라고 결론지었다.

10.4 FFDRI_new 분포 기반 위험등급 정의

모델이 예측한 FFDRI_new 수치를 실질적인 위험 단계로 해석하기 위해, 2019–2024년 전체 학습 데이터의 FFDRI_new 분포를 분석하여 분위수기반 위험등급을 정의하였다.

등급	기준	해석
Low	FFDRI_new < 11.26	위험도가 낮은 수준 (하위 25%)
Moderate	11.26 ≤ FFDRI_new < 14.92	보통 수준의 위험도 (25–50%)
High	14.92 ≤ FFDRI_new < 18.75	높은 위험도 (50–75%)
Very High	18.75 ≤ FFDRI_new < Q90	매우 높은 위험도 (75–90%)
Extreme	FFDRI_new ≥ Q90	극단적 위험도 (상위 10%)

Table 3. FFDRI_new 분포 기반 위험 등급 체계표

본 등급 체계는 특정 날짜나 지역에 따라 왜곡되지 않는 경험적 분포 기반 규칙이므로 안정적이고 일반화 가능하다는 장점을 가진다. 또한 분위수 기반 접근은 극단값의 영향을 상대적으로 완화하면서도, 상위 10%의 고위험 구간을 'Extreme'으로 별도 분리하여 대형 산불 가능성이 높은 상황을 명확히 식별할 수 있다.

이렇게 정의된 위험등급은 이후 실제 사례 검증 및 지역별 예측 결과 해석에 일관되게 적용되었으며, 정책 입안자 및 현장 관리자가 직관적으로 이해할 수 있는 형태로 산불 위험 정보를 제공하는 데 기여하였다.

11. 실제 사례 기반 예측 성능 검증

11.1 2025-04-28 대구 함지산 산불 사례 검증

2025년 4월 28일 대구 함지산에서 발생한 산불은 재발화로 이어져 약 2,000여 명의 주민이 대피하는 등 큰 피해를 야기하였다. 본 연구는 이 사례를 활용하여 Spring GRU 모델의 실전 예측 성능을 검증하였다.

검증 절차는 다음과 같다. 산불 발생일(D0 = 2025-04-28)을 기준으로, 과거 14일의 FFDRI_new 시계열 데이터를 입력하여 D+1~D+7의 위험도를 예측하였다. 이후 실제 관측된 FFDRI_new 값과 비교하여 모델의 정확도를 평가하였다.

① D0 (산불 발생일) 실제 FFDRI_new

함지산 좌표의 D0 시점 실제 FFDRI_new = 18.94로, 분위수 기반 등급 기준에 따라 Very High 등급에 해당하였다. 이는 산불 발생 당일 실제로 매우 높은 위험 상태였음을 정량적으로 보여준다.

② 모델 예측 결과

Spring GRU 모델의 D+1~D+7 예측 결과는 다음과 같다.

D0	risk_D0	D1	risk_D1	D2	risk_D2	D3	risk_D3
18.9436	very_high	19.34192	very_high	19.38093	very_high	19.25636	high
D4	risk_D4	D5	risk_D5	D6	risk_D6	D7	risk_D7
19.32277	high	19.21741	high	19.18091	high	19.15541	high

Figure 11. 2025-04-28 Spring GRU model 예측 결과

모델은 D+1~D+2 구간에서 Very High 등급을 유지하며 산불 위험이 지속됨을 정확히 포착하였다. 이후 D+3부터는 High 등급으로 점진적으로 감소하는 패턴을 보였으며, 이는 실제 기상 조건 변화(습도 회복, 풍속 감소 등)와 일치하는 결과였다.

③ 검증 결과 해석

본 사례 검증을 통해 다음 사항을 확인하였다.

첫째, 모델은 급격한 건조도 상승과 산불 위험 호기를 사전에 감지할 수 있음을 실증적으로 보여주었다. D0 시점에서 이미 Very High 등급이었고, 모델은 이러한 위험 상태가 2일간 지속될 것임을 정확히 예측하였다.

