

석사학위논문

Google Earth Engine의  
기계학습을 통한 KOMPSAT-3/5  
위성영상의 고해상도 토지피복  
분류 및 SHAP 기반 결과 분석

2026년

한 성 대 학 교 대 학 원

융 합 보 안 학 과

융 합 보 안 전 공

이 정 희



석사학위논문  
지도교수 이기원

Google Earth Engine의  
기계학습을 통한 KOMPSAT-3/5  
위성영상의 고해상도 토지피복  
분류 및 SHAP 기반 결과 분석

Land Cover Classification by Machine Learning in  
Google Earth Engine with KOMPSAT-3/5 Satellite  
Imagery and SHAP-Based Results Analysis

2025년 12월 일

한 성 대 학 교 대 학 원

융 합 보 안 학 과

융 합 보 안 전 공

이 정 희

석사학위논문  
지도교수 이기원

Google Earth Engine의  
기계학습을 통한 KOMPSAT-3/5  
위성영상의 고해상도 토지피복  
분류 및 SHAP 기반 결과 분석

Land Cover Classification by Machine Learning in  
Google Earth Engine with KOMPSAT-3/5 Satellite  
Imagery and SHAP-Based Results Analysis

위 논문을 공학 석사학위 논문으로 제출함

2025년 12월 일

한 성 대 학 교 대 학 원

융 합 보 안 학 과

융 합 보 안 전 공

이 정 희

이정희의 공학 석사학위 논문을 인준함

2025년 12월 일

심사위원장 서 화 정 (인)

심 사 위 원 이 기 원 (인)

심 사 위 원 박 명 서 (인)

# 국 문 초 록

## Google Earth Engine의 기계학습을 통한 KOMPSAT-3/5 위성영상의 고해상도 토지피복 분류 및 SHAP 기반 결과 분석

한 성 대 학 교 대 학 원  
융 합 보 안 학 과  
융 합 보 안 전 공  
이 정 희

최근 Google Earth Engine (GEE) 플랫폼에서 제공하는 기계학습 알고리즘을 활용한 연구가 점점 활발히 이루어지고 있다. 본 연구에서는 GEE에서 애플리케이션 프로그래밍 인터페이스(application programming interface, API)를 통해 제공하는 중·저해상도의 위성 데이터와 GEE에서 직접 제공하지 않는 고해상도 위성 데이터를 직접 올려 융합하여 사용하였다. 색상정보를 제공하는 광학영상인 Sentinel2, KOMPSAT-3A, Landsat8와 구름이나 대기 조건에 영향을 받지 않는 레이더 SAR영상인 Sentinel1, KOMPSAT-5 데이터를 이용하였다. 또한 Sentinel2 데이터와 Landsat8 데이터를 이용해 계산한 정규 식생 지수(Normalized Difference Vegetation Index, NDVI)자료를 추가해 Random Forest 알고리즘으로 연구 지역에 대한 토지피복(LULC) 분류에 사용하였다. 토지피복 분류 결과에 대하여 Accuracy와 Kappa 계수를 비교 분석하였으며 분류 성능 지표로 F-1 Score를 이용하였다. SHAP 기법은 다수의 입력 변수와 다중 대상 객체를 포함하는 복합적인 기계학습 모델을

해석하는데 적합하며, 단일 객체 추출 중심의 딥러닝 모델보다 다양한 지표 정보를 다루는 기계학습 기반 분석에 더욱 효과적으로 활용될 수 있다. 토지 피복 분류 결과를 바탕으로 SHAP 분석을 진행하였으며, 각 피복 유형에 가장 영향을 미치는 특성을 식별하는데 유용한 정보를 제공함을 확인하였다. 이 연구는 설명 가능한 기계학습 기법이 다중 센서 고해상도 위성영상을 이용한 지표 환경의 이해 및 모니터링에 효과적인 도구로 활용 될 수 있음을 시사한다.

**【주요어】** KOMPSAT-3A, KOMPSAT-5, Google Earth Engine, SHAP, Image classification, Random Forest

# 목 차

|   |    |
|---|----|
| 제 1 장 서 론 .....                               | 1  |
| 제 1 절 연구 배경 .....                             | 1  |
| 제 2 절 연구 목적 .....                             | 2  |
| 제 2 장 클라우드 기반 처리 플랫폼 및 XAI 기법 연구 동향 .....     | 4  |
| 제 1 절 클라우드 기반 처리 플랫폼 .....                    | 4  |
| 제 2 절 GEE 및 활용 가능한 위성데이터와 기계학습 알고리즘 .....     | 6  |
| 제 3 절 XAI 기법과 SHAP .....                      | 8  |
| 1. Stacked Summary Bar Plot .....             | 10 |
| 2. Bee Swarm Plot .....                       | 10 |
| 3. Dependency Plot .....                      | 11 |
| 4. Waterfall Plot .....                       | 12 |
| 제 3 장 연구 자료 및 방법 .....                        | 13 |
| 제 1 절 토지피복지도(Land Use Land Cover, LULC) ..... | 13 |
| 제 2 절 연구 지역과 연구 자료 .....                      | 14 |
| 제 3 절 연구 방법 .....                             | 17 |
| 제 4 장 토지피복 분류 결과 및 SHAP 결과 .....              | 20 |
| 제 1 절 토지피복 분류 결과 .....                        | 20 |
| 제 2 절 Stacked Summary Bar Plot 결과 .....       | 24 |
| 제 3 절 Bee Swarm Plot 결과 .....                 | 27 |
| 제 4 절 Dependency Plot 결과 .....                | 31 |

|                               |    |
|-------------------------------|----|
| 제 5 절 Waterfall Plot 결과 ..... | 36 |
| 제 5 장 결 론 .....               | 40 |
| 참 고 문 헌 .....                 | 42 |
| ABSTRACT .....                | 47 |

## 표 목 차

|  |    |
|--|----|
| [표 2-1] 클라우드 기반 위성처리 플랫폼 비교 .....  | 5  |
| [표 2-2] GEE에서 지원하는 Supervised classification .....   | 7  |
| [표 3-1] 연구 자료 .....  | 16 |
| [표 4-1] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8<br>and NDVI(Sentinel-2) F1-Score .....          | 21 |
| [표 4-2] Sentinel-1, Sentinel-2, KOMPSAT3A, Landsat 8 and<br>NDVI(Sentinel-2, Landsat 8) F1-Score .....           | 22 |
| [표 4-3] Sentinel-1, Sentinel-2, KOMPSAT3A, KOMPSAT-5 Landsat 8<br>and NDVI(Sentinel-2, Landsat 8) F1-Score ..... | 23 |

## 그림 목 차

|  |    |
|--|----|
| [그림 2-1] Shapely Value 공식 .....  | 9  |
| [그림 2-2] SHAP 모델 분해 공식 .....   | 9  |
| [그림 3-1] 환경공간정보서비스에서 확인할 수 있는 토지피복지도 정보 .....  | 13 |
| [그림 3-2] 위성 데이터별 연구지역 .....  | 15 |
| [그림 3-3] 연구지역 .....  | 16 |
| [그림 3-4] 연구 흐름도 .....  | 17 |
| [그림 3-5] (a) Sentinel-2 NDVI 계산 결과, (b) Landsat 8 NDVI 계산 결과 .....   | 18 |
| [그림 3-6] 학습 자료 및 검증 자료 .....   | 18 |
| [그림 3-7] GEE에서 활용할 수 있는 RF 알고리즘 .....  | 19 |
| [그림 4-1] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) 토지피복 분류 결과 .....                        | 21 |
| [그림 4-2] Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) 토지피복 분류 결과 .....                        | 22 |
| [그림 4-3] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5 Landsat 8 and NDVI(Sentinel-2, Landsat 8) 토지피복 분류 결과 .....              | 23 |
| [그림 4-4] 세 가지 경우에 대한 전체 정확도 및 Kappa 계수 .....   | 24 |
| [그림 4-5] Stack Summary Bar Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) .....           | 25 |
| [그림 4-6] Stack Summary Bar Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....           | 26 |
| [그림 4-7] Stack Summary Bar Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5 Landsat 8 and NDVI(Sentinel-2, Landsat 8) ..... | 26 |
| [그림 4-8] Bee Swarm Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) .....                   | 28 |

|   |    |
|---|----|
| [그림 4-9] Bee Swarm Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....                          | 29 |
| [그림 4-10] Bee Swarm Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5 Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....               | 30 |
| [그림 4-11] Dependency Plot(Water Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) · .....          | 33 |
| [그림 4-12] Dependency Plot(Water Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) ...              | 33 |
| [그림 4-13] Dependency Plot(Water Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8) ..... | 33 |
| [그림 4-14] Dependency Plot(Veg. Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) ... ..            | 34 |
| [그림 4-15] Dependency Plot(Veg. Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) ...               | 34 |
| [그림 4-16] Dependency Plot(Veg. Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....  | 34 |
| [그림 4-17] Dependency Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) ... ..                 | 35 |
| [그림 4-18] Dependency Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) ...                    | 35 |
| [그림 4-19] Dependency Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....       | 35 |
| [그림 4-20] Waterfall Plot(Water): Sentinel-1, Sentinel-2,  |    |

|           |   |    |
|-----------|---|----|
|           | KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) ...   | 36 |
| [그림 4-21] | Waterfall Plot(Water): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) ...              | 37 |
| [그림 4-22] | Waterfall Plot(Water): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8) ..... | 37 |
| [그림 4-23] | Waterfall Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) .....             | 38 |
| [그림 4-24] | Waterfall Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....             | 38 |
| [그림 4-25] | Waterfall Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8) .....  | 39 |

# 제 1 장 서 론

## 제 1 절 연구 배경

클라우드 컴퓨팅 기술의 발전으로 많은 플랫폼들이 나타나고 있다. Google Earth Engine(GEE)은 2010년에 Google에서 공개된 클라우드 기반 위성영상 처리 플랫폼으로, 900개 이상의 분석 지원 데이터셋과 50 페타바이트에 달하는 위성영상정보와 지구과학 정보를 애플리케이션 프로그래밍 인터페이스(Application Programming Interface, API)를 통해 무료로 제공한다. 또한 GEE 기능 중 하나인 Code Editor를 이용하여 자료들을 실시간으로 처리하고 결과를 확인할 수 있다. Per-pixel 연산, Kernel 연산, 배열/행렬 연산, Reducer를 기반으로 한 통계 수치 정보 처리 및 기계학습 함수 등을 제공하며, 개인이 만든 소스코드 또한 공유할 수 있기 때문에 여러 활용 분야에 필요한 전문 함수(Specialized function)들이 개발자 커뮤니티를 중심으로 계속 개발되고 있다. Amani et al., 2020에 따르면 주로 식생, 농업, 수문학, 토지 피복, 도시 연구 분야 순으로 활용된다. 토지 피복은 현재 지표면이 어떤 물리적 특성을 가지는지 나타낸다. 토지 피복은 생태 환경 평가, 자원 관리, 도시 계획 및 기후 변화를 포함한 다양한 환경 연구에 필수적이다(Wang et al., 2023). 최근 중·고해상도 위성영상에 대한 접근성이 높아지면서 토지피복 분류 연구가 빠른 속도로 증가하고 있다. 특히 GEE와 같은 클라우드 기반 지리정보 플랫폼이 등장하며 대규모 데이터 처리와 분류를 효율적으로 할 수 있게 됨에 따라 기계학습과 딥러닝 기반의 토지피복 분류 연구가 활발히 진행되고 있다(Alshehri et al., 2025).

