

GEE기반 위성 자료를 활용한 경상북도, 대구 지역 특화 산불 위험 예측 모델 개발

산업수학 기말 최종 발표

호환가능

팀장: 현대호

팀원: 장동환, 안가은



목차

- 01 연구 배경 및 중간과정 요약
- 02 위험 유형별 좌표 선정과 예측 모델 설계
- 03 FFDRI_NEW 설계
- 04 시계열 모델 비교
- 05 딥러닝 기반 FFDRI 예측 모델
- 06 최종 결론

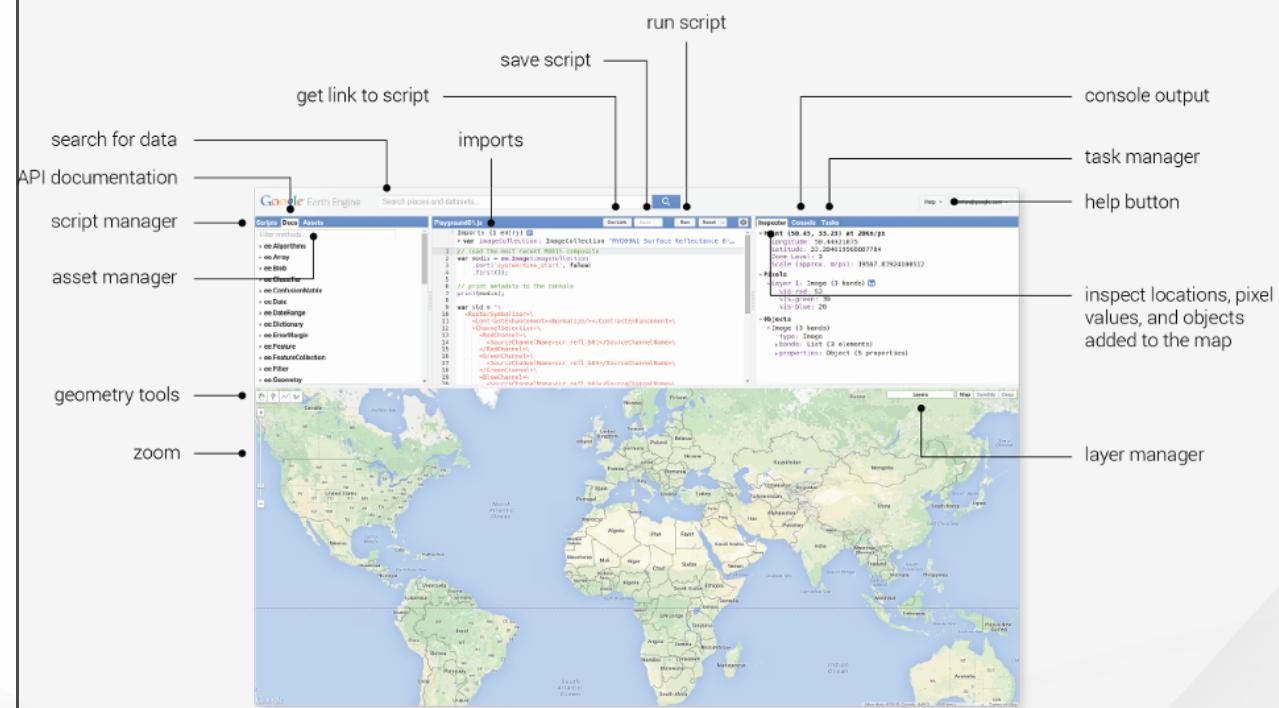
1

연구 배경 및 중간과정 요약

연구 배경

연구배경

기후 변화로 인한 국내 산불의 발생 빈도 증가 및 대형화·장기화 추세
2025년 3월 말 대구 함지산에서 발생한 산불의 재발화



연구 필요성

현재 산림청에서는 전국에 동일한 알고리즘을 적용하여 산불 위험 지수를 산출함

지역별 환경 특성을 반영한
정밀 산불위험도 예측 모델을 구축

연구 목표



GEE 활용
: 다양한 환경요인 변수
실시간 추출



지역 특화 FFDRI
산출



산불 위험 예측 모델 개발
&
위험 수준 지도화

데이터 구성

주요 요인	변수명 (한글)	변수명 (영문 약어)
기상 요인	일평균기온	Tmean
	일최고기온	Tmax
	일최저기온	Tmin
	상대습도	RH
	실효습도	EH
	풍속	WSPD
	강수량	TP_mm
	고도	DEM
지형 요인	경사	Slope
	지표면온도	LST
	식생지수	NDVI
식생 및 토양 요인	토양수분	Soil Moisture

기존 FFDR_I

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

DWI (기상지수), FMI (임상지수), TMI (지형지수)

DWI

preDWI(Tmax, RH, EH, WSPD), RNE

FMI

임상도(2019, 2022 2024)

TMI

DEM, Slope

일 가중치

산림청 정보 기준 일 가중치 사용

| 중간 결과 분석과 보완

① 주요 문제: 랜덤 좌표로 인한 시공간 불연속

FFDRI 데이터셋

산불 발생, 비발생(랜덤샘플링) 시공간이 불연속적

	A	B	C
1	date	lon	lat
2	2023-11-30	128.1582	35.60781
3	2023-12-03	129.0026	35.69456
4	2023-10-18	128.6537	35.69612
5	2023-10-18	128.7057	35.72681
6	2023-10-24	128.6522	35.85625
7	2023-11-30	128.6522	35.86524
8	2023-11-23	128.4236	35.90755