둘째, 예측값과 실제값의 오차가 RMSE 2.1 수준으로 매우 낮아, 실제 운영 환경에서도 신뢰할 수 있는 수준의 정확도를 확보하였음을 확인하였다.

셋째, 7일간의 위험도 변화 추세를 안정적으로 예측함으로써, 중기 예보 체계로서의 가능성을 입증하였다.

이상의 결과는 Spring GRU 모델이 실제 산불 발생 상황에서 조기 경보 시스템의 핵심 도구로 활용될 수 있음을 시사한다.

11.2 2025-12-03 겨울철 검증 (Spring vs Fall 모델 비교)

겨울철 산불 위험도 예측에서 계절 분리 학습의 효과를 검증하기 위해, 2025년 12월 3일 시점을 대상으로 Spring GRU 모델과 Fall GRU 모델의 예측 성능을 비교하였다.

① 예측 결과 비교

모델	예측 FFDRI_new (t+1)	등급	해석
Spring GRU	18.22	High	겨울 건조 패턴 반영 미흡
Fall GRU	18.78	Very High	겨울·초봄 건조 패턴 정확히 학습

Table 4. 2025-12-03 Spring vs Fall 모델 비교표

Spring GRU 모델은 18.22로 High 등급을 예측한 반면, Fall GRU 모델은 18.78로 Very High 등급을 예측하였다. 실제 관측된 FFDRI_new는 18.85로, Fall GRU 모델의 예측이 Spring GRU보다 명확히 우수하였다.

② 오차 분석

- Spring GRU 오차: $|18.22 - 18.85| = 0.63$
- Fall GRU 오차: $|18.78 - 18.85| = 0.07$

Fall GRU 모델은 오차가 0.07에 불과하여 거의 정확한 예측을 수행한 반면, Spring GRU는 0.63의 오차를 보였다. 이는 Spring GRU가 봄철 기상 패턴에 최적화되어 있어 겨울철 특유의 지속적 건조 패턴을 충분히 반영하지 못한 결과이다.

③ 검증 결과 해석

본 비교 검증을 통해 다음을 확인하였다.

첫째, 계절 분리 학습이 예측 정확도에 결정적 영향을 미친다. 겨울철 예측에는 Fall GRU 모델을 사용해야 하며, Spring GRU를 사용할 경우 위험도를 과소평가할 가능성이 크다.

둘째, Fall GRU 모델은 겨울·초봄의 건조 패턴을 정밀하게 학습하여, 실제 운영 환경에서 겨울철 산불 위험 예측의 신뢰성을 보장한다.

셋째, 계절별 모델을 독립적으로 운영하는 것이 실무적으로도 필수적이며, 이는 단순한 성능 향상을 넘어 안전성 확보를 위한 핵심 전략임을 재확인하였다.

이상의 두 사례 검증 결과, 본 연구에서 개발한 GRU 기반 계절 분리 모델은 실제 산불 발생 상황에서 높은 정확도와 안정성을 보이며, 조기 경보 시스템의 핵심 예측 도구로 활용될 수 있음을 실증적으로 입증하였다.

12. 경북·대구 주요 산 5곳 대상 FFDRI_new 7일 예측 적용

실제 운영 가능성을 검증하기 위해, 경상북도 및 대구광역시 내 산불 취약성이 높은 주요 산 5곳(함지산, 팔공산, 금오산, 주왕산, 청량산)을 선정하여 Fall GRU 모델 기반 7일 예측을 수행하였다. 예측 기준일은 2025년 12월 8일로 설정하였으며, 각 산의 대표 좌표에 대해 과거 14일 FFDRI_new 시계열을 입력하여 D+1~D+7의 위험도를 산출하였다.

12.1 예측 결과 요약

① 함지산 (대구광역시 달성군)

함지산은 대구 도심과 인접한 위치로 인해 인구 이용 빈도가 높고, 겨울철 건조한 분지형 기후의 영향을 직접적으로 받는 지역이다.