이렇게 기계학습과 딥러닝을 적용하는 경우 보편적으로 분류 정확도의 값으로 결과를 분석하지만 어떤 입력 자료가 어떻게 분류 산출물에 영향을 주는 지가 블랙박스(Black-box)로 다루어지는 문제가 있다(Gevaert, 2022). 특히 서로 다른 특성을 가진 다양한 입력 변수를 동시에 사용하게 되면서 모델이 특정 클래스를 어떻게 구분하고, 어떤 변수가 분류에 주요하게 작용했는지

를 이해하는 필요성이 더욱 커지고 있다(Temenos et al., 2023). SHAP(SHapley Additive exPlanations)는 각 특성이 예측에 기여한 정도를 정량적으로 제시하는 대표적인 설명가능한 인공지능(eXplainable AI: XAI) 기법으로 주목받고 있다. 또한 변수 간 상호작용을 시각적으로 제시할 수 있다는 점에서 복잡한 블랙박스 모델의 작동 원리를 해석하는데 핵심적인 역할을 한다. 모델의 작동 원리에 대한 해석이 가능할 경우 모델의 신뢰도 향상과 더불어 분류 결과의 타당성을 제시할 수 있다(Rodríguez-Pérez et al., 2020). 그러나 현재 기계학습과 딥러닝을 이용한 토지피복 분류와 XAI 기법을 함께 적용하는 연구는 적은 실정이다.

## 제 2 절 연구 목적

최근 GEE의 Random Forest (RF) 기계학습 알고리즘을 적용한 여러 사례 연구가 발표되고 있다. Vizzari. (2022)는 브라질 열대 지역을 대상으로 NICFI PlanetScope, Sentinel-2, Sentinel-1을 GEE에서 융합하고 RF 알고리즘을 이용해 픽셀 기반과 객체 기반 토지 사용 및 토지 피복(Land Use and Land Cover, LULC) 분류를 수행한 결과 높은 정확도를 보였고, Aji et al. (2023)은 인도네시아 지역 내 맹그로브(mangrove) 종의 시계열 분류를 위해 Sentinel-2 영상과 정규식생지수(Normalized Difference Vegetation Index, NDVI) 자료를 RF 알고리즘에 적용한 연구를 수행하였다. Ahmed et al. (2024)은 독일 연구 지역의 20년간의 LULC 주제도를 제작하기 위해 Landsat-7과 Landsat-9 영상정보를 RF 알고리즘에 적용하였으며, 이 연구 결과가 지역 계획 및 정책 결정에 중요한 정보로 활용될 수 있다고 보고하였다. Seitkazy et al. (2024)은 카자흐스탄 지역에서 Landsat 위성영상과 RF 알고리즘을 적용하여 토지이용 변화를 분석하는 연구를 수행하였다. 또한, Fentaw and Abegaz (2024)는 에티오피아 유역의 1990-2021년 31년간의 LULC를 RF 알고리즘을 이용해 분류하였고, 그 결과 토지 황폐화를 생태·자원관리와 연계해 시사점을 제시하였다. Amindin et al. (2024)은 이란 지역에서 생태 안정성 분석을 위해 여러 시기의 LULC 주제도를 작성하면서

Landsat 영상정보와 RF 알고리즘을 활용하였다.

이와 같이 GEE에서 RF 알고리즘을 사용한 연구는 GEE에서 제공하는 위성영상을 활용하여 10년 이상의 장기 변화를 분석하거나 광범위한 연구 지역을 대상으로 분류도를 제작하는 연구가 주로 수행되고 있다. 연구의 목적으로는 먼저 국내 고해상도 위성인 KOMPSAT-3A 광학영상과 SAR 영상인 KOMPSAT-5의 위성영상정보를 GEE에서 제공하는 다른 중·저 해상도의 위성영상정보와 함께 활용하여 GEE 플랫폼에서 RF 알고리즘을 이용해 다중 센서 영상 분류를 수행하고 전체 정확도 Overall Accuracy(OA) 값과 Kappa 계수의 계산 값을 이용하여 정확도 결과를 분석하였다, 이후 수행한 결과를 XAI의 대표적인 기법 중 하나인 SHAP 기법을 이용하여 분류에 사용한 데이터들이 결과에 미치는 영향과 밴드들 간의 상관관계를 확인하고 분석해 보고자 한다.

## 제 2 장 클라우드 기반 처리 플랫폼 및 XAI 기법 연구 동향

### 제 1 절 클라우드 기반 처리 플랫폼

최근 위성자료의 공간적·시간적 해상도가 향상되고 수집 주기가 단축되면서 전 지구 규모의 지리공간 데이터는 페타바이트 단위에 이를 정도로 빠르게 증가하고 있다(Höhn et al., 2025). 이러한 방대한 데이터를 기존의 데스크탑 기반 환경에서 처리하기에는 연산 자원, 저장 공간, 전처리 요구 사항 등의 한계가 존재했다. 이에 따라 대규모 위성영상 및 공간정보를 효율적으로 저장, 관리, 분석하기 위한 클라우드 기반 위성처리 플랫폼에 대한 중요성이 연구 및 산업 분야에서 더욱 커지고 있다(Xu et al., 2022). 클라우드 플랫폼은 사용자가 로컬 컴퓨터의 성능에 의존하지 않고, 서버 사이드에서 병렬 연산을 수행할 수 있도록 지원함으로써 대용량 시계열 분석, 다중센서 융합, 기계학습 기반 분류 작업 등을 보다 효율적으로 처리할 수 있게 한다.

대표적인 클라우드 기반 위성처리 플랫폼으로는 GEE, Microsoft Planetary Computer(MPC), Sentinel Hub, Amazon Web Service(AWS) Open Data Program 등이 있으며 각 플랫폼은 서로 다른 장점을 가지고 있다. GEE는 페타바이트 규모의 Landsat, Sentinel, MODIS 등 다중센서 위성 자료를 제공하고 있으며, 사용자는 브라우저 또는 API 환경에서 직접 영상 전처리, 시계열 분석과 RF, Support Vector Machine(SVM)과 같은 기계학습 기반 분류를 수행할 수 있다는 장점이 있다. 특히 데이터 접근과 연산이 용이하다는 점에서 광범위하게 활용되고 있다(Xu et al., 2022). GEE, MPC, Sentinel Hub, AWS 플랫폼 비교는 [표 2-1]에서 확인할 수 있다.

[표 2-1] 클라우드 기반 위성처리 플랫폼 비교

| Category                 | GEE   | MPC   | Sentinel Hub   | AWS Open Data Program  |
|--------------------------|---|---|--|--|
| Provider                 | Google  | Microsoft<br>Azure  | Sinergise /<br>ESA   | Amazon Web<br>Service  |
| Data Catalog             | Landsat,<br>Sentinel-1/2,<br>MODIS,<br>ASTER,<br>NAIP, DEM,<br>etc. | Sentinel-1/2/<br>5P, Landsat,<br>NOAA,<br>ERA5, NAIP,<br>etc. (STAC<br>Catalog) | Sentinel-2,<br>Sentinel-1<br>(partial),<br>Sentinel-3<br>(limited),<br>Landsat,<br>DEM | Sentinel,<br>Landsat,<br>MODIS,<br>NOAA, etc.<br>(raw data on<br>S3) |
| Access Interface         | Web IDE,<br>JS/Python<br>API  | Python API,<br>STAC API,<br>Azure ML<br>integration                             | Web API,<br>WMS/WMTS<br>, Python/JS<br>API   | Direct S3<br>access;<br>custom<br>processing via<br>EC2/SageMaker    |
| Machine Learning Support | Built-in ML<br>(RF, SVM,<br>CART, GBT)                              | No built-in<br>ML (uses<br>Azure ML)  | No built-in<br>ML  | No built-in<br>ML (ML via<br>SageMaker/E<br>C2)                      |

## 제 2 절 GEE 및 활용 가능한 위성데이터와 기계학습 알고리즘

GEE는 Google이 구축한 클라우드 기반 위성처리 플랫폼으로 대규모 위성영상 및 환경 데이터를 효율적으로 처리하고 분석하기 위한 대표적인 플랫폼이다. 기존의 데스크탑 기반 GIS와 같은 도구는 사용자가 영상 데이터를 직접 다운로드하고 전처리 해야하며, 수십 기가바이트(GB) 이상의 대용량 시계열 자료를 처리하는데 연산과 저장같은 자원의 한계가 존재했다. 하지만 GEE는 페타바이트(Petabyte) 규모의 위성영상 및 지구과학 정보를 클라우드 환경에서 바로 접근할 수 있도록 제공하고 모든 연산을 플랫폼에서 바로 수행할 수 있기 때문에 대규모 시공간 분석, 변화 탐지, 다중 센서 융합 연구에서 큰 이점을 가진다(Gorelick et al., 2017).

GEE는 다양한 위성의 원시자료 및 전처리된 표면반사도(Surface Reflectance) 제품을 포함하여 900개 이상의 데이터셋을 제공한다. JRC Global Surface Water, Hansen Forest Change, CHIRPS 강수량, ERA5 기상자료, SRTM/ALOS DEM과 같은 환경자료도 함께 제공하여 공간분석 연구에 활용할 수 있다. 이러한 데이터들을 별도의 다운로드 없이 즉시 활용할 수 있도록 하여 재현성과 효율성을 동시에 향상시킬 수 있다(Velastegui-Montoya et al., 2023). GEE에서 사용자가 작성한 코드는 서버 기반 병렬 처리가 되어 Google의 대규모 분산 컴퓨팅 인프라에서 병렬 처리 되어 수백 장의 위성 영상 합성, SAR 데이터를 활용한 변화 탐지 등 수 시간이 소요되는 분석을 수 초 또는 수 분 내에 수행할 수 있다(Mutanga et al., 2019). 또한 동일한 코드를 실행하면 항상 동일한 결과를 산출하는 구조이므로 오류를 최소화하고 연구의 재현성을 보장할 수 있다.

RF, SVM, CART(classification and Regression Tree), Gradient Tree Boosting 등 대표적인 기계학습 알고리즘을 지원하고 있어 토지피복 분류, 변화 탐지 등 다양한 원격탐사 기반 분류를 플랫폼 내부에서 직접 수행할 수 있다. 반면 딥러닝 기반 알고리즘은 플랫폼 내부에서 지원하지 않기 때문에 Python과 같은 외부 환경에서 U-Net, CNN과 같은 모델을 학습한 뒤 예측 결과를 다시 GEE로 가져와 시각화하는 방식으로 확장해 활용할 수 있으며

[표 2-2]에는 GEE에서 제공하는 기계학습 알고리즘과 지원하는 출력모드 (CLASSIFICATION, PROBABILITY, MULTIPROBABILITY 등)를 정리한 것이다. 또한 웹 기반 code editor와 Python API를 모두 제공하기 때문에 높은 접근성과 확장성을 가지며, 코드 및 분석 결과를 URL로 공유할 수 있어 협업 및 재현성을 높인다(Developers Guide, 2025).

[표 2-2] GEE에서 지원하는 Supervised classification

| Classifier                           | CLASSIFICATION | REGRESSION | PROBABILITY | MULTIPROBABILITY | RAW | RAW_REGRESSION |
|--------------------------------------|----------------|------------|-------------|------------------|-----|----------------|
| ee.Classifier.amnhMaxent             | X              | X          | ✓           | X                | X   | X              |
| ee.Classifier.minimumDistance        | ✓              | ✓          | X           | X                | ✓   | X              |
| ee.Classifier.smileCart              | ✓              | ✓          | ✓           | ✓                | X   | X              |
| ee.Classifier.smileGradientTreeBoost | ✓              | ✓          | ✓           | ✓                | X   | X              |
| ee.Classifier.smileKNN               | ✓              | X          | ✓           | X                | X   | X              |
| ee.Classifier.smileNaiveBayes        | ✓              | X          | ✓           | ✓                | X   | X              |
| ee.Classifier.smileRandomForest      | ✓              | ✓          | ✓           | ✓                | ✓   | ✓              |
| ee.Classifier.libsvm C_SVC           | ✓              | X          | ✓           | ✓                | X   | X              |
| ee.Classifier.elibsvm NU_SVC         | ✓              | X          | ✓           | ✓                | X   | X              |
| ee.Classifier.elibsvm ONE_CLASS      | ✓              | X          | X           | X                | X   | X              |
| ee.Classifier.elibsvm EPSILON_SVR    | ✓              | X          | X           | X                | X   | X              |
| ee.Classifier.elibsvm NU_SVR         | X              | ✓          | X           | X                | X   | X              |

### 제 3 절 XAI 기법과 SHAP

최근 인공지능 및 기계학습 알고리즘은 다양한 분야에서 높은 예측 성능을 보이며 널리 활용되고 있다. 그러나 이러한 모델들은 복잡한 내부 구조와 비선형적 의사 결정 방식으로 인해 왜 이러한 예측을 내렸는가(Why)를 이해하기 어려운 블랙박스(black-box) 특성을 가진다(Guidotti et al., 2018). 특히 딥러닝 기반 모델의 경우 계층 구조가 깊어질수록 해석 가능성이 감소하며, 이는 결과의 신뢰성, 모델 검증, 과학적 해석 등을 요구하는 실제 응용 분야에서 중요한 한계로 나타난다(Zhang et al., 2020). 이러한 배경에서 설명가능 인공지능(XAI)이 등장했다.