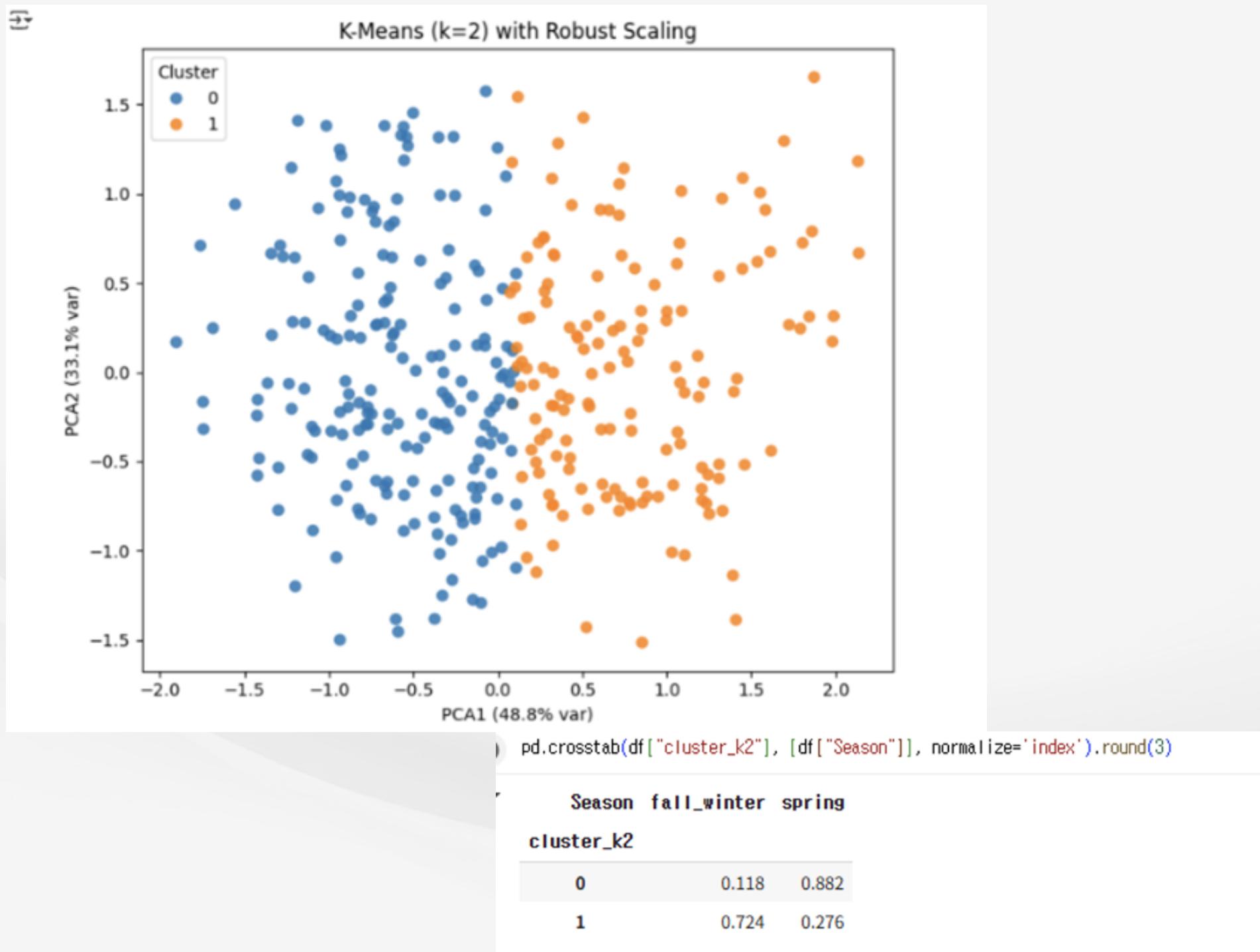
보완 방향

위험 대표 좌표 15개 pid 고정 후
시계열 데이터로 만들기

- 딥러닝 모델이 요구하는 '고정된 공간에서의 연속적인 시계열' 조건을 충족
- 과거 산불 발생 지점과 경사·풍속이 높은 고위험 산림지역을 기반으로 대표 좌표를 다시 선별
- 산불과 지형·기상 요인이 일관되게 반영되는 15개의 고정 좌표를 확보하고, 각 좌표에서 2019–2024년의 연속적인 일 단위 시계열을 수집

중간 결과 분석과 보완 ① FFDRI의 계절별 분포 차이

클러스터링 결과



보완 방향

계절 별로
모델 분리하여 학습시키기

- NDVI, FFDRI, sunlight 변수 조합으로 클러스터링 진행
- Cluster 0은 봄(Spring) 데이터 비율이 88.2%로 매우 높고,
Cluster 1은 가을·겨울(Fall/Winter) 데이터가 72.4%로 다수를 차지
- 즉, 계절적 특성에 따라 군집이 분리됨

②

위험 유형별 좌표 선정과 예측 모델 설계

1. 산불 위험 후보 지점 도출 과정

중간까지 구축된 전체 데이터 셋을 기반으로 두 가지 조건을 모두 만족하는 지점들을 “산불 위험 후보 좌표”로 도출

	조건 ① 과거 실제 발생 좌표	조건 ② 산림 지형·기상 조건 위험 구간
1	2019~2024년 산불 피해면적 및 관련 위경도 활용	경사도 상위 20%
2	실제 산불 발생 지점을 전체 데이터 셋에서 직접 추출	평균 풍속 상위 30%

두 조건의 합집합을 최종 후보 좌표로 선정

① 실제 산불이 발생한 지점 \cup ②지형·기상 위험도가 높은 지점
→ 최종 산불 위험 후보 좌표

2. 위험 유형별 대표 좌표 선별 과정

구축된 위험 후보 좌표들을 연구논문 기반 가중치에 따라 세 가지 위험 유형으로 점수화하여 분류하고 각 **유형별 대표 지점을 균형 있게 선정**

위험 유형 분류 기준 (Score-Based Classification)

- 건조 취약형(DryScore)

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

- 강풍 확산형(WindScore)

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

- 굽경사 확산형(SlopeScore)

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

→ 가중치는 아래 두 논문을 참고하여 설정

- (1) Scientific Reports (Nature Portfolio): Forest Fire Hazard
- (2) Forest Fire Risk Prediction in South Korea using GEE

최종 선정 좌표: 유형별 대표 포인트

lat	lon	type
35.78743	129.1584	dry
35.81058	129.1764	dry
36.0076	129.1011	dry
36.01729	129.0982	dry
36.07185	128.751	dry
35.79539	128.6214	wind
35.87422	128.3558	wind
35.91733	129.4419	wind
36.48568	128.398	wind
36.78152	128.9487	wind
36.16176	128.9703	slope
36.8748	128.4311	slope
36.94743	128.465	slope
37.02858	128.7424	slope
37.07099	129.2116	slope

세 유형을 모두 포함한 대표 좌표 선정으로 모델이 다양한 위험 패턴을 학습하고 특정 유형에 편향되는 것을 방지

3. 시계열 데이터 구축 과정

2019-2024년 기간 동안 선정된 15개 지점에서 모든 기상·식생·지형 변수를 날짜 기준으로 정렬하여 ‘시계열 데이터셋’을 구축

기상요인

- Tmean
- RH
- TP_mm
- WSPD
- Sunlight_ERAS

식생지표

- NDVI

지형요인

- Slope
- Elevation

건조지수

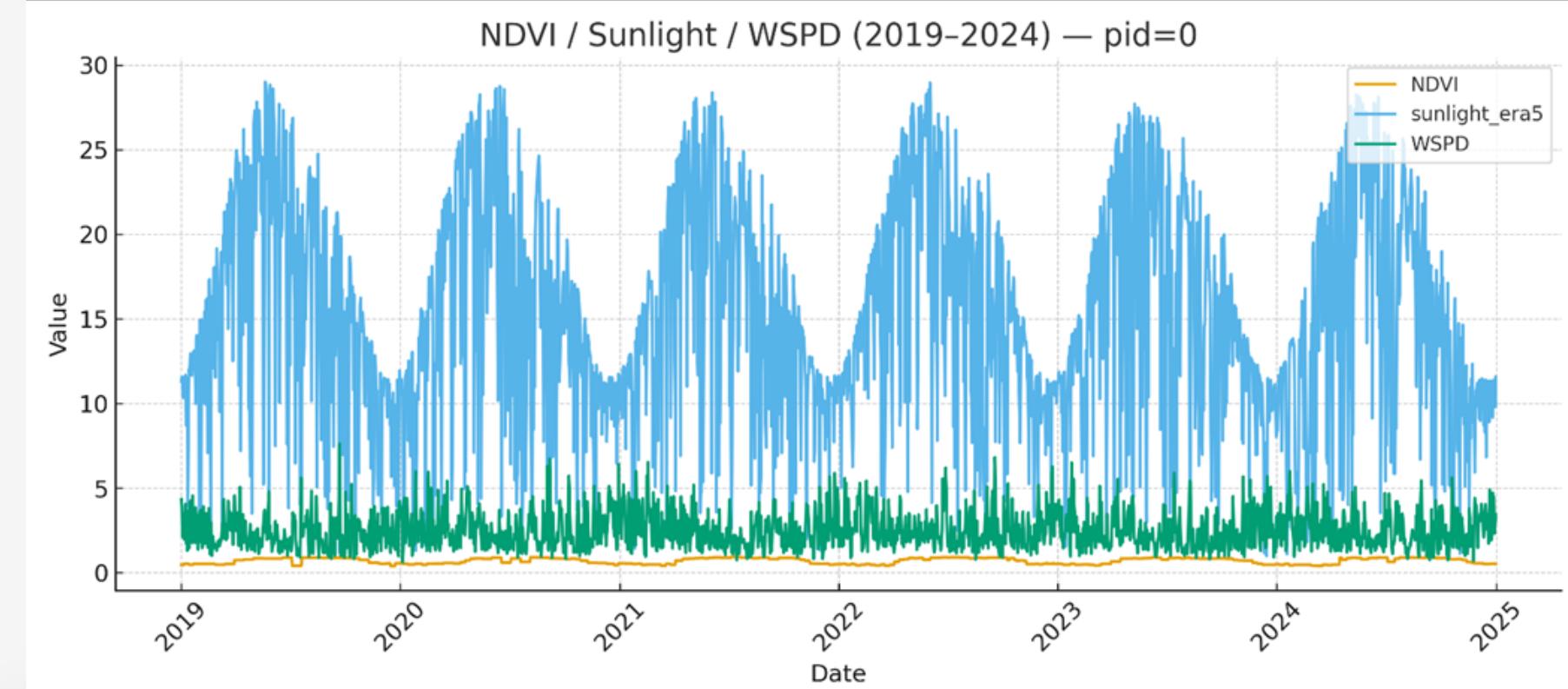
- DWI
- FMI
- TMI
- FFDRI

3. 시계열 데이터 구축 과정

6년 × 15좌표 × 365일의 일 단위 완전 시계열 데이터 구축 (총 약 32,880 rows)