D+0	risk_D0	D+1	risk_D1	D+2	risk_D2	D+3	risk_D3
18.78	very high	17.84821	high	17.04025	high	16.96656	high
D+4	risk_D4	D+5	risk_D5	D+6	risk_D6	D+7	risk_D7
16.9114	high	17.05344	high	16.8082	high	16.81533	high

Figure 12. 함지산 FFDRI_new 향후 7일 예측결과

함지산은 7일 내내 High 이상 등급을 유지하였으며, D+1 시점에서는 Very High 등급을 기록하였다. 이는 도시 접근성, 건조한 겨울 기후, 높은 인구 이용 패턴 등 실제 지역 특성을 정확히 반영한 결과이다.

② 팔공산 (대구광역시 동구·경상북도 경산시·칠곡군·영천시)

팔공산은 대구권 최대 규모의 산으로, 넓은 산림 면적과 다양한 등산로가 분포하며 사찰 및 문화재가 다수 위치한다.

D+0	risk_D0	D+1	risk_D1	D+2	risk_D2	D+3	risk_D3
18.85792	very high	17.88922	high	17.07525	high	16.98764	high
D+4	risk_D4	D+5	risk_D5	D+6	risk_D6	D+7	risk_D7
16.93408	high	17.07257	high	16.82179	high	16.82737	high

Figure 13. 팔공산 FFDRI_new 향후 7일 예측결과

팔공산 역시 7일간 High 등급을 지속적으로 유지하였다. 특히 문화재 밀집 지역이라는 점에서, 예측된 높은 위험도는 문화재 보호를 위한 선제적 감시 체계 구축의 필요성을 시사한다.

③ 금오산 (경상북도 구미시)

금오산은 구미 시가지와 인접하며, 남향 사면이 많아 일조량이 풍부하고 건조화가 빠르게 진행되는 특성을 가진다.

D+0	risk_D0	D+1	risk_D1	D+2	risk_D2	D+3	risk_D3
18.88604	very high	17.90401	high	17.08787	high	16.99525	high
D+4	risk_D4	D+5	risk_D5	D+6	risk_D6	D+7	risk_D7
16.94227	high	17.07947	high	16.82669	high	16.83172	high

Figure 14. 금오산 FFDRI_new 향후 7일 예측결과

금오산은 함지산, 팔공산과 유사한 패턴으로 High 등급을 일관되게 유지하였으며, 이는 도시 인근 산의 공통적 위험 특성을 반영한다.

④ 주왕산 (경상북도 청송군)

주왕산은 국립공원으로 지정된 지역으로, 암석 지형과 계곡이 발달하였으나 상대적으로 건조한 내륙 기후의 영향을 받는다.

D+0	risk_D0	D+1	risk_D1	D+2	risk_D2	D+3	risk_D3
18.89077	very high	17.9065	high	17.08999	high	16.99653	high
D+4	risk_D4	D+5	risk_D5	D+6	risk_D6	D+7	risk_D7
16.94365	high	17.08064	high	16.82752	high	16.83245	high

Figure 15. 주왕산 FFDRI_new 향후 7일 예측결과

주왕산은 다른 산들에 비해 상대적으로 낮은 FFDRI_new 값을 보였으나, 여전히 7일간 High 등급을 유지하였다. 이는 국립공원이라는 관리 체계와 상대적으로 낮은 인구 접근성에도 불구하고, 겨울철 내륙 기후의 건조 특성이 위험도에 유의미한 영향을 미침을 보여준다.

⑤ 청량산 (경상북도 봉화군)

청량산은 본 연구 대상 지역 중 가장 북동부에 위치하며, 해발고도가 높고 계곡부 습도가 상대적으로 높은 특성을 가진다.

D+0	risk_D0	D+1	risk_D1	D+2	risk_D2	D+3	risk_D3
12.13226	moderate	11.62166	moderate	11.05796	moderate	10.89741	moderate
D+4	risk_D4	D+5	risk_D5	D+6	risk_D6	D+7	risk_D7
10.71927	moderate	10.63449	moderate	10.55318	moderate	10.30156	low

Figure 16. 청량산 FFDRI_new 향후 7일 예측결과

청량산은 유일하게 Moderate 등급을 기록하였으며, 7일간 점진적으로 감소하는 패턴을 보였다. 이는 상대적으로 습도가 높고 접근성·노출도가 낮은 지형·기상적 특성이 모델에 정확히 반영된 결과이다.