XAI는 모델의 예측 결과 및 내부 의사결정 과정을 인간이 이해할 수 있도록 설명하는 것을 목표로 하며 모델의 신뢰성을 향상시키기 위한 중요한 접근법으로 대두되고 있다(Räuker et al., 2022).

XAI를 구현하는 대표적 기법으로는 LIME(Local Interpretable Model-agnostic Explanations), SHAP, Grad-CAM(Gradient-weighted Class Activation Mapping), Integrated Gradients, Layer-wise Relevance Propagation(LRP) 등이 있다. LIME은 모델의 국소 영역을 선형 모델로 근사하여 중요 변수를 설명한다(Ribeiro et al., 2016). 그리고 Grad-CAM 계열은 주로 CNN 기반 딥러닝 모델에서 관심 영역(Region of Interest, ROI)을 시각적으로 강조하는 방법에 사용된다(Selvaraju et al., 2017).

SHAP은 이들 기법 중 가장 널리 활용되고 있다. SHAP은 Shapley Value를 기반으로 하며, 각 입력 특징이 최종 예측에 기여한 정도를 공정하게 분배하는 원리는 사용한다. [그림 2-2]과 [그림 2-3]에서 공식을 확인할 수 있다. Shapley Value는 특정한 특징이 포함되었을 때와 제외되었을 때 모델의 예측 값 변화량을 계산하여, 각 특징의 기여도를 정량적으로 보여준다(Shapley, L. S., 1953). SHAP기법의 식은 특정한 특징이 포함된 경우와 제외된 경우의 차

이를 모든 가능한 조합에 대해 특징이 기여한 정도를 계산한다(Lundberg et al., 2017). 대표적인 SHAP 시각화 기법으로는 Stacked Summary Bar Plot, Bee Swarm Plot, Dependence Plot, Waterfall Plot이 있다.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! \cdot (|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)]$$

$\phi_i$ : Feature  $i$ 의 SHAP 값

$N$ : 전체 Feature 집합

$S$ : Feature  $i$ 를 제외한 부분 집합

$v(S)$ : 부분 집합  $S$ 를 기반으로 한 모델의 예측 함수 값

[그림 2-1] Shapely Value 공식

$$f(x) = \phi_0 + \sum_{j=1}^M \phi_j x_j'$$

$\phi_0$ : 베이스라인 예측값

$\phi_j$ : Feature  $j$ 의 SHAP 값

$x_j$ : Feature 존재 여부를 나타내는 이진 변수

[그림 2-2] SHAP 모델 분해 공식

## 1. Stacked Summary Bar Plot

Stacked Summary Bar Plot은 특징의 평균 절대 SHAP 값(mean |SHAP value|)을 기준으로 전체 특징의 중요도를 비교하는 feature-level의 중요도 시각화 기법이다(Molnar, 2022). Stacked Bar Plot은 Bee Swarm Plot과 다르게 특징의 기여도가 어떤 방향으로 작용했는지나 값들의 세부 분포를 표현하지 않고 중요도만을 단순화 하여 시각화한다. 따라서 각 특징이 예측 값에 얼마나 기여 하는지의 크기를 상대적으로 확인할 수 있다.

이 그래프에서 막대(bar)는 각 특징의 평균 절대 기여도를 나타내며 값이 클수록 해당 특징이 전체 모델 예측에 중요한 역할을 했다는 것을 의미한다. 게다가 다중 클래스 분류일 경우 클래스별 SHAP 값을 누적형태로 하나의 막대에 표현한다. 따라서 하나의 이미지로 각 클래스에 대한 기여도를 동시에 비교할 수 있다. 이 방식은 입력 자료의 각 밴드에 대한 특징 중요도를 명확하게 보여주기 때문에 모델의 전반적인 구조를 이해하는데 도움이 될 수 있다. Stacked Summary Bar Plot과 Bee Swarm Plot은 서로 상호보완적인 관계이며, 입력 자료에 대한 각 특징의 중요도를 간단하게 제시할 수 있다.

## 2. Bee Swarm Plot

Bee Swarm Plot은 SHAP에서 가장 널리 사용되는 feature-level 설명의 시각화로 전체 데이터셋에 대해 각 특징의 SHAP 값을 점의 형태로 표시하여 모델 내의 전반적인 중요도와 영향 방향을 나타낸다(Molnar, 2022).

이 플롯은 Summary Plot이라고도 불리며 대표적으로 특징의 전체 중요도와 특징 값에 따른 영향 방향을 보여준다. y축에 특징을 중요도 기준으로 상위에서 하위 순서로 나열하고 x축에는 해당 특징의 SHAP 값을 나타낸다. 각 점은 하나의 샘플을 의미하며 점의 색상은 특징의 실제 입력 값의 높고 낮음

을 표현한다. 이를 통해 특징 값이 높은 경우 양의 SHAP값을 가지는지 또는 낮은 값에서 음의 기여도를 가지는지를 시각적으로 판단할 수 있다. 따라서 점의 위치와 색상으로 특정한 특징이 모델의 예측에 어떤 방식으로 영향을 미치는지를 확인할 수 있다.

특히, 모든 특징의 전반적인 영향력을 하나의 그림으로 비교할 수 있다는 장점을 가진다. RF와 같은 트리 기반의 모델에서 자주 사용되며, 전체적인 동작을 이해하는데 유용한 해석 도구로 활용될 수 있다.

### 3. Dependency Plot

Dependence Plot은 특정 특징의 실제 입력 값(feature value)에 따라 SHAP 값이 어떻게 변화하는지를 보여주는 시각화 기법으로 특정한 특징과 모델 예측 간의 관계를 직접적으로 해석할 수 있는 도구이다(Molnar, 2022). x축은 해당 특징의 입력 값, y축은 해당 특징의 SHAP 값을 나타내고 점의 색상은 선택된 다른 보조 특징의 값이다. 이를 통해 두 특징 간의 상호작용을 시각적으로 확인할 수 있다.

특정한 특징이 예측을 증가시키거나 감소시키는 경향을 직접 확인할 수 있으며, 특징 간의 상호작용이 존재할 경우 자동으로 이를 반영하여 특징이 영향을 끼치는 패턴을 보여준다. 예를 들어, NIR과 같은 특정 밴드의 값이 낮을 때 SHAP 값이 높은 양수를 보이는 경우, 이 특징이 수체로 분류하는데 강하게 기여한다는 해석을 할 수 있다. Dependence Plot은 비선형적 관계 또한 시각적으로 확인할 수 있기 때문에 모델의 이해와 변수 간의 관계를 파악하는데 유용하다.

#### 4. Waterfall Plot

Waterfall Plot은 SHAP 값 기반의 개별 샘플(local instance)에 대한 국소적 설명(Local Explanation)을 제공하는 시각화 도구이다(Molnar, 2022). 특정한 하나의 예측이 어떻게 구성되었는지를 단계적으로 나타내며 모델 예측 결과의 합성 구조를 이해하는데 사용된다. Waterfall Plot은 모델의 전체 데이터 평균 예측 값인 기본 값에서 시작하여 각 특징의 SHAP 기여도를 차례대로 더하거나 빼면서 최종 예측 값에 도달하는 과정을 시각적으로 확인할 수 있게 해준다.

그래프는 일반적으로 붉은색과 파란색의 막대로 구성되고 각 특징이 예측 결과를 증가 또는 감소시키는 역할을 하는지 명확하게 시각적으로 보여준다. 이 그래프는 특정 분류 결과가 “왜 그렇게 판정되었는가”를 설명하는데 유용하며, 의사결정의 투명성이 중요한 분야에서 널리 활용된다. Waterfall Plot은 개별 입력 샘플에 대한 설명이므로 모델 검증 또는 특정한 사례의 해석에 사용하는 것이 적합하다.

# 제 3 장 연구 자료 및 방법

## 제 1 절 토지피복지도(Land Use Land Cover, LULC)

LULC는 지표면의 물리적, 인위적 특성과 토지 이용 형태를 나타내는 핵심 공간정보이다. 환경 모니터링, 도시계획, 농업 및 수자원 관리, 생태계 관리 등 다양한 분야에서 활용되는 지표이다. Land Cover는 수체, 식생, 토양과 같은 토지의 자연적 상태를 의미하며 Land Use는 농경지, 도시지역, 산업시설 등의 인간의 사회·경제적 활동에 따른 토지 이용 형태를 나타낸다(Cai et al., 2019). LULC는 지표 변화를 장기적으로 평가할 수 있는 기반자료로 기후변화와 재난 예측, 수문학적 분석 등 정책 및 과학적 의사결정에 이용될 수 있으며, 원격탐사는 넓은 지역을 관측할 수 있어 LULC 분석에서 널리 활용된다(Wang et al., 2023). [그림 3-1]은 환경공간정보서비스에서 확인할 수 있는 토지피복지도이다.



[그림 3-1] 환경공간정보서비스에서 확인할 수 있는 토지피복지도 정보

## 제 2 절 연구 지역과 연구 자료

연구 지역은 전북특별자치도 부안군부터 정읍시 일대 지역으로 서해안 고속도로를 일부 포함한 지역이다. 도로 주변으로 농림 지역과 마을이 분포해있으며, 산림과 수체가 포함되어 있다. 가로 13.8km, 세로 10.5km로 전체 면적은 약 145.37km<sup>2</sup> 이다.

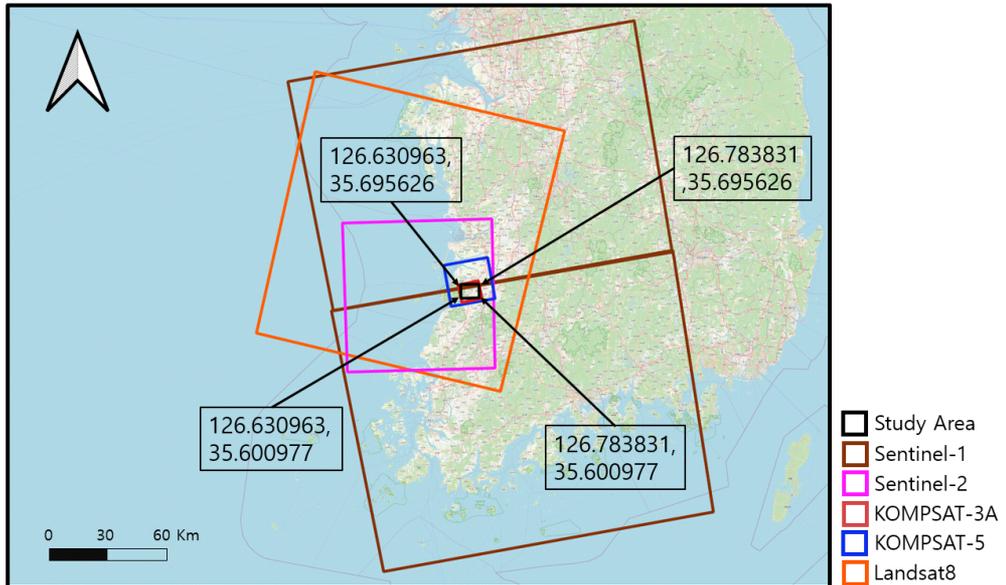
토지피복 분류에 있어, 자연 초지, 방목지, 잔디밭 등을 포함하는 초지 (grassland)와 논, 밭, 작물 재배지 등 농경지(cropland)는 별도로 구분하지 않고 모두 Veg. Zone으로 통합하여 분류하였다. Water Zone은 하천, 연못, 습지, 개울 등을 포함하고 있으며, Bare Zone은 초지, 농림, 산림 등과 같이 식생이 거의 없고 지표면이 노출된 지역을 의미한다. 또한 Forest는 자연림과 조림지를 포함하며, road는 고속도로, 국도, 지방도는 물론 임도, 농로와 같은 소규모 비포장도로와 교량까지 포함한다. 이외에 분류에 사용한 다섯 가지 분류 항목에 포함되지 않는 지역은 별도로 Unclassified Zone으로 정의하였다. 각 클래스는 환경부에서 제공하는 토지피복지도 분류 코드 및 색상표준 (<https://egis.me.go.kr/bbs/landcover.do>)을 참고해 설정하였다.