32862	2024-12-12	12	12	11~20	0.83	fall	14	129.2116
32863	2024-12-13	12	13	11~20	0.83	fall	14	129.2116
32864	2024-12-14	12	14	11~20	0.83	fall	14	129.2116
32865	2024-12-15	12	15	11~20	0.83	fall	14	129.2116
32866	2024-12-16	12	16	11~20	0.83	fall	14	129.2116
32867	2024-12-17	12	17	11~20	0.83	fall	14	129.2116
32868	2024-12-18	12	18	11~20	0.83	fall	14	129.2116
32869	2024-12-19	12	19	11~20	0.83	fall	14	129.2116
32870	2024-12-20	12	20	11~20	0.83	fall	14	129.2116
32871	2024-12-21	12	21	21~31	0.83	fall	14	129.2116
32872	2024-12-22	12	22	21~31	0.83	fall	14	129.2116
32873	2024-12-23	12	23	21~31	0.83	fall	14	129.2116
32874	2024-12-24	12	24	21~31	0.83	fall	14	129.2116
32875	2024-12-25	12	25	21~31	0.83	fall	14	129.2116
32876	2024-12-26	12	26	21~31	0.83	fall	14	129.2116
32877	2024-12-27	12	27	21~31	0.83	fall	14	129.2116
32878	2024-12-28	12	28	21~31	0.83	fall	14	129.2116
32879	2024-12-29	12	29	21~31	0.83	fall	14	129.2116
32880	2024-12-30	12	30	21~31	0.83	fall	14	129.2116
32881	2024-12-31	12	31	21~31	0.83	fall	14	129.2116
32882								

변수 시계열 분포 그래프



4. 1차 예측 모델 구성

FFDRI의 시계열 패턴을 모델이 학습할 수 있는지 진단하기 위해
기상, 식생, 지형 요인을 입력으로 한 다음날 FFDRI 예측 LSTM 베이스라인 모델을 구축

Step 1

Input variables

기상 요인(Tmean, RH, TP_mm, WSPD, Sunlight_ERA5), 식생 지표(NDVI), 지형 요인(Slope, Elevation)

Step 2

Target

다음날 FFDRI ($t+1$)

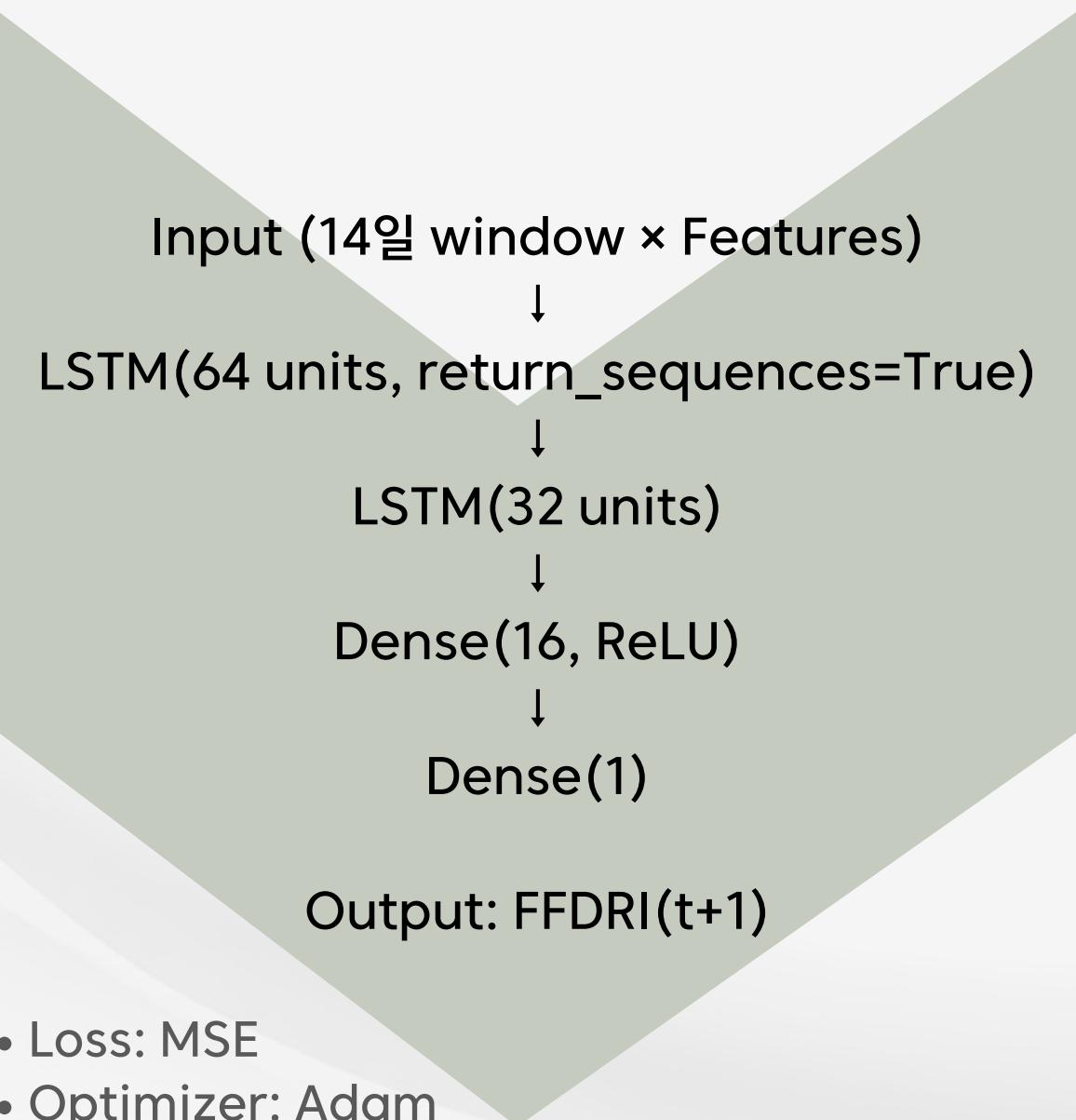
Step 3

window와 예측 범위 설정

14일 입력 → 다음날 FFDRI 예측

4. 1차 예측 모델 구성

FFDRI의 시계열 패턴을 모델이 학습할 수 있는지 진단하기 위해
기상, 식생, 지형 요인을 입력으로 한 다음날 FFDRI 예측 LSTM 베이스라인 모델을 구축



- Loss: MSE
- Optimizer: Adam
- Metrics: MAE
- EarlyStopping(patience=3)

```

def build_lstm(input_shape):
    m = models.Sequential()
    m.add(layers.Input(shape=input_shape))
    m.add(layers.LSTM(64, return_sequences=True))
    m.add(layers.LSTM(32))
    m.add(layers.Dense(16, activation="relu"))
    m.add(layers.Dense(1))

    m.compile(
        loss="mse",
        optimizer=tf.keras.optimizers.Adam(1e-3),
        metrics=["mae"])
    return m
  
```

모델 설계 목적이 “진단”이기 때문에 단순한 구조로 진행

4. 1차 예측 모델의 한계 및 원인 분석

Spring MAE: 4.6298683927393105

Spring R2 : 0.00012083320039957268

1 R²가 0에 근접

- 모델이 FFDRI 변동성을 설명하지 못함

2 MAE가 FFDRI 값의 스케일 대비 높은 수준

- 실제 FFDRI 흐름을 따라가지 못함

4. 1차 예측 모델의 한계 및 원인 분석

FFDRI
26.89372
26.43856
25.95776
26.36736
26.0364
26.32554
26.72986
26.25004
26.93602
27.29987
24.1328
24.19393
24.65216

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

임상지수(FMI) : 침엽수림 2, 혼효림 3, 침엽수림 10

지형지수(TMI)