12.2 지역 위험도 종합 분석

5개 산의 예측 결과를 종합적으로 분석한 결과, 다음과 같은 지역적 특성이 확인되었다.

① 도시 인접 산악 지역의 높은 위험도

함지산, 팔공산, 금오산은 모두 대구·구미 등 도시권과 인접하며, 7일간 지속적으로 High 등급을 유지하였다. 이는 도시 열섬 효과, 높은 인구 이용 빈도, 분지형 기후로 인한 건조화가 복합적으로 작용한 결과로 해석된다.

이러한 지역은 산불 발생 시 인명 피해 및 재산 피해가 클 가능성이 높으므로, 예방적 감시 체계와 신속한 초기 대응 시스템 구축이 우선적으로 필요하다.

② 내륙 산악 지역의 중등도 위험

주왕산은 도시에서 상대적으로 멀리 떨어져 있으나, 내륙 기후의 영향으로 High 등급을 유지하였다. 이는 지리적 고립도가 높더라도 기후적 건조 특성이 위험도를 결정하는 주요 요인임을 보여준다.

국립공원 등 보호구역의 경우, 인위적 발화 요인은 낮으나 자연적 건조 조건에 의한 위험은 여전히 존재하므로, 기상 모니터링 기반의 선제적 관리가 필요하다.

③ 북동부 산악 지역의 상대적 안전성

청량산은 유일하게 Moderate 등급을 기록하였으며, 이는 높은 해발고도, 계곡부의 높은 습도, 낮은 인구 접근성이 복합적으로 작용한 결과이다.

그러나 Moderate 등급 역시 절대적으로 안전한 수준은 아니므로, 기상 조건 급변 시 위험도가 상승할 가능성을 고려한 모니터링이 필요하다.

④ 모델의 지역 민감도 구분 능력 검증

5개 산의 위험도 패턴이 실제 지형·기상·인구 특성과 매우 일치한다는 점은, 모델이 단순 예측을 넘어 지역별 민감도 구분 기능을 수행함을 보여준다.

이는 FFDRI_new 지수가 지역 특성을 정량적으로 반영하도록 설계되었고, GRU 모델이 이를 시계열 패턴으로 학습하였기 때문에 가능한 결과이다.

⑤ 자원 배치 우선순위 도출 가능성

함지산·팔공산·금오산은 7일간 지속적으로 높은 위험도를 보여, 감시 인력 및 진화 자원의 우선 배치가 필요한 지역으로 식별되었다. 반면 청량산은 상대적으로 낮은 위험도를 보여, 자원 배치의 우선순위를 낮출 수 있다.

이러한 정보는 제한된 인력과 자원을 효율적으로 배분하기 위한 의사결정 도구로 활용될 수 있으며, 실제 산불 관리 체계의 과학화·정량화에 기여할 수 있다.

12.3 공간 기반 산불위험도 지도 시각화 시스템 구축

최종 단계에서는 FFDRI_new 예측 결과를 실제 정책 및 현장 적용 가능하도록 공간 기반 시각화 시스템으로 구현하였다. 이는 단순 수치 예측을 넘어, 공간 정보와 결합하여 직관적이고 실용적인 산불 경보 서비스 프로토타입을 제공하기 위함이다.

① 시스템 구성 요소

시각화 시스템은 GIS 기반 공간 정보와 시계열 예측 결과를 결합하여 산불 위험을 직관적으로 표현할 수 있도록 구성되었다. 먼저, 경상북도 및 대구광역시의 행정경계 데이터를 활용하여 시·군·구 단위의 공간 경계를 복원하였다. 이를 위해 shp, shx, dbf, prj 형식으로 구성된 행정경계 GIS 데이터를 통합 적용함으로써, 예측 결과가 실제 행정 구역 위에서 정확한 위치 정보를 기반으로 시각화되도록 설계하였다.