연구지역은 아래의 [그림 3-2]와 [그림 3-3]에서 확인할 수 있으며, 연구에 사용한 데이터인 Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 영상 등의 촬영 일자 는 모두 2022년 5월로 크게 차이나지 않는다. 이 연구에서 사용한 자료에 대한 촬영 날짜 및 해상도 등의 데이터 정보는 [표 3-1]에 정리하였다.

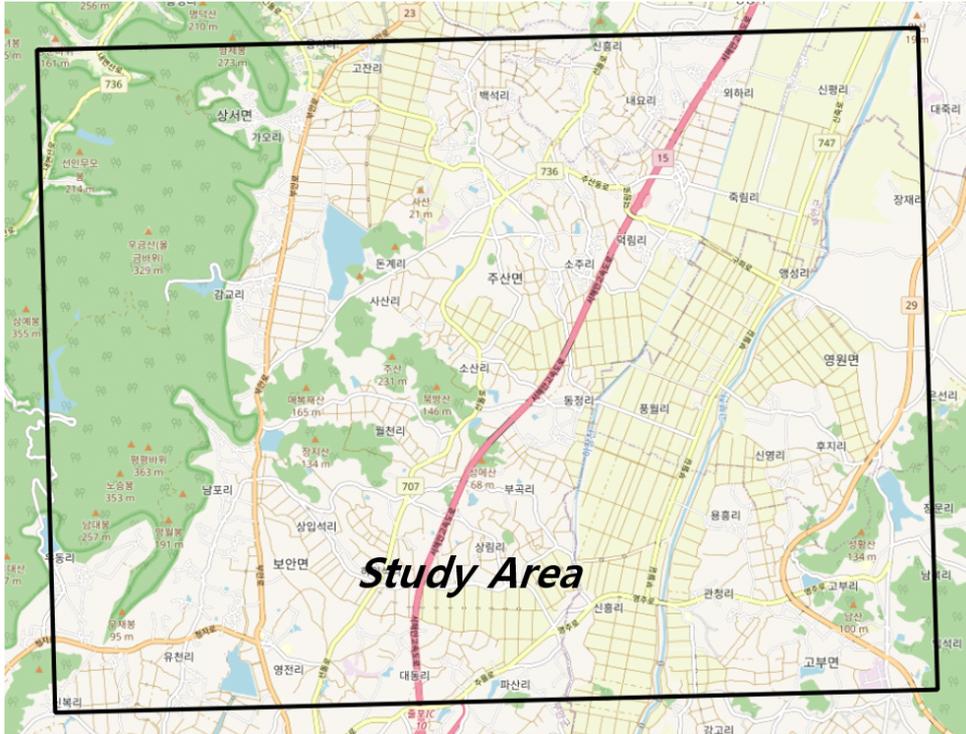
GEE에서 제공하는 Sentinel-1 ground range detected (GRD) SAR 영상은 radiometric calibration geometric correction, thermal noise removal, multilooking과 terrain correction 등의 전처리 과정을 거친 자료를 플랫폼에서 적합한 위성영상 분석 준비 데이터(Analysis Ready Data: ARD) 형식으로 변환한 것이다(Mullissa et al., 2021). Sentinel-2와 Landsat 8 광학영상의 경우는 각 밴드의 지표반사도 처리 영상을 사용하였다.

고해상도 영상인 KOMPSAT-3A와 KOMPSAT-5 위성 데이터는 외부에서 수집한 후, GEE에 asset 기능을 이용하여 업로드 하였고 연구 대상 지역

에 맞게 공간 범위를 설정해(clipping) 필요한 영상 처리 작업을 수행하였다. 추가적으로 SAR 영상인 KOMPSAT-5 데이터는 SNAP을 이용하여 Calibrate, Speckle filtering, Multilooking, Terrain Correction 등 전처리를 완료한 후 사용하였다.



[그림 3-2] 위성 데이터별 연구지역



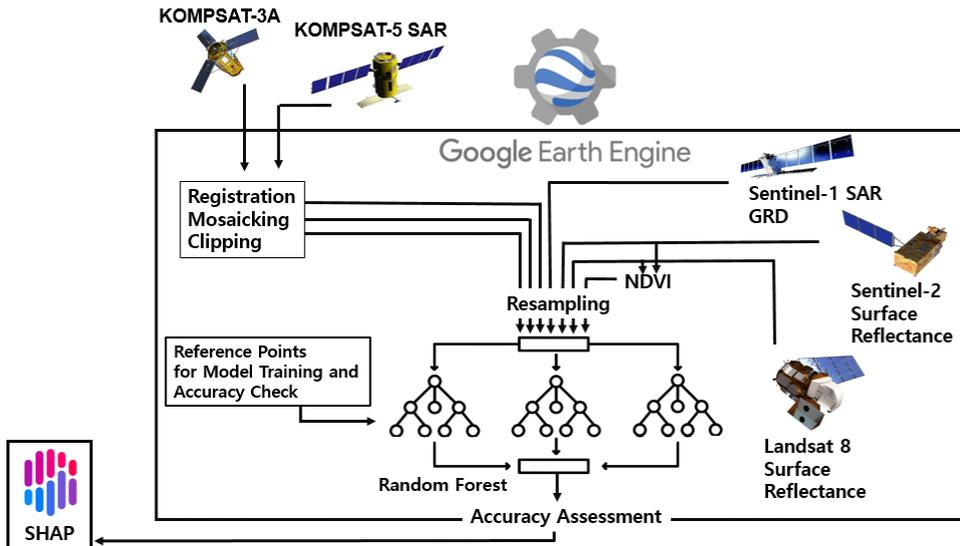
[그림 3-3] 연구지역

[표 3-1] 연구 자료

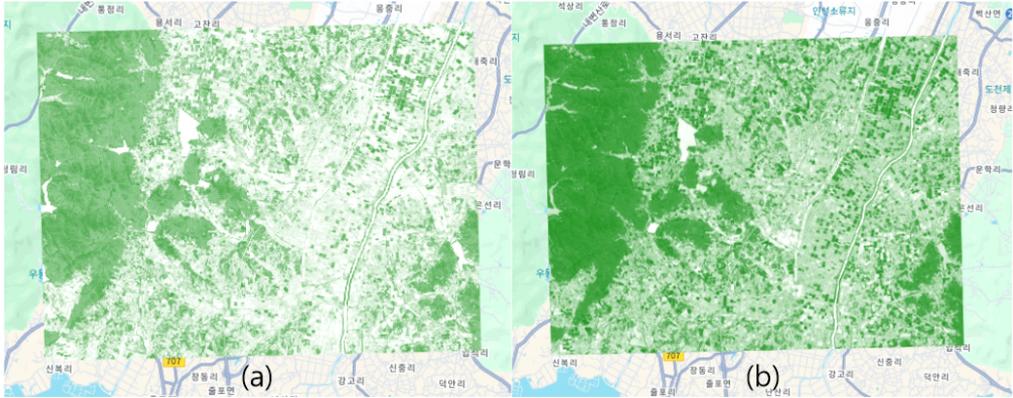
| Data sensor       | Data type                                  | Acquisition date | Source  | Resolution |
|-------------------|--|------------------|---------|------------|
| Sentinel-1        | VV   | 2022-05-05       | GEE API | 10m        |
| Sentinel-2        | R, G, B, and NIR                           | 2022-05-17       | GEE API | 10m        |
| KOMPSAT-3A        | R, G, B, and NIR                           | 2022-05-01       | KARI    | 2.8m       |
| KOMPSAT-5         | VH   | 2022-05-02       | KARI    | 1m         |
| Landsat8          | R, G, B, and NIR                           | 2022-05-04       | GEE API | 30m        |
| NDVI (Sentinel-2) | Computed by Sentinel-2 Surface Reflectance | 2022-05-17       | GEE API | 10m        |
| NDVI (Landsat 8)  | Computed by Sentinel-2 Surface Reflectance | 2022-05-04       | GEE API | 30m        |

### 제 3 절 연구 방법

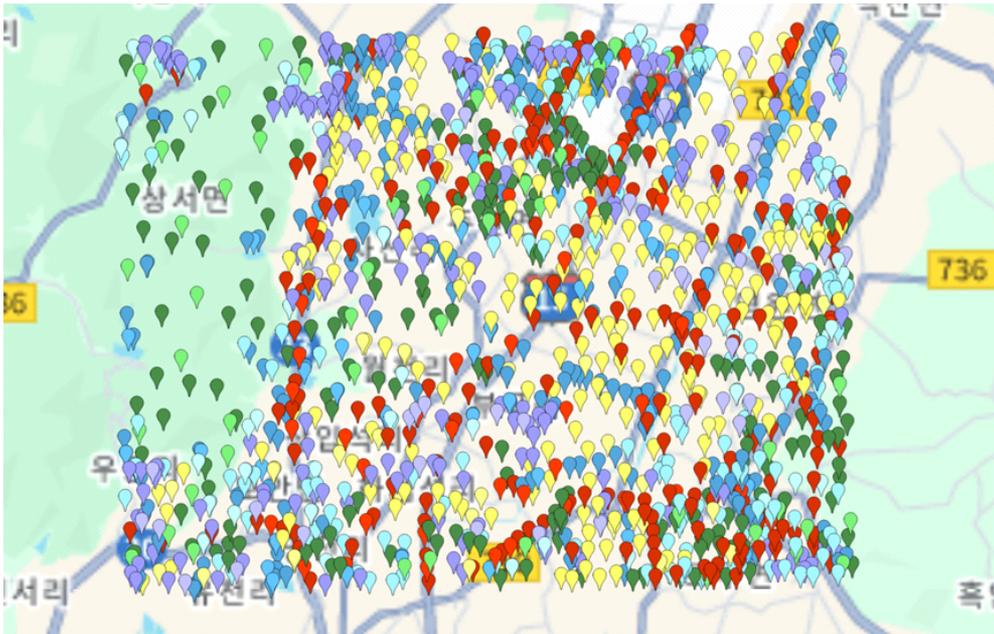
[그림 3-4]는 본 연구에서 수행한 자료 처리 과정과 전반적인 연구 흐름을 대략적으로 정리한 것이다. 연구에서 중·저해상도 영상인 Sentinel-1, Sentinel-2, Landsat 8과 고해상도 영상인 KOMPSAT-3A, KOMPSAT-5 영상을 이용하였고, Sentinel-2와 Landsat 8의 지표반사도 데이터를 이용하여 NDVI를 직접 계산한 분류 입력 자료를 [그림 3-5]에서 확인할 수 있다. 분류 대상으로는 Water Zone, Forest, Bare Zone, Veg. Zone, Road, Unclassified 총 6개의 클래스로 설정하였다. 각 클래스마다 200개의 학습 자료와 50개의 검증 자료를 설정해 전체 학습 자료는 1200개, 검증 자료는 300개로 구성하였으며 [그림 3-6]에 제시하였다. 분류 알고리즘으로는 GEE에서 활용 가능하고 기존 선행연구들에서 높은 정확도를 입증한 RF 알고리즘을 이용 하였으며, [그림 3-7]에서 자세한 내용을 확인할 수 있다. 입력 자료의 구성에 따라 총 세 가지 실험으로 진행하였다. 각 실험에 대해 분류 정확도를 평가한 후, 전체 정확도 값과 Kappa 계수의 값이 높은 세 가지 실험에 대해 SHAP기법을 통한 XAI 결과를 확인하였다.