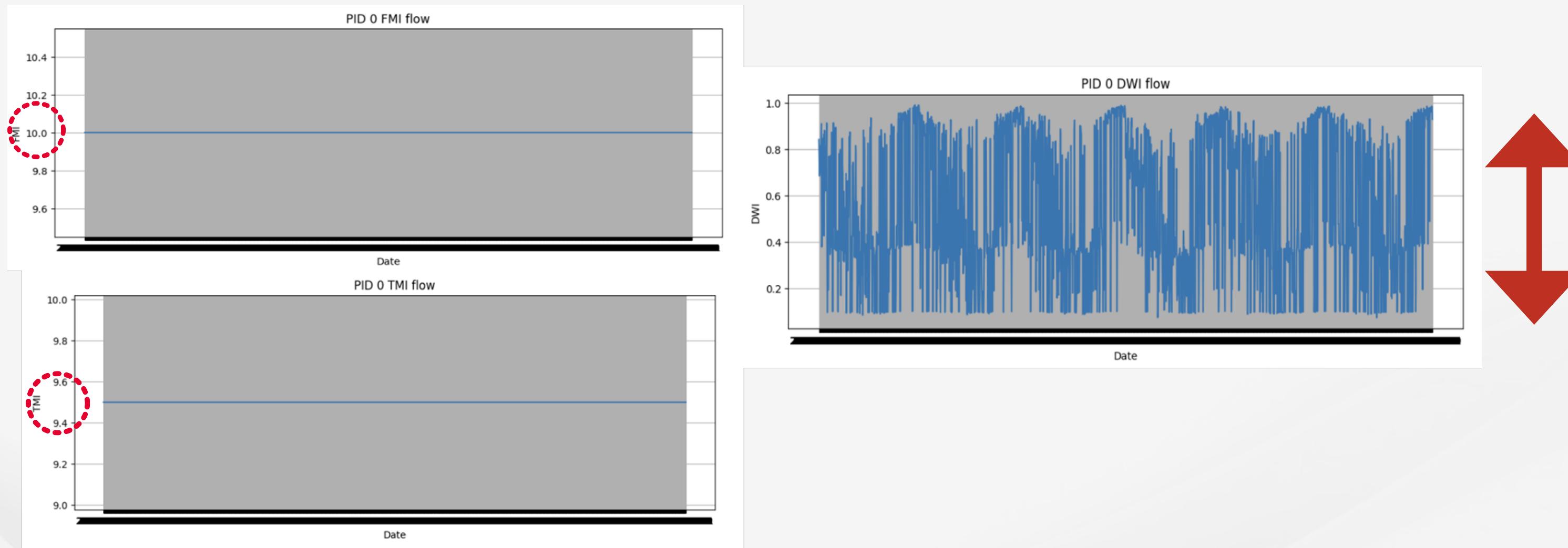
방위	지수	발화지점	지수	고도분포(m)	지수
E	1.5	산정하부	0.5	876 이상	1
N/W	2.5	산복상부	0.5	628~876 미만	2
SE/S	4	산복하부	1	380~628 미만	3
NW/NE	4.5	산록상부	1.5	132~380 미만	4
SW	5	산록하부	5	132 미만	5

3 FFDRI는 변동성이 있는 유동적인 지수

- Regression을 통해서 파악할 수 있는 구조일 것이라고 단순하게 생각

4 FMI, TMI는 점수화를 통해 산출되는 구조

4. 1차 예측 모델의 한계 및 원인 분석



4 FMI, TMI는 점수화를 통해 산출되는 구조

- FMI, TMI가 DWI에 비해 수치가 많이 높음
- DWI(기상지수)는 Regression을 통해 예측이 가능한 구조이지만, 기상지수를 예측하는 것만 고려하면 프로젝트의 의미가 기상 예측과 별 다를게 없어짐

③ FFDRI_new 설계

FFDRI_new 설계

FFDRI_new 개발 배경

1차 모델 실험 결과

- DWI, FMI, TMI만으로는 FFDRI 변동을 설명하는 데 한계가 있음

시계열적 문제

- 기존 FFDRI는 지역별 특성 및 시계열 변화 패턴이 충분히 반영되지 않은 지수 구조

지역 특화 FFDRI 필요

- 기존 지수를 그대로 예측하기보다, 지역형 FFDRI를 새롭게 정의하고 가중치를 재추정

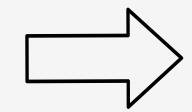
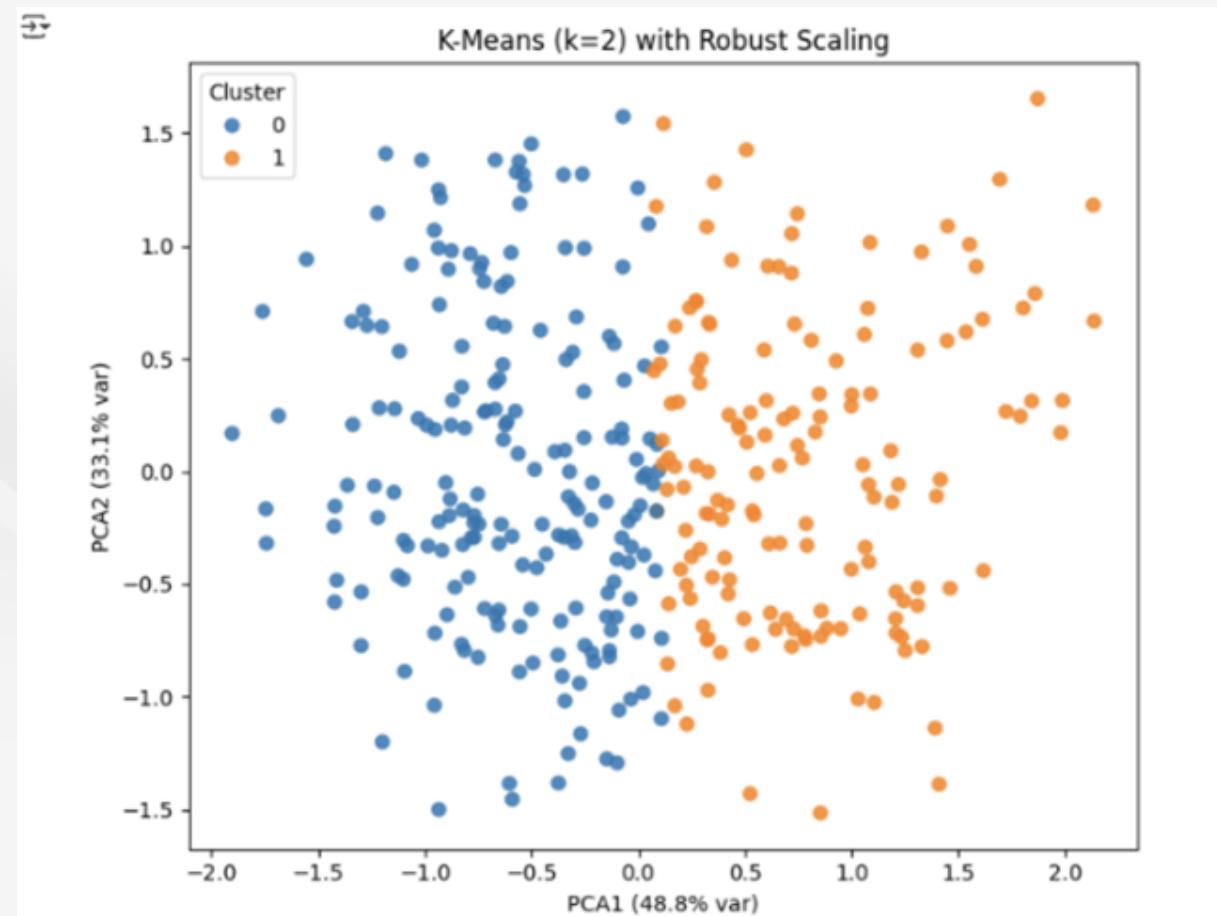
FFDRI_new 설계

FFDRI_new 주요 개념

기존 FFDRI 구성요소(DWI, FMI, TMI)에 지역성 반영 변수(NDVI, sunlight_era5)를 추가하여
지역 특성과 계절적 패턴을 함께 내재화한 연속형 지수