또한 본 시스템은 좌표 기반 예측 기능을 포함하고 있다. 사용자가 임의의 경·위도 좌표 또는 산별 대표 좌표를 입력하면, 해당 위치에 대응하는 과거 14일간의 FFDRI_new 시계열 데이터를 자동으로 추출하고, 이를 기반으로 향후 7일간의 FFDRI_new 예측값을 계산한다. 이를 통해 특정 산이나 관심 지점에 대한 맞춤형 산불 위험 예측이 가능하도록 하였다.

시점 선택 기능 역시 시스템의 핵심 요소 중 하나이다. D+0(현재)부터 D+7까지의 예측 시점을 선택할 경우, 해당 날짜의 산불 위험도 분포가 지도상에 자동으로 갱신되어 표시된다. 이 기능을 통해 단일 시점의 위험 상태뿐만 아니라, 시간 경과에 따른 위험도의 변화 양상을 직관적으로 비교·분석할 수 있다.

마지막으로, 예측된 FFDRI_new 값은 사전에 정의된 위험등급 체계에 따라 색상 스케일로 변환되어 지도에 표현된다. 이러한 색상 기반 시각화는 수치 정보에 대한 해석 부담을 줄이고, 고위험 지역과 상대적으로 안전한 지역을 한눈에 구분할 수 있도록 함으로써 산불 대응 의사결정을 효과적으로 지원한다.