[그림 3-4] 연구 흐름도



[그림 3-5] (a) Sentinel-2 NDVI 계산 결과,  
(b) Landsat 8 NDVI 계산 결과



[그림 3-6] 학습 자료 및 검증 자료

| Usage  | Returns    |
|--|------------|
| <code>ee.Classifier.smileRandomForest(numberOfTrees, variablesPerSplit, minLeafPopulation, bagFraction, maxNodes, seed)</code> | Classifier |

| Argument                                  | Type                          | Details  |
|---|-------------------------------|--|
| <u><a href="#">numberOfTrees</a></u>      | Integer                       | The number of decision trees to create.  |
| <u><a href="#">variablesPerSplit</a></u>  | <i>Integer, default: null</i> | <i>The number of variables per split. If unspecified, uses the square root of the number of variables.</i> |
| <u><a href="#">minLeaf Population</a></u> | <i>Integer, default: 1</i>    | Only create nodes whose training set contains at least this many points.                                   |
| <u><a href="#">bagFraction</a></u>        | <i>Float, default: 0.5</i>    | The fraction of input to bag per tree.   |
| <u><a href="#">maxNodes</a></u>           | <i>Integer, default: null</i> | The maximum number of leaf nodes in each tree. If unspecified, defaults to no limit.                       |

[그림 3-7] GEE에서 활용할 수 있는 RF 알고리즘

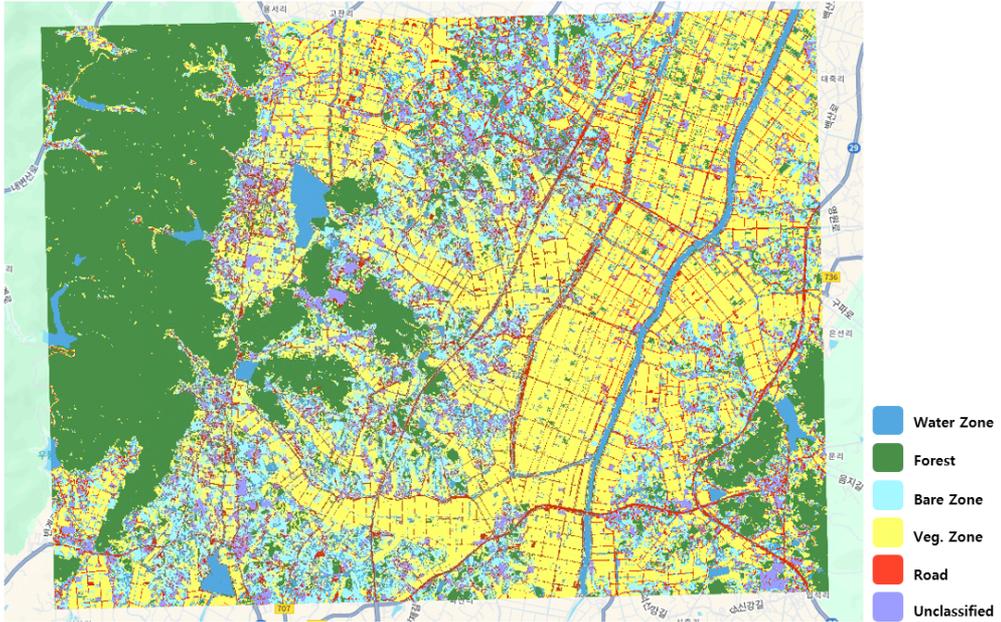
## 제 4 장 토지피복 분류 결과 및 SHAP 결과

### 제 1 절 토지피복 분류 결과

이 연구에서는 세 가지 경우에 대한 RF 알고리즘을 이용한 분류 작업을 진행하였다. 첫 번째는 Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 그리고 Sentinel-2 데이터로 NDVI를 계산한 영상 정보를 적용한 경우이다. 두 번째는 Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 그리고 Sentinel-2와 Landsat8 데이터로 NDVI를 계산한 경우, 세 번째는 두 번째 경우에 KOMPSAT-5영상을 추가한 경우이다.

첫 번째와 두 번째 경우의 전체 정확도와 Kappa 계수는 각각 89%, 0.87과 89.33%, 0.87로 두 번째 경우의 전체 정확도가 첫 번째의 경우보다 0.33% 높은 것을 확인하였다. 그리고 전체 정확도와 Kappa 계수의 값이 가장 높은 것은 91%와 0.89로 세 번째 경우인 것으로 확인할 수 있었다.

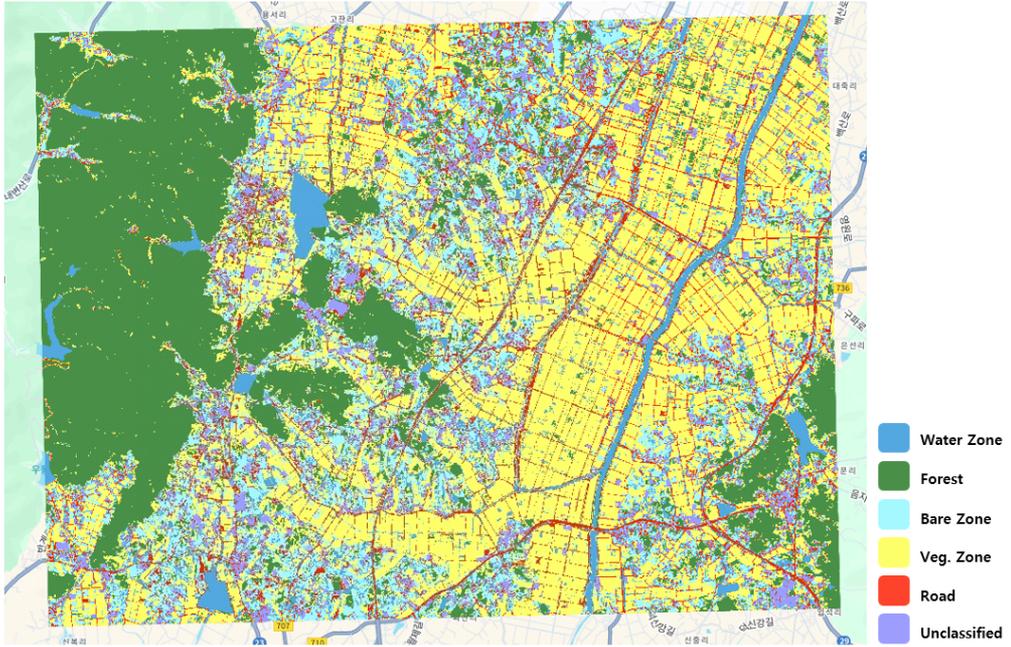
세 가지 경우에 대한 토지피복 분류 결과와 F1-Score는 각각 [그림4-1], [그림4-2], [그림4-3], [표4-1], [표4-2], [표4-3]에서 확인할 수 있으며 전체 정확도 값과 Kappa 계수로 계산한 분류 정확도는 [그림 4-4]에서 확인할 수 있다.



[그림 4-1] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) 토지피복 분류 결과

[표 4-1] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2) F1-Score

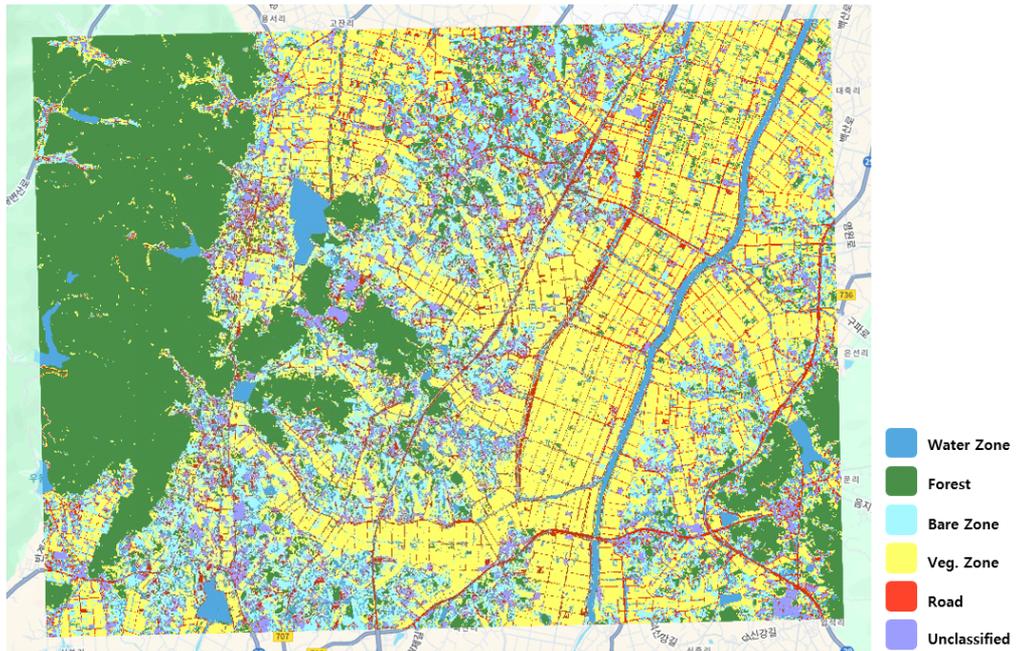
|              | Water Zone | Forest | Bare Zone | Veg. Zone | Road | Unclassified | F1-Score |
|--------------|------------|--------|-----------|-----------|------|--------------|----------|
| Water Zone   | 86         | 6      | 2         | 6         | 0    | 0            | 0.860    |
| Forest       | 2          | 96     | 0         | 2         | 0    | 0            | 0.950    |
| Bare Zone    | 4          | 0      | 86        | 6         | 2    | 0            | 0.846    |
| Veg. Zone    | 2          | 0      | 6         | 92        | 0    | 0            | 0.860    |
| Road         | 6          | 0      | 4         | 4         | 86   | 0            | 0.905    |
| Unclassified | 0          | 0      | 8         | 4         | 2    | 86           | 0.925    |



[그림 4-2] Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) 토지피복 분류 결과

[표 4-2] Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8) F1-Score

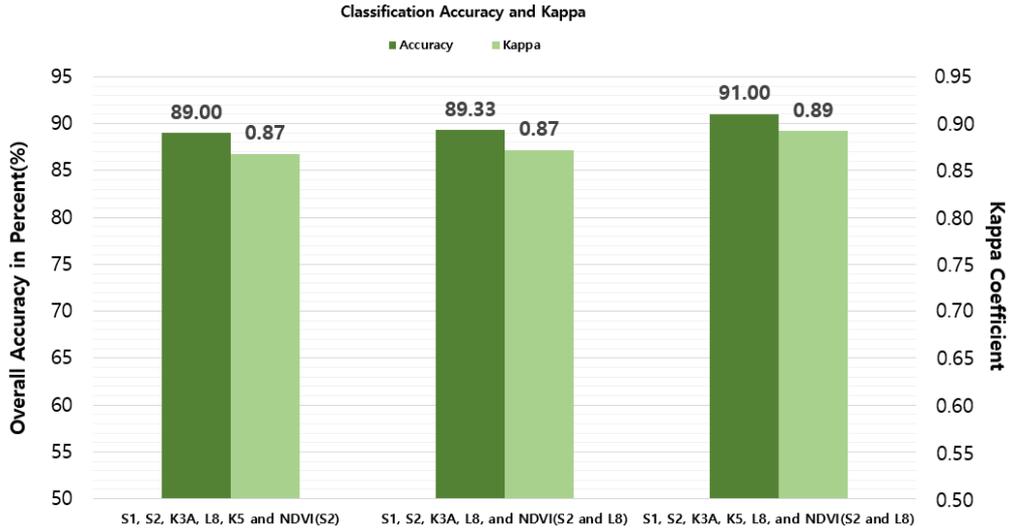
|              | Water Zone | Forest | Bare Zone | Veg. Zone | Road | Unclassified | F1-Score |
|--------------|------------|--------|-----------|-----------|------|--------------|----------|
| Water Zone   | 88         | 4      | 2         | 6         | 0    | 0            | 0.871    |
| Forest       | 2          | 94     | 0         | 4         | 0    | 0            | 0.949    |
| Bare Zone    | 6          | 0      | 88        | 4         | 2    | 0            | 0.846    |
| Veg. Zone    | 2          | 0      | 6         | 92        | 0    | 0            | 0.868    |
| Road         | 4          | 0      | 4         | 4         | 88   | 0            | 0.907    |
| Unclassified | 0          | 0      | 8         | 2         | 4    | 86           | 0.925    |