NDVI와 일사량(sunlight_era5)은 경북·대구 지역에서 산불 위험 시계열 패턴을 가장 잘 분리하는 변수이다



NDVI, sunlight_era5를 FFDRI에 반영하는 것을 핵심 요소로 판단

FFDRI_new 설계

Step 1

입력 변수

DWI, FMI, TMI, NDVI, sunlight_era5

Step 2

타깃

기존 FFDRI

Step 3

방법

Ridge Regression으로 기존 FFDRI를 가장 잘 설명하는 회귀 계수를 추정 표준화 후 역변환해 최종 공식 도출

Step 4

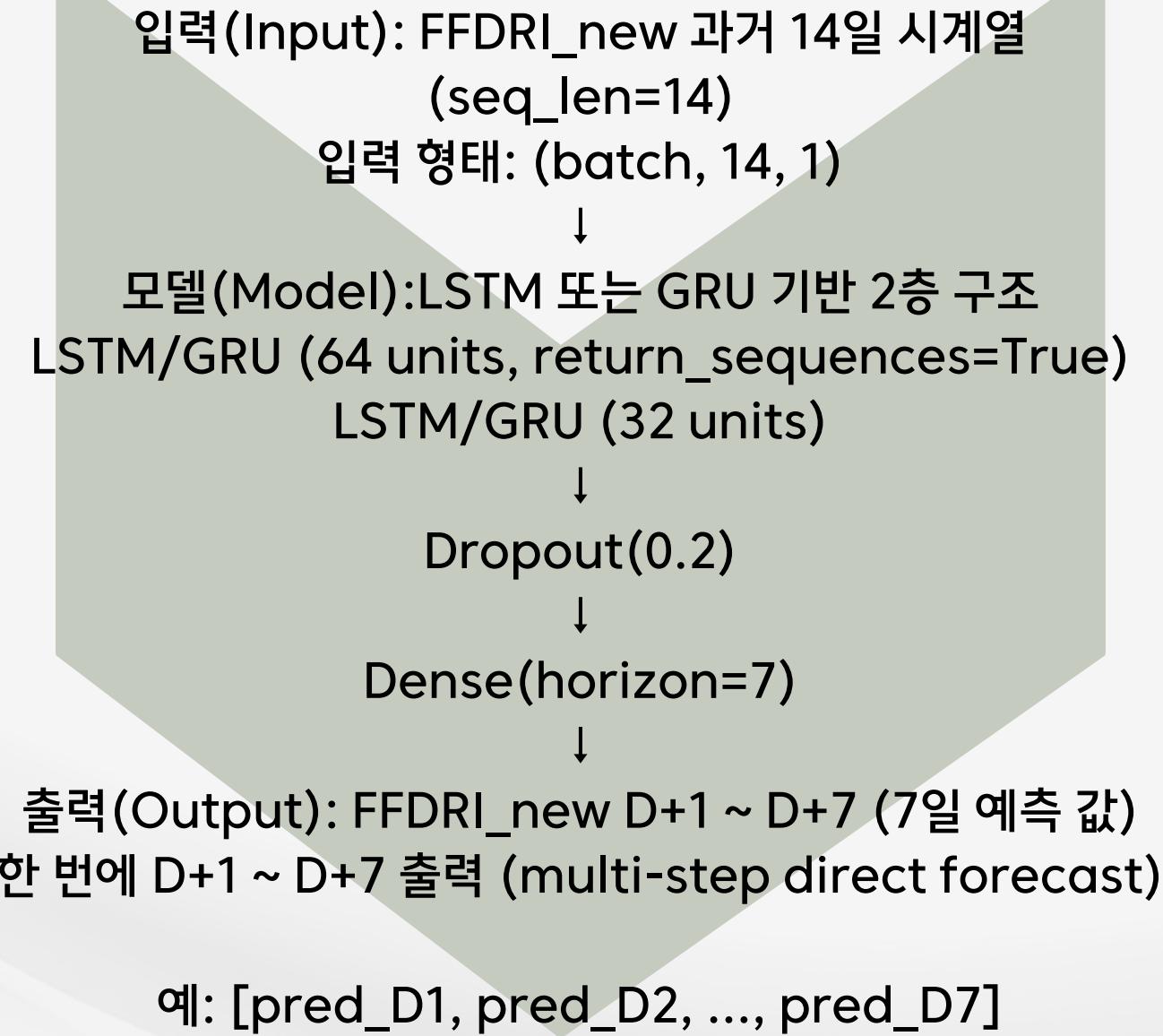
최종 산출 공식

$$\text{FFDRI}_{\text{new}} = 5.3012 + 6.6232 \cdot \text{DWI} + 0.7657 \cdot \text{FMI} + 1.0994 \cdot \text{TMI} + 0.0906 \cdot \text{sunlight} - 10.6796 \cdot \text{NDVI}$$

4

시계열 모델 비교

FFDRI_new 기반 LSTM/GRU 시계열 예측 모델 구조

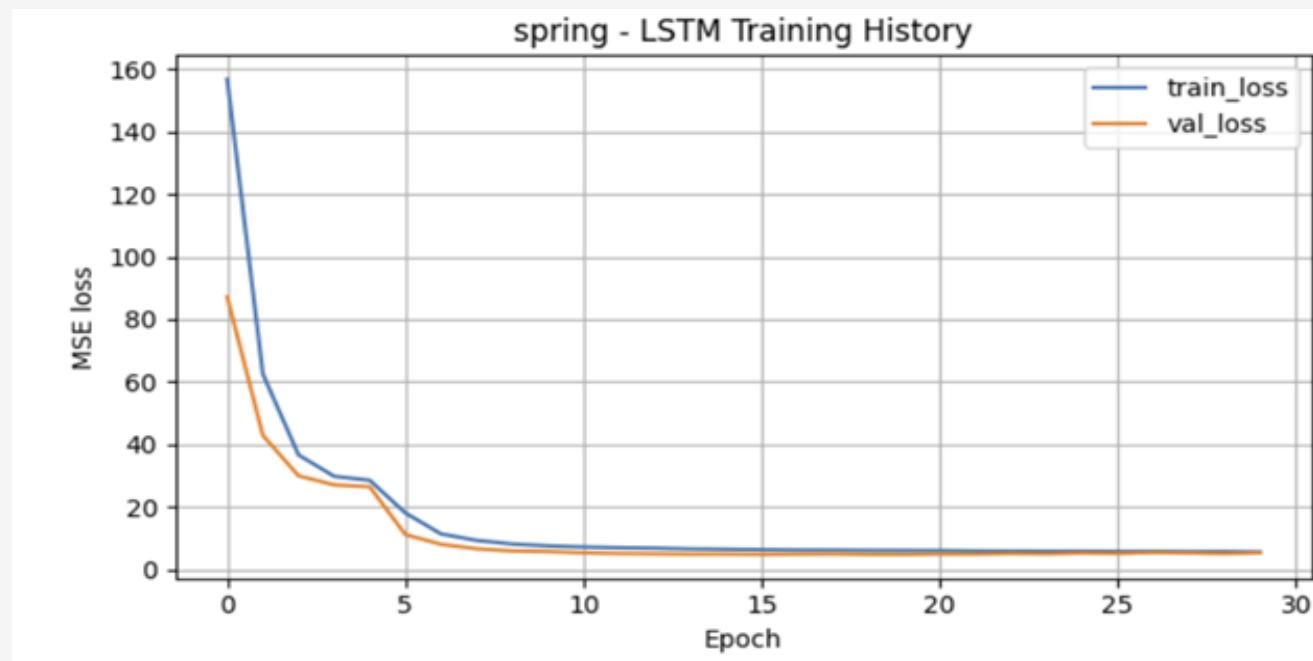


```

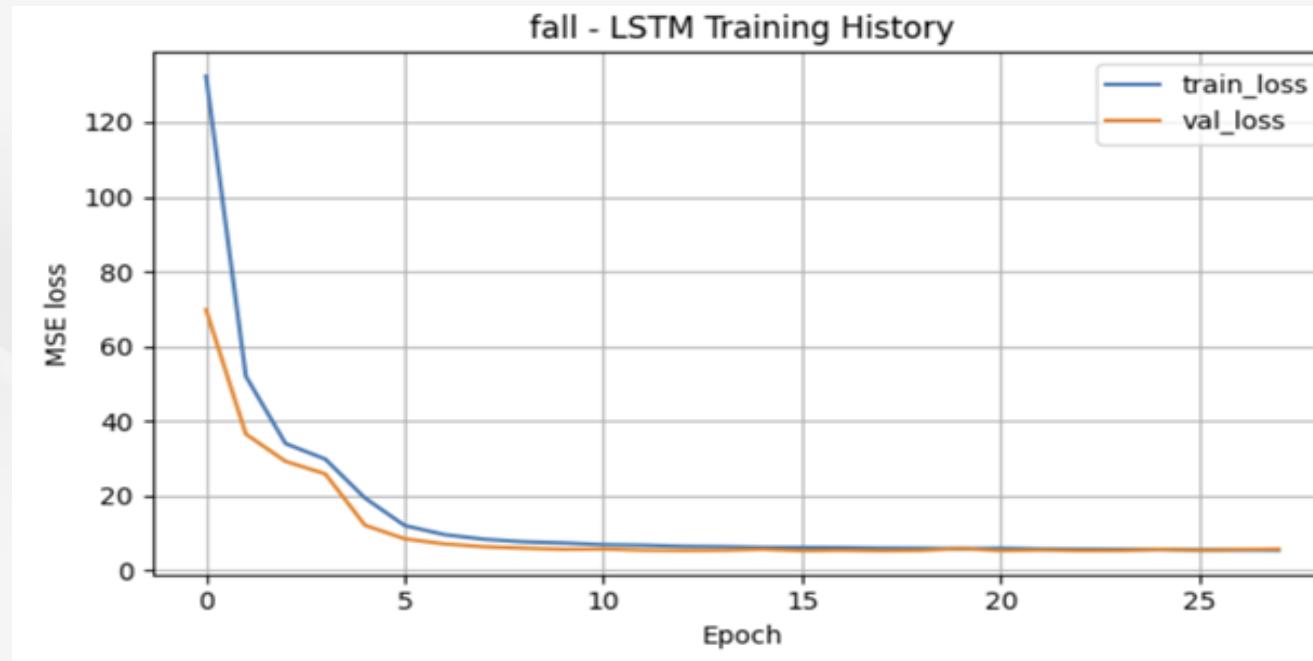
# 6. LSTM / GRU 모델 정의
def build_lstm_model(timesteps, n_features, horizon):
    model = Sequential([
        LSTM(64, return_sequences=True,
             input_shape=(timesteps, n_features)),
        LSTM(32),
        Dropout(0.2),
        Dense(horizon)
    ])
    model.compile(
        optimizer=tf.keras.optimizers.Adam(1e-3),
        loss="mse",
        metrics=["mae"]
    )
    return model

def build_gru_model(timesteps, n_features, horizon):
    model = Sequential([
        GRU(64, return_sequences=True,
             input_shape=(timesteps, n_features)),
        GRU(32),
        Dropout(0.2),
        Dense(horizon)
    ])
    model.compile(
        optimizer=tf.keras.optimizers.Adam(1e-3),
        loss="mse",
        metrics=["mae"]
    )
    return model
  