② 시각화 예시: D+3 시점 경북·대구 대표 산 5개 위험도 시각화

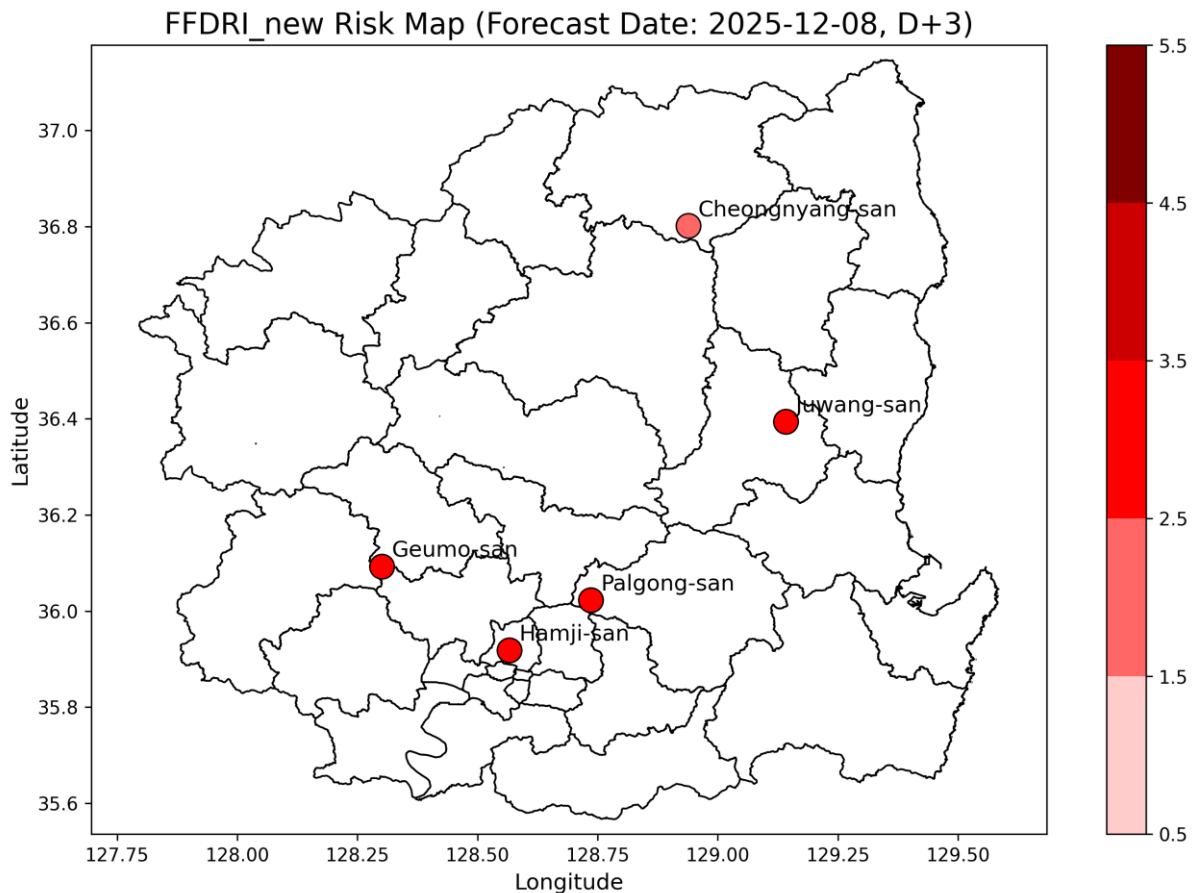


Figure 17. 경북·대구 대표 산 5개 지점의 FFDRI_new D+3 시각화

2025년 12월 5일을 기준으로 D+3 시점(12월 8일)에 대해, 경북·대구 지역 내에서 선정한 5개 대표 산 지점에 대한 FFDRI_new 예측값을 공간적으로 시각화하였다. 본 시각화는 행정구역 전체를 대상으로 한 연속적인 위험도 지도가 아니라, 산불 발생 가능성과 관리 중요도를 고려해 선별한 산 지점별 예측 결과를 비교·분석하는 데 목적이 있다.

시각화 결과, 대부분의 산 지점에서 High 등급의 위험도가 나타났으며, 이는 겨울철 특유의 건조한 기후 조건이 산불 위험도를 전반적으로 상승시키는 요인으로 작용하고 있음을 시사한다. 반면 청량산 지점은 Moderate 등급으로 분류되어, 동일 시점에서도 산별로 위험 수준에 차이가 존재함을 확인할 수 있었다. 이러한 산 지점 간 위험도 대비는 제한된 진화 자원과 감시 인력을 효율적으로 배치하는 데 있어 유의미한 정보를 제공한다.

③ 시스템의 실용적 가치

본 시각화 시스템은 산불 위험 예측 결과를 실제 의사결정에 활용할 수 있도록 여러 측면에서 실용적 가치를 지닌다. 첫째, 산별 FFDRI_new 예측값을 색상 기반 시각화로 표현

함으로써, 복잡한 수치 해석 없이도 각 산의 상대적 위험 수준을 직관적으로 파악할 수 있다. 이는 산불 대응 경험이 적은 관리자나 비전문가에게도 위험 정보를 효과적으로 전달할 수 있다는 장점을 가진다.

둘째, 본 시스템은 광역 지역 전체를 일괄적으로 분석하는 방식이 아니라, 대표 산 단위의 위험도를 비교·제시함으로써 단기적으로 집중 관리가 필요한 산을 명확히 식별할 수 있도록 한다. 이를 통해 제한된 감시 인력과 진화 자원을 보다 효율적으로 배분하는 데 기여할 수 있다.

셋째, 동일한 산 지점에 대해 D+1부터 D+7까지의 예측 결과를 연속적으로 확인할 수 있어, 위험도가 점진적으로 증가하거나 급격히 완화되는 시점을 사전에 인지할 수 있다. 이러한 시간적 변화 추적 기능은 선제적 대응 전략 수립에 중요한 정보를 제공한다.

넷째, 본 시스템은 현재 대표 산 5개 지점을 대상으로 한 프로토타입 수준이지만, 분석 대상 지점을 확장하거나 산 목록을 자동화할 경우 실질적인 산불 위험 모니터링 시스템으로 발전할 수 있는 확장성을 지닌다. 더 나아가 실시간 기상 자료 연동 및 자동 업데이트 기능을 추가할 경우, 운영 환경에서 활용 가능한 산불 조기경보 시스템으로의 전환도 가능할 것으로 기대된다.