[그림 4-3] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8) 토지피복 분류 결과

[표 4-3] Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5 Landsat 8 and NDVI(Sentinel-2, Landsat 8) F1-Score

|              | Water Zone | Forest | Bare Zone | Veg. Zone | Road | Unclassified | F1-Score |
|--------------|------------|--------|-----------|-----------|------|--------------|----------|
| Water Zone   | 90         | 4      | 2         | 4         | 0    | 0            | 0.891    |
| Forest       | 2          | 96     | 0         | 2         | 0    | 0            | 0.960    |
| Bare Zone    | 2          | 0      | 92        | 4         | 2    | 0            | 0.885    |
| Veg. Zone    | 2          | 0      | 8         | 90        | 0    | 0            | 0.865    |
| Road         | 6          | 0      | 2         | 4         | 88   | 0            | 0.917    |
| Unclassified | 0          | 0      | 4         | 4         | 2    | 90           | 0.947    |



[그림 4-4] 세 가지 경우에 대한 전체 정확도 및 Kappa 계수

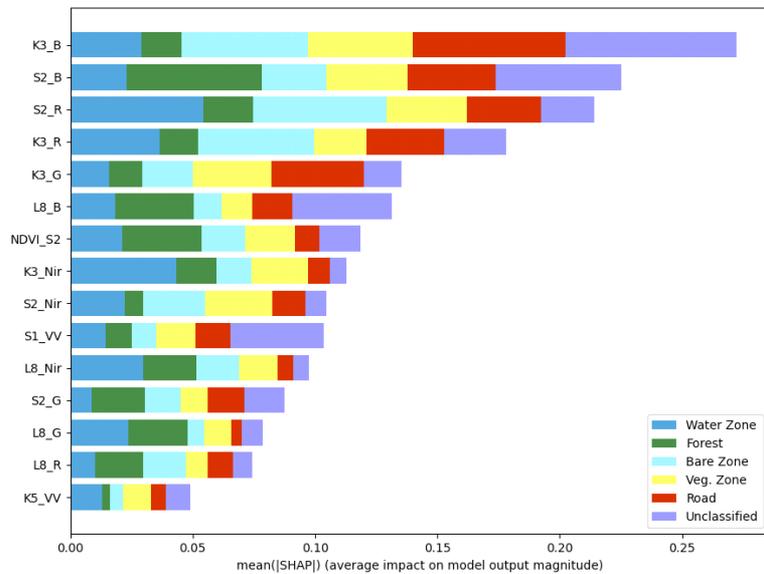
## 제 2 절 Stacked Summary Bar Plot 결과

Stack Summary Bar Plot의 경우에는 가장 영향을 많이 준 입력 자료 순서로 상단부터 나열하였다. [그림 4-5], [그림 4-6], [그림 4-7] 세 가지 경우 모두 입력 자료 중 영향을 많이 끼친 상위 네 가지 입력 자료가 KOMPSAT-3A의 B 밴드, Sentinel-2의 B 밴드, Sentinel-2의 R 밴드, KOMPSAT-3A의 R 밴드 순으로 동일하게 나타났다. 이는 위의 네 가지의 밴드가 Water Zone, Veg. Zone, Bare Zone 등 클래스 간 스펙트럼의 차이를 구분하는데 중요한 역할을 한다고 해석할 수 있다. 특히 KOMPSAT-3A의 B 밴드는 세 경우에서 모두 가장 높은 SHAP 값을 가지므로 각 클래스의 차이에 가장 민감하게 반응했다고 볼 수 있다.

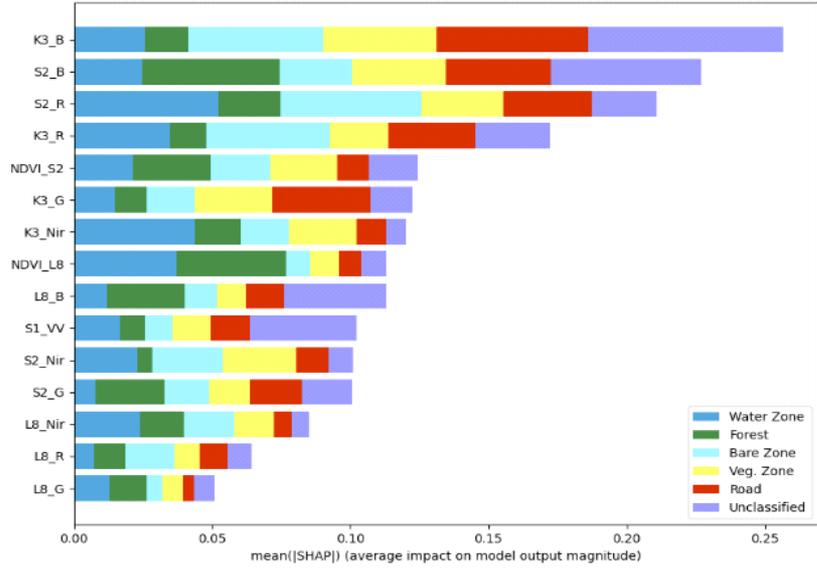
반대로 세 가지 경우에서 가장 영향을 끼치지 않은 입력 자료는 순서대로 KOMPSAT-5의 VV 밴드, Landsat 8의 G 밴드와 R 밴드인 것을 확인할 수 있다. 이는 레이더 SAR 영상이 광학 영상에 비해 미세한 스펙트럼의 차이를

설명하는데 한계가 있다고 해석될 수 있으며, Landsat 8의 밴드의 경우 해상도가 30m로 비교적 낮기 때문에 경계를 판별하는데 필요한 세부 정보가 다른 위성 데이터들에 비해 부족했기 때문이라고 볼 수 있다.

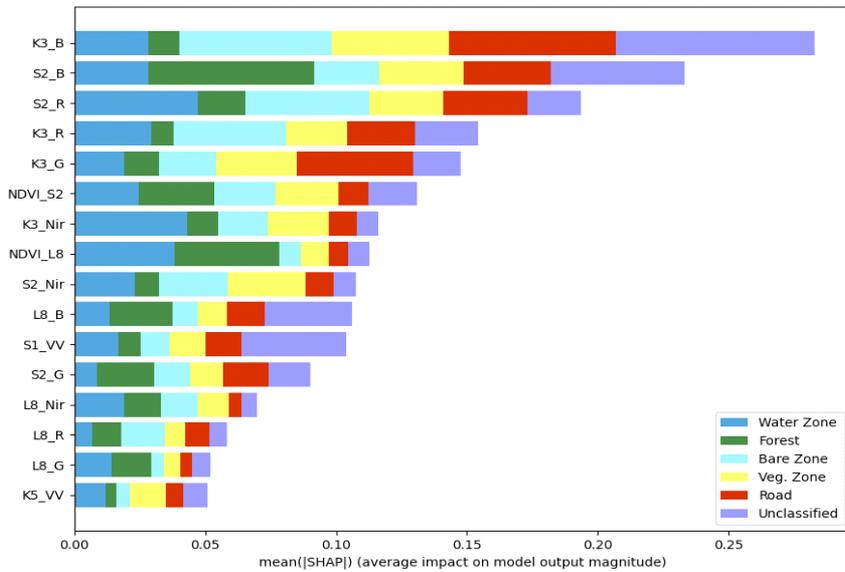
또한 전체적인 막대의 색상 구성 비율을 확인했을 때 B 밴드와 G 밴드의 비중이 Water Zone과 Forest에서 높게 나타나고, R 밴드와 NIR 밴드의 비중이 Road에서 상대적으로 높은 것을 확인할 수 있다.



[그림 4-5] Stack Summary Bar Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



[그림 4-6] Stack Summary Bar Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



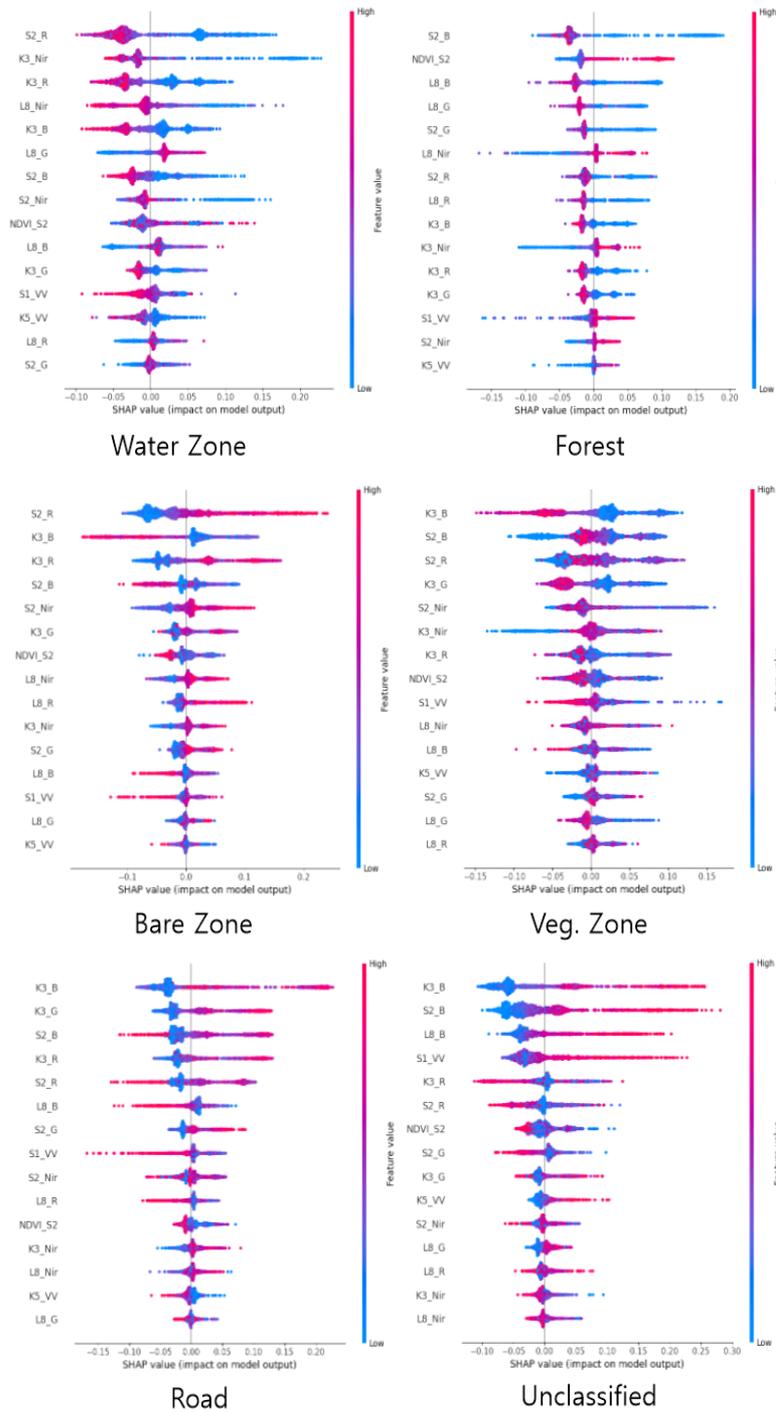
[그림 4-7] Stack Summary Bar Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5 Landsat 8 and NDVI(Sentinel-2, Landsat 8)