```

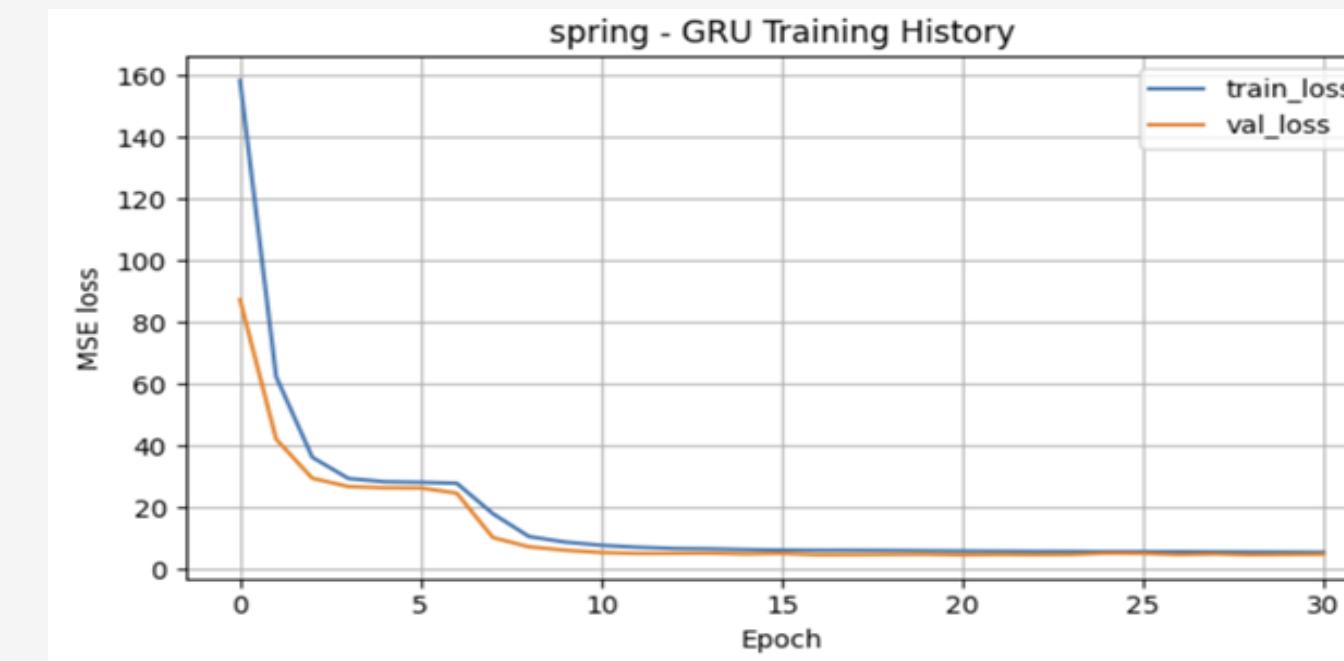
LSTM vs GRU 성능 비교 및 운영 모델 선정



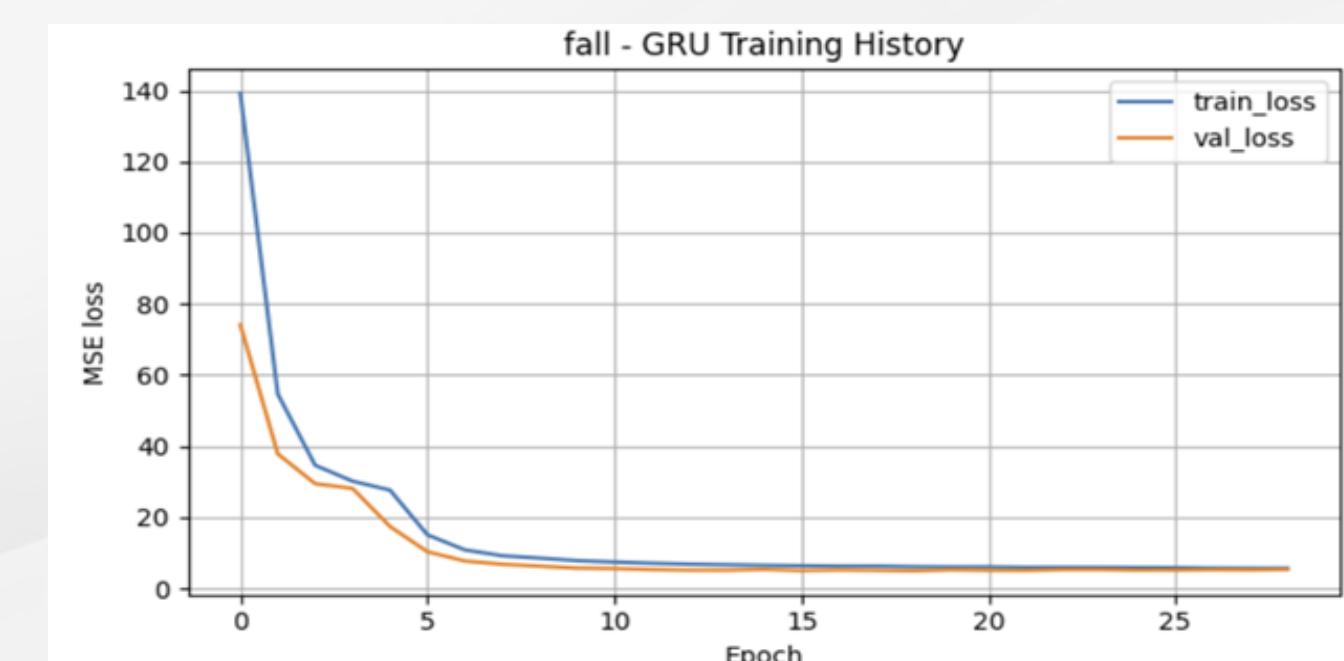
[spring][LSTM] 테스트셋 성능
(t+1) RMSE: 2.338332463502439 MAE: 1.9832006692886353 Corr: 0.8929729279072603
(1~7) RMSE: 2.4737413384824203 MAE: 2.0727391242980957