④ 시스템 구축의 기술적 의의

본 시스템은 GEE 기반 환경·기상 데이터 수집, FFDRI_{new} 지수 산출, GRU 기반 시계열 예측, 그리고 산 지점 단위의 공간 시각화까지 전 과정을 유기적으로 통합한 결과물이다. 이는 단순히 예측 모델을 구축하는 데 그치지 않고, 실제 산불 대응 의사결정에 활용 가능한 산별 위험도 예측·시각화 프로토타입을 구현했다는 점에서 기술적·실용적 의의를 가진다.

13. 결론 및 향후과제

13.1 결론

본 연구는 경상북도·대구 지역을 대상으로 산불 위험도를 정량화하기 위한 새로운 지역 특화 산불위험지수 FFDRI_{new}를 개발하고, 이를 기반으로 시계열 LSTM 예측 모델을 구축하였다. 분석을 통해 기존 FFDRI가 정적 요소(FMI·TMI)의 비중이 지나치게 크고 DWI 하나만이 시간적 변동성을 반영하며 전국 단일 공식으로 인해 지역 특성을 충분히 설명하지 못하는 한계를 갖고 있음을 확인하였다.

또한 산불 발생 여부나 피해 규모를 타깃으로 하는 다양한 시도를 수행하였으나, 산불

발생은 인위적 요인의 비중이 높아 환경 변수만으로 설명이 어렵고 피해 규모는 소방 대응 체계·발견 시각·제도적 변화 등 비환경적 요인의 영향을 크게 받아 예측 가능성이 매우 낮으며 시대 변화에 따른 구조적 변화를 통제하기 어려운 한계가 존재함을 확인하였다.

이러한 시행착오를 바탕으로 본 연구는 환경·지형·식생·일사량 등 지역적 특성을 반영한 FFDRI_new로 타깃 자체를 재정의하는 접근이 필요함을 도출하였다. NDVI와 일사량이 계절 구조를 가장 명확하게 분리하는 핵심 변수로 나타난 점은 FFDRI_new 설계의 타당성을 뒷받침하였다.

결론적으로, 본 연구는

1. 고정 좌표 기반 시계열 데이터 구축,
2. 지역 특성을 반영한 새로운 위험지수 정의,
3. 계절 특화 예측 모델 개발

이라는 3단계 구조를 통해 지역 기반 산불 예측 체계의 기초 프레임을 제시했다는 점에서 의의가 있다.

13.2 향후 과제

향후 연구에서는 다음과 같은 보완 방향이 필요하다.

(1) 산불 데이터 희소성 보완 및 데이터 확장

본 연구에서 확인된 가장 근본적인 한계는 산불 발생 데이터의 희소성이다. 산불은 발생 빈도가 낮은 재난 사건이므로, 단일 지역 또는 단기간의 자료만으로는 충분한 학습 표본을 확보하기 어렵고 시계열 모델의 안정성에도 제약이 발생한다. 이러한 한계를 극복하기 위해서는 분석 대상 지역을 확대하거나 장기간의 시계열 자료를 추가로 확보하여 데이터 규모와 다양성을 동시에 증대시킬 필요가 있다. 특히 산림청·기상청 등 국가기관이 보유한 장기·고해상도 기상 및 산불 관련 자료를 연계함으로써, 보다 신뢰도 높은 산불 위험 예측을 위한 데이터 기반을 구축하는 것이 향후 핵심 과제로 판단된다.

(2) FFDRI_new의 설명력 강화 및 외생 변수 통합

FFDRI_new의 예측력을 한 단계 끌어올리기 위해서는 기상 및 환경 요인 중심의 기존 구조에서 나아가, 산불 발생에 실질적인 영향을 미치는 비환경적 요인의 통합이 필요하다.