### 제 3 절 Bee Swarm Plot 결과

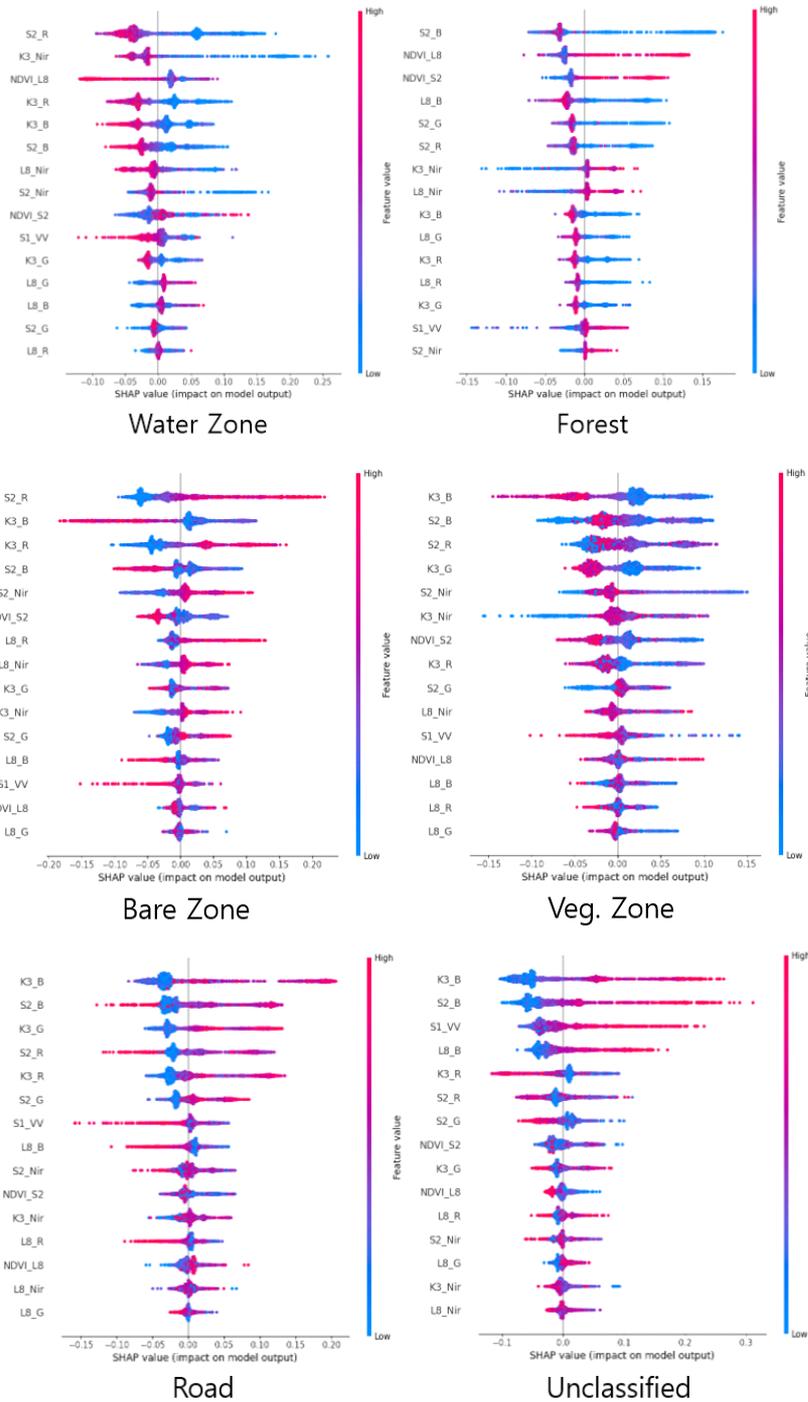
클래스에 대하여 전체 feature의 중요도와 값의 방향성을 시각화하여 제시한다. 빨간색 점은 각 클래스의 특징에서 높은 값을 나타내고, 파란색 점은 낮은 값을 나타낸다. 마찬가지로 가장 영향을 많이 준 입력 자료 순서로 상단부터 나열하였으며, 각 위성 자료의 B 밴드가 Water Zone에는 음의 상관도, Bare Zone에는 양의 상관도를 나타낸다.

Water Zone에서는 Sentinel-2의 B 밴드, R 밴드와 KOMPSAT-3A의 B 밴드의 값이 낮을수록 SHAP 값이 증가하는 특징을 보인다. 이는 수체 영역의 낮은 분광 반사율이 해당 클래스를 판별하는 주요 기여 요인인 것을 의미한다. 특히 NIR 밴드의 값들이 낮을수록 양의 SHAP 값을 가지는데, 이는 수체의 물리적 특성이 반영되어 Water Zone으로 분류될 확률을 높인다고 해석될 수 있다. Forest에서는 높은 NDVI 값일수록 양의 SHAP 값을 가진다. 반대로 B 밴드에서는 낮은 값을 가질수록 높은 SHAP 값을 나타내며, Veg. Zone은 NIR 밴드의 반사율이 높을수록 SHAP 값이 높게 나오는 것을 확인할 수 있다. 이는 식물이 광합성을 위해 가시광선을 흡수하고 근적외선은 반사하는 특성을 보여주는 결과이다. Bare Zone에서는 B 밴드를 포함한 대부분의 광학 밴드에서 높은 feature 값이 양의 SHAP 값을 가진다. 따라서 Bare Zone은 Water Zone과는 반대로 전반적으로 높은 반사율을 가진다는 것을 보여준다. Road는 R 밴드와 B 밴드가 양의 SHAP 값을 가지며, 특히 SAR 데이터의 영향력이 타 클래스에 비해 Road와 Unclassified에서 증가하는 경향을 보인다. 이는 인공 구조물의 후방산란 특성이 보조적으로 사용될 수 있다는 것을 의미한다.

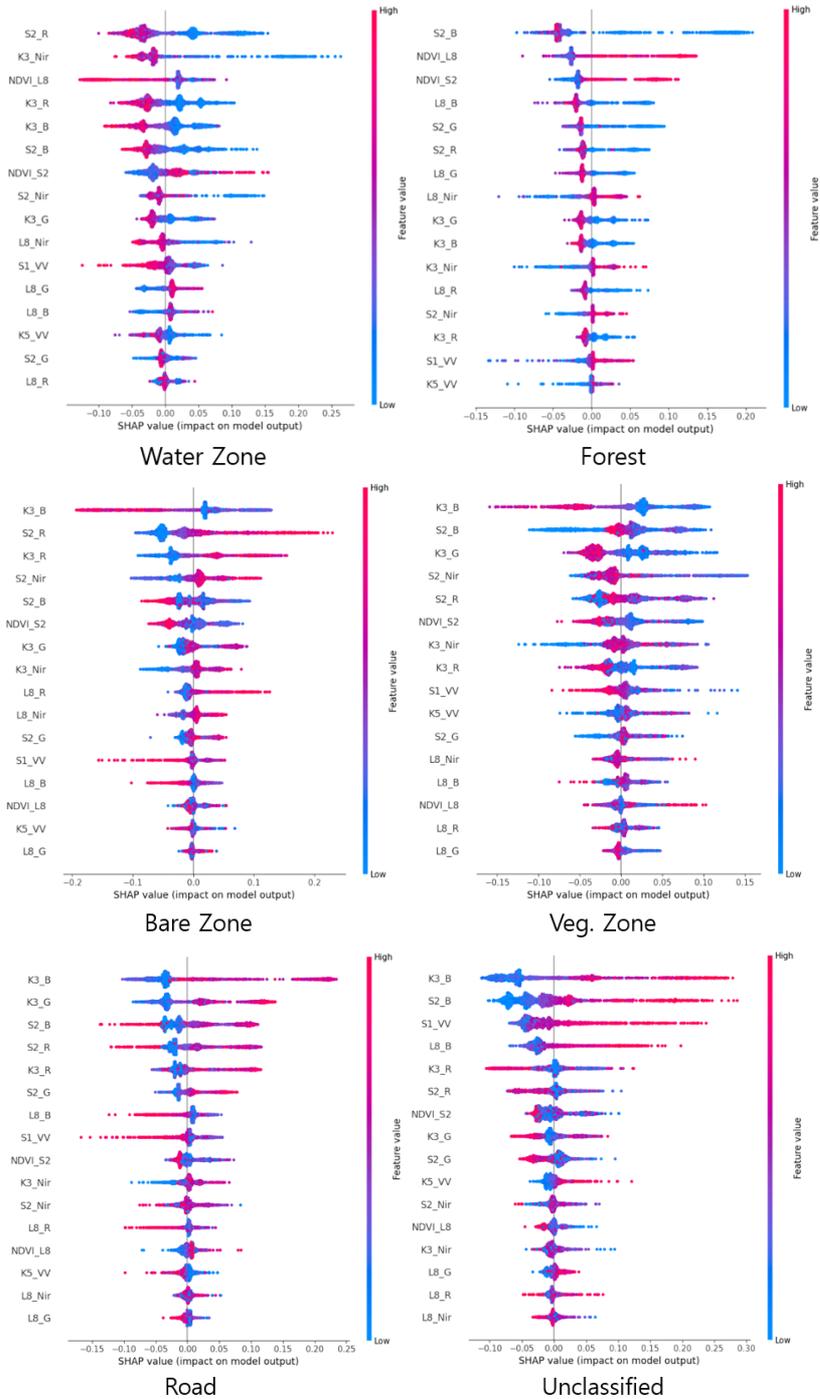
[그림 4-8], [그림 4-9], [그림 4-10] 세 가지 경우의 Bee Swarm Plot을 비교해 보았을 때, 입력 자료가 달라지더라도 클래스 분류에서 나타나는 각 밴드들의 공통점이 존재한다. 이는 모델이 실제 지표의 물리적·분광적 특성을 기반으로 분류를 수행한다는 의미를 가진다.



[그림 4-8] Bee Swarm Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



[그림 4-9] Bee Swarm Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



[그림 4-10] Bee Swarm Plot: Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8)

## 제 4 절 Dependency Plot 결과

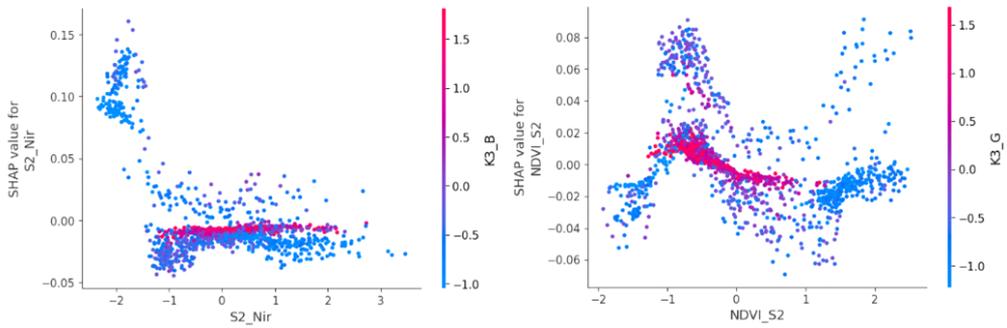
Dependency Plot에서는 x축이 주요 feature 값이며 y축은 주요 feature의 SHAP 값이다. [그림 4-11], [그림 4-12], [그림 4-13] Water Zone의 경우 KOMPSAT-3A B 밴드의 값이 0 이하일 때 SHAP 값은 0~0.075의 양수 영역에 분포하지만, 0을 넘어 증가할수록 SHAP 값은 급격하게 감소해 음수로 이동한다. 또한 Sentinel-2 NIR 밴드값이 -2 이하일 때 SHAP 값이 0.10~0.15 수준으로 높게 나타나지만 값이 -1 이상으로 증가하면 0~0.02 범위의 SHAP 값을 가진다. 이는 특히 수체의 특성인 낮은 NIR 밴드의 반사율을 의미한다. KOMPSAT-3A B 밴드의 값 중 0을 기점으로 SHAP 값이 음수로 변하고 같은 구간에서 NDVI의 데이터들이 하단에 분포하는 것 또한 식생 특성이 관찰되는 지표면을 Water Zone 클래스에서 배제함으로써 수체 판별의 신뢰도를 높이고 있음을 보여준다. 또한, Sentinel-2 NIR 밴드의 값이 -2 이하일 때 KOMPSAT-3A B 밴드가 낮은 값을 가지고 있다는 점은 수체 추출 시 두 밴드 간의 상호작용을 보여준다.