[fall][LSTM] 테스트셋 성능
(t+1) RMSE: 2.207037623573485 MAE: 1.7838232517242432 Corr: 0.9031385274808519
(1~7) RMSE: 2.286399359728201 MAE: 1.8576616048812866



[spring][GRU] 테스트셋 성능
(t+1) RMSE: 2.269524421615965 MAE: 1.8996083736419678 Corr: 0.8987932092578198
(1~7) RMSE: 2.4787375348538045 MAE: 2.0512471199035645



[fall][GRU] 테스트셋 성능
(t+1) RMSE: 2.1240242232397972 MAE: 1.7182897329330444 Corr: 0.9109366161665555
(1~7) RMSE: 2.290253105002402 MAE: 1.8593662977218628

LSTM vs GRU 성능 비교 및 운영 모델 선정

계절별 성능 요약

봄(Spring): GRU가 더 낮은 val_loss·RMSE, 더 높은 Corr

가을(Fall): GRU가 LSTM 대비 모든 지표에서 우세

공통: 두 모델 모두 학습·검증 손실 추이 유사 → 안정적인 수렴

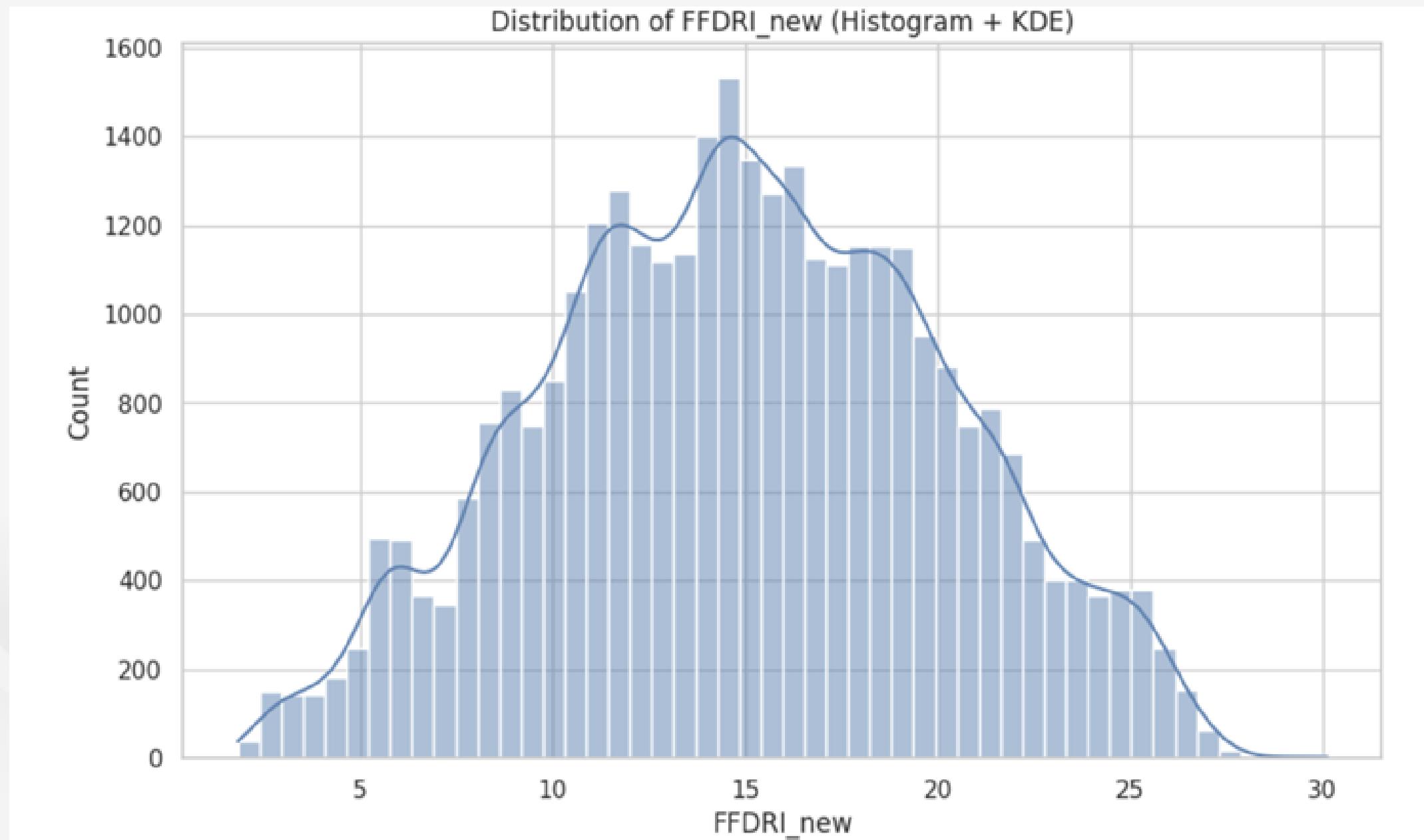
정량적 비교

상관계수 Corr: 0.89–0.91 수준 (양호)

RMSE: 2.1–2.5 (FFDRI 평균 대비 적절한 오차 수준)

GRU가 계산 효율성과 성능 모두 우수하여 최종 운영 모델로 선정

FFDRI_new 위험등급 구분 기준



count 32880.000000
mean 14.970236
std 5.228975
min 1.807475
25% 11.256842
50% 14.915760
75% 18.750781
max 30.127765

* FFDRI_new 분포 토대로 등급을 매김

2025-04-28 (대구 함지산 대형산불 일자) 테스트

항목	값
base date	2025-04-28
위도(lon)	35.91818
경도(lat)	128.56667
D0 FFDRI_new	18.9436
D0 위험등급	very_high

예측일	FFDRI_new	위험등급
D+1	19.34192	very_high
D+2	19.38093	very_high
D+3	19.25636	high
D+4	19.32277	high
D+5	19.21741	high
D+6	19.18091	high
D+7	19.15541	high

- GRU(Spring)은 실제 산불 발생 시점을 놓치지 않고 위험도를 안정적으로 포착
 - 운영 모델로 채택한 이유를 실증적으로 보여주는 사례
- > 향후 실시간 지도 시각화 및 위험 알림 시스템 적용 가능성 ↑

함지산 12월 일자(2025.12.3) 테스트

예측일	FFDRI_new	위험등급
D0	18.78215	very_high
D+1	18.61684	very_high
D+2	18.71068	very_high
D+3	18.39303	very_high
D+4	18.37815	very_high
D+5	18.46237	very_high
D+6	18.41466	very_high
D+7	18.32278	very_high

가을 GRU 모델 예측 결과

예측일	FFDRI_new	위험등급
D0	18.22778	high
D+1	16.50277	high
D+2	16.46316	high
D+3	16.29112	high
D+4	16.36241	high
D+5	16.24227	high
D+6	16.14547	high
D+7	16.18823	high

봄 GRU 모델 예측 결과

| 함지산 12월 일자(2025.12.3) 테스트

항목	봄 모델(Spring)	가을 모델(Fall)
D0_FFDRI_new	18.22778	18.78215
D0 위험등급	high	very_high
계절 반응	12월에 비교적 낮게 예측	12월 건조 패턴 반영해 매우 높게 예측
해석	계절 불일치로 위험도 억제	계절 일치로 예측 정확도 ↑

가을 모델은 10~2월의 건조 및 산불 증가 패턴을 학습

-> 12월 실제 환경과 잘 맞아 떨어지며 위험등급을 높게 예측

봄 모델과 달리 seasonality를 더 정확하게 반영한 결과, 계절 분리 학습이 효과적으로 이루어짐.