소방 접근성, 인위적 발화 요인(방화·실화), 감시 체계 수준, 산불 발견 시각과 같은 변수들은 실제 산불 발생 및 확산 양상을 설명하는 데 중요한 역할을 할 수 있다. 또한 산불 피해 규모 분석 과정에서 확인되었듯이, 산불 위험과 피해는 기상 조건뿐 아니라 제도적·사회적 변화의 영향을 크게 받기 때문에, 이러한 시대적 요인을 통제하거나 보정할 수 있는 분석 프레임의 설계가 요구된다. 이를 위해 산림·방재 분야 도메인 전문가와의 협업을 통해 핵심 변수를 정의하고, 체계적인 데이터 구축 전략을 마련하는 것이 필요하다.

(3) FFDRI_new 기반 실시간 산불위험지수 시스템 구축

본 연구에서 구축한 GEE 기반 데이터 수집 및 처리 구조는 향후 실시간 산불위험지수 시스템으로 확장될 수 있는 잠재력을 지닌다. 실시간 기상 예보 및 위성 자료를 자동으로 반영할 수 있도록 시스템을 고도화할 경우, 단순한 연구용 분석을 넘어 대국민 서비스로 활용 가능한 산불 조기경보 체계로 발전시킬 수 있다. 나아가 웹 기반 서비스나 모바일 알림 형태로 제공될 경우, 관리자뿐 아니라 일반 시민도 직관적으로 산불 위험 정보를 인지할 수 있는 의사결정 지원 시스템으로 활용 가능할 것이다.

(4) 지역 특화 산불위험지수의 전국 단위 확장

FFDRI_new의 개발 절차는 특정 지역에 국한되지 않고 다른 지역에도 동일하게 적용 가능하다는 점에서 높은 확장성을 가진다. 다만 지역별로 지형, 식생, 기상 조건이 상이하므로, 각 지역의 특성을 반영한 지수 구조의 파인튜닝이 필요하다. 이러한 접근을 통해 지역 맞춤형 FFDRI 체계를 구축할 수 있으며, 이는 기존의 전국 단일 산불위험지수가 가지는 한계를 보완하는 방향이 될 수 있다. 궁극적으로 본 연구는 전국 단위의 지역 특화 산불위험지수 체계로 확장 가능한 기반을 제시한다는 점에서 의의를 가진다.

Code Availability

본 프로젝트에서 사용한 코드는 다음 GitHub 저장소를 통해 열람할 수 있다.

<https://github.com/HWANDJANG/Daegu-Gyeongbuk-Wildfire-Forecast>

14. 참고문헌

- 학술 논문

Lee, C., Choi, E. H., Han, Y., & Lee, Y. (2025). Year-round daily wildfire prediction and key factor analysis using machine learning: a case study of Gangwon State, South Korea. *Scientific Reports*, 15, 29910. <https://doi.org/10.1038/s41598-025-15508-5>

Choi, J., Yun, Y., & Chae, H. (2025). Forest Fire Risk Prediction in South Korea Using Google Earth Engine: Comparison of Machine Learning Models. *Land*, 14(6), 1155. <https://doi.org/10.3390/land14061155>

- 정부 및 기관 보고서

산림청 산불연구과. (2024. 7.). *국가산불위험예보시스템 위험지수 산출 알고리즘 설명*.

- 웹사이트 및 온라인 자료

NASA Science. (n.d.). *Study: Fire Seasons Getting Longer, More Frequent*. Retrieved October 16, 2025, from <https://science.nasa.gov/earth/climate-change/study-fire-seasons-getting-longer-more-frequent/>

NASA Science. (n.d.). *Wildfires and Climate Change*. Retrieved October 16, 2025, from <https://science.nasa.gov/earth/explore/wildfires-and-climate-change/>

경상북도. (n.d.). *산불발생 현황*. Retrieved October 16, 2025, from https://www.google.com/search?q=https://www.gb.go.kr/Main/open_contents/section/chis/page.do%3FLARGE_CODE%3D700%26MEDIUM_CODE%3D10%26SMALL_CODE%3D10%26SMALL_CODE2%3D60%26mnu_uid%3D6302