5

딥러닝 기반 FFDRI 예측 모델

5개 산 FFDRI_new 7일 예측

산 이름	오늘	위험도	1일후	위험	2일후	위험도	3일후	위험도	4일후	위험도	5일후	위험도	6일후	위험도	7일후	위험도	7일 중 최대 위험
Hamji-san (함지산)	18.78	high	17.85	high	17.04	high	16.97	high	16.91	high	17.05	high	16.81	high	16.81	high	high
Pal-gong-san (팔공산)	18.85	high	17.89	high	17.07	high	16.99	high	16.93	high	17.07	high	16.82	high	16.83	high	high
Geumo-san (금오산)	18.89	high	17.90	High	17.09	high	16.99	high	16.94	high	17.08	high	16.83	high	16.83	high	high
Juwang-san (주왕산)	18.88	high	17.91	High	17.09	high	16.99	high	16.94	high	17.08	high	16.83	high	16.83	high	high
Cheongnyang-san (청량산)	12.13	moderate	11.62	moderate	11.06	moderate	10.89	moderate	10.72	moderate	10.63	moderate	10.55	moderate	10.30	low	moderate

| 결과 요약

대구·경북 주요 산은
공통적으로 'high' 등급 유지

- FFDRI_new ≈ 16.8 ~ 18.9 범위
 - 1~7일 동안 위험 수준이 크게 떨어지지 않고 안정적으로 높은 상태 유지
- 겨울철 건조 패턴이 반영된 정상적인 결과



청량산만 유일하게
'moderate → low'로
상대적으로 낮은 위험도

- 해발, 기후 영향으로 습도가 상대적으로 높고 건조도가 낮은 지역 특성 반영
- 실제 패턴과 일치하는 예측



모델이 산별·지역별 특성을
잘 반영하고 있음

- 도심·교통 접근성 높은 산 (함자·팔공·금오·주왕)은 high 지속
- 산악 고원 지역인 청량산은 중간 수준 유지



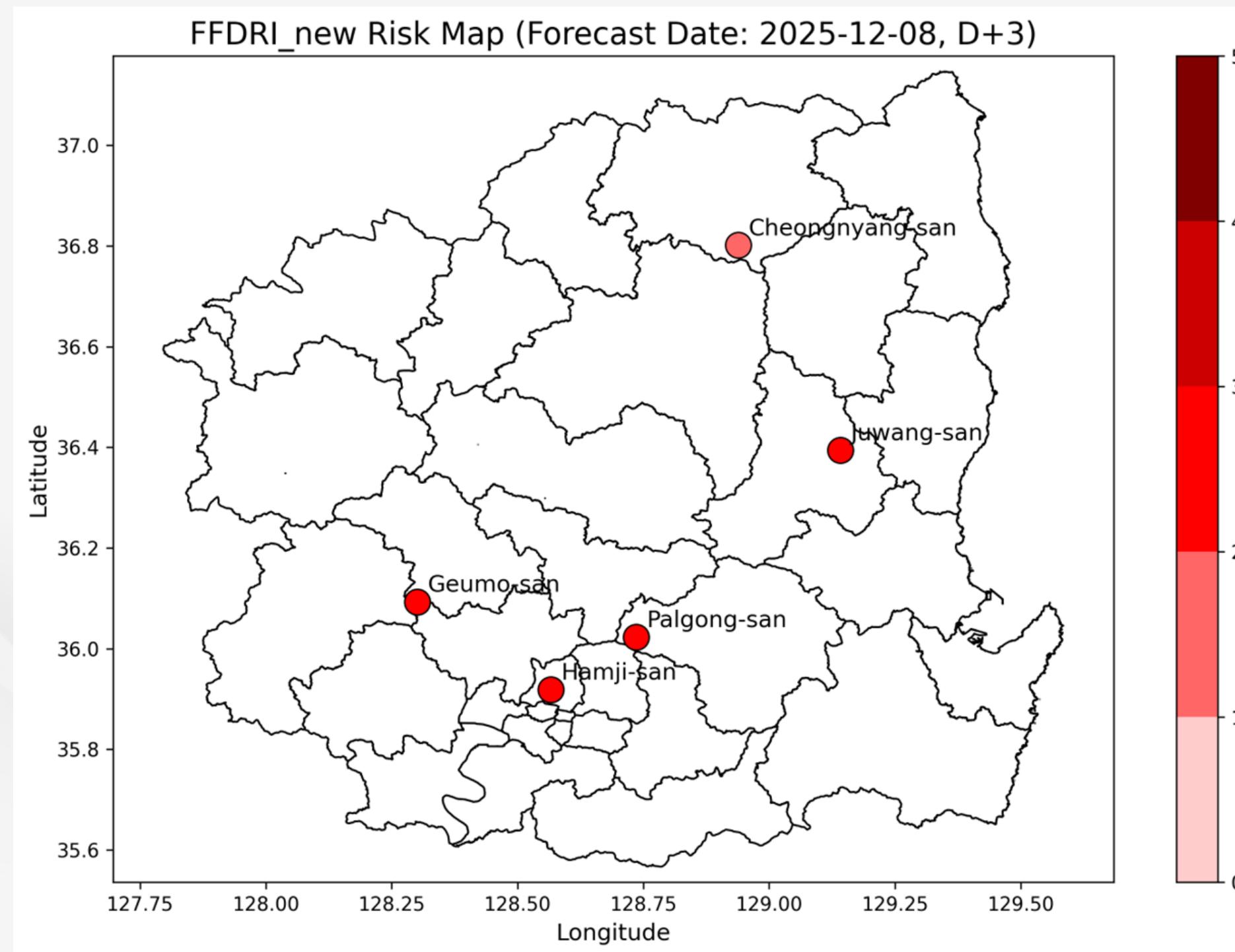
7일 내 최대 위험등급이
명확히 지역 특성을 구분

- 4개 산 = high
- 청량산 = moderate



대구·경북 대표 산 5개 지역의 7일 예측 결과, 지역별 산불위험 패턴이 구분되며
모델이 지형·기상·노출 특성을 안정적으로 반영함을 확인

경북·대구 5대 주요 산 산불위험도 지도 시각화 (D+3 Forecast)



shp(경계), shx(인덱스), dbf(속성), prj(좌표계) 파일을 통합하여 경상북도·대구 행정경계를 정확히 복원한 GIS 지도 위에서 모델 예측값을 표현

- FFDRI_new 예측 결과를 기반으로 산불 위험도를 지도에 시각화
- 원하는 위치의 좌표를 수동 입력 가능
- 기준일(D) 입력 시 예측일기준 +7일까지 위험등급 시각화 가능
- 색상으로 공간적 위험 수준을 한눈에 비교 가능
- GIS 기반이라 확장성 높음

6

최종 결론

| 최종 결론



1 산불 데이터는 희소·불규칙적이며 자연·인위 요인이 복합적으로 작용

- 단순 ML/DL로 학습하기 어려운 비정형 구조적 특성을 가짐.

2 대표 좌표 15개를 선정해 활용

- 공간·시간 요소를 동시에 고려해야 하는 도메인 특성으로
좌표 수가 증가하면 시계열 데이터가 기하급수적으로 폭증 → 전처리·정합 과정에서 큰 계산적 부담 발생.
- 경북·대구 전 지역의 지형·기상 다양성을 완전히 커버하기는 어려움(공간적 제약)

3 현실적인 산불 예측 모델 구축의 기반을 마련

- FFDRI_new 기반 좌표·시계열 구조 재설계
- 계절별 GRU 예측 모델 구축, 위험도 시각화 및 지도 기반 분석

4 산불 경보 시스템(Prototype)을 구현

- 모델 예측 → 위험등급 판정 → 지도 시각화: 초기 운영 가능 형태의 경보 체계 프로토타입을 제시함
- 산림좌표는 수동적으로 입력해야한다는 한계가 있음

감사합니다

산업수학 기말 최종 발표 | 호환가능 | 현대호, 장동환, 안가은