[그림 4-14], [그림 4-15], [그림 4-16] Veg. Zone에서는 NDVI의 값이 -1 부근에서 가장 많이 분포되어 있으며, NDVI 값이 0 이상으로 증가할 경우 SHAP 값은 점차 감소하는 경향을 보인다. 이는 음수에서 양수로 값이 증가할수록 SHAP 값이 감소하는 것을 보여주며, 특정 임계 범위를 벗어난 NDVI 값에 대해 판별 기여도를 낮게 평가하고 있음을 의미한다. 이 구간에서 KOMPSAT-3A G 밴드와의 상호작용을 보면, 동일한 NDVI 값 내에서도 높은 값을 가진 G 밴드의 데이터들이 높은 값의 SHAP 값을 가진 것 또한 알 수 있다. 이는 G 밴드의 특성과 식생의 특성을 보조 지표로 결합하고 있음을 보여준다. 그리고 Sentinel-2 R 밴드의 값이 1 이상으로 증가함에 따라 SHAP 결과값이 감소하는 곡선 형태의 변화를 보인다. R 밴드의 반사율이 NDVI의 특정 구간일 경우에는 기여도가 올라가지만, 해당 구간을 벗어나면 SHAP 값을 낮추며 Sentinel-2 R 밴드와 NDVI의 상호보완적인 관계를 보여준다.

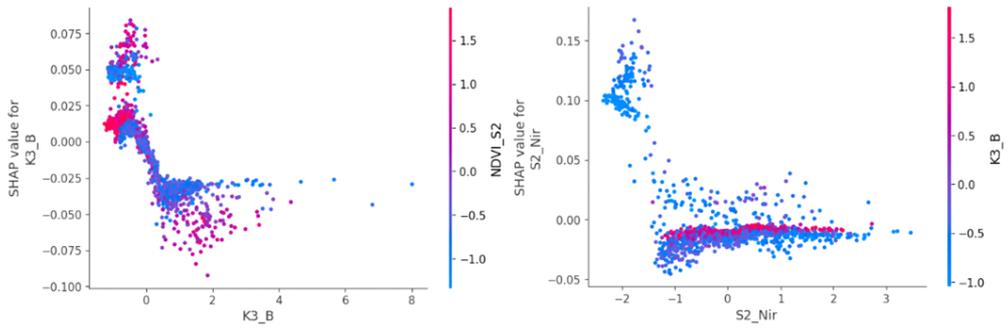
[그림 4-17], [그림 4-18], [그림 4-19] Road에서 S2\_G 값이 0 이하일

때 SHAP 값은  $-0.02$  수준에 위치하지만, 값이 증가할수록 SHAP 값은  $0.02\sim 0.07$ 까지 상승한다. 이후 특정 값 이상에서는 SHAP 값이 다시  $0\sim 0.02$  근처로 감소하는 변화를 보인다. S1\_VV 변수에서는 0 부근에서 SHAP 값이  $0\sim 0.05$  사이의 양수로 나타난다. 이는 Road 클래스의 물리적인 특성인 매끄러운 표면을 반영한 결과이며, 반대로 S1\_VV 값이 1~3 범위로 증가할 경우 SHAP 값이  $-0.05$ 에서  $-0.15$ 까지 크게 감소하는 부분은 후방산란 계수가 높게 나타나는 거친 지표면을 Road 클래스에서 제외하고 있는 것으로 분석 가능하다. 따라서 Road 클래스의 물리적인 특성을 반영했다는 것을 알 수 있으며, 이와 반대로  $-1$  이하의 음의 구간에서 SHAP 값이 0 근처에 밀집되는 것은 Road 클래스에서 기여도가 낮다고 해석할 수 있다. KOMPSAT-5 VV 밴드는 SHAP 값의 변화가 거의 없다는 특징을 가진다.

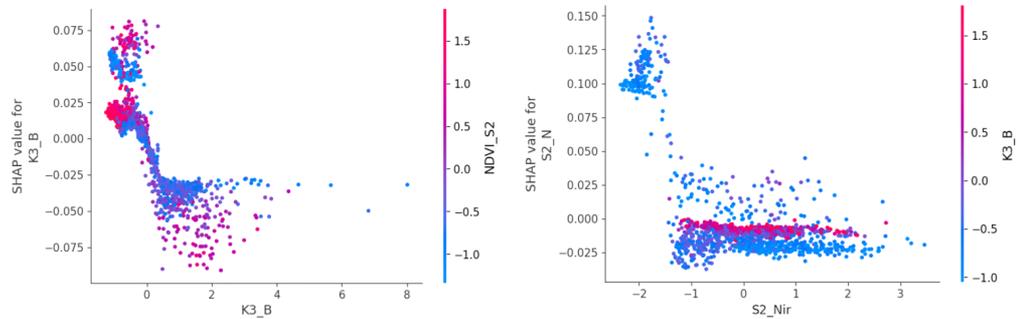
Dependency Plot에서는 단순한 선형 관계가 아닌 구간별로 SHAP 값의 증가 또는 감소하는 것을 볼 수 있었다. 이는 각 클래스에서 입력 자료가 가지는 영향이 다르고, feature의 상호작용과 구간별 기여도 변화를 시각화하여 확인할 수 있음을 나타낸다.



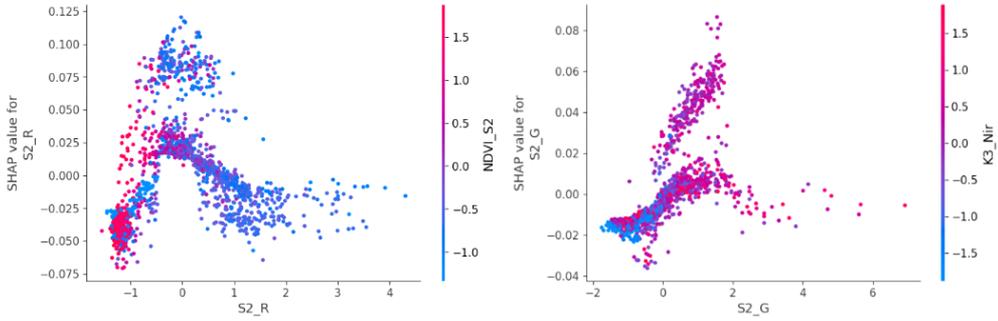
[그림 4-11] Dependency Plot(Water Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



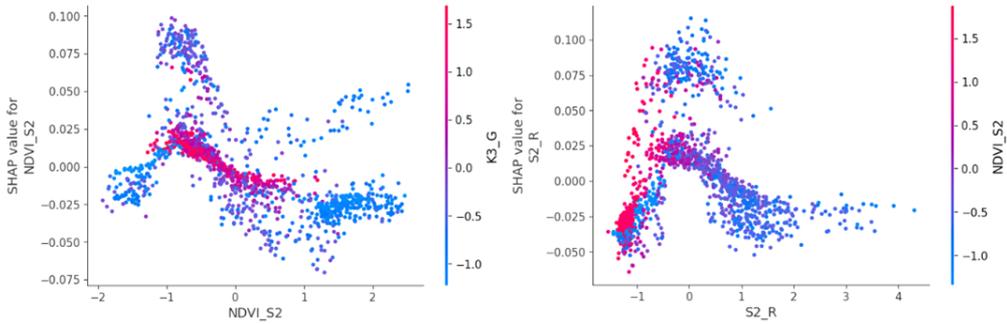
[그림 4-12] Dependency Plot(Water Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



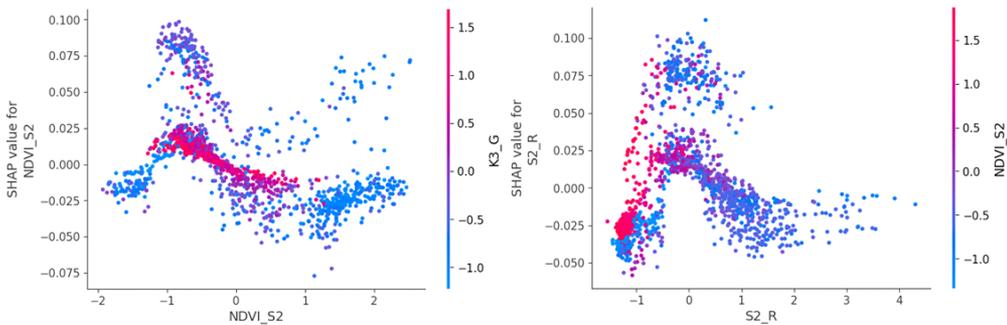
[그림 4-13] Dependency Plot(Water Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



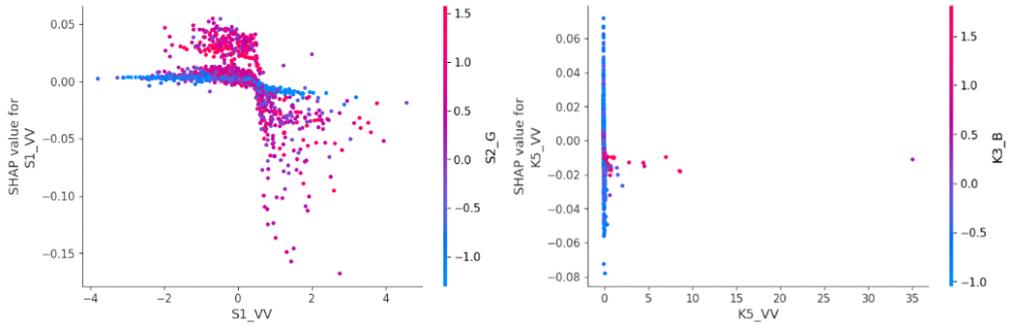
[그림 4-14] Dependency Plot(Veg. Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



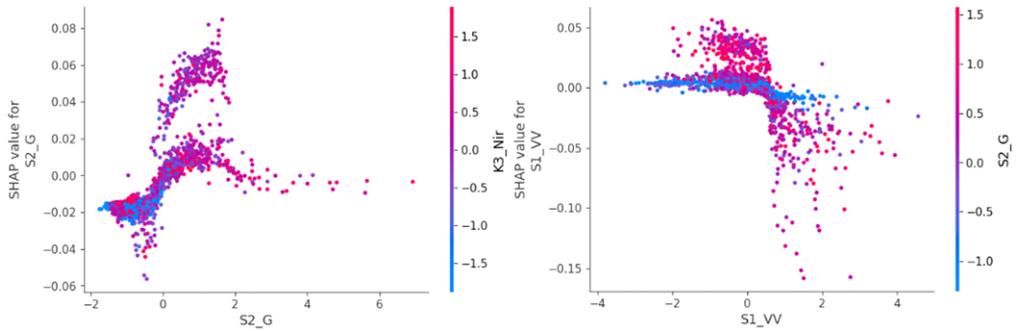
[그림 4-15] Dependency Plot(Veg. Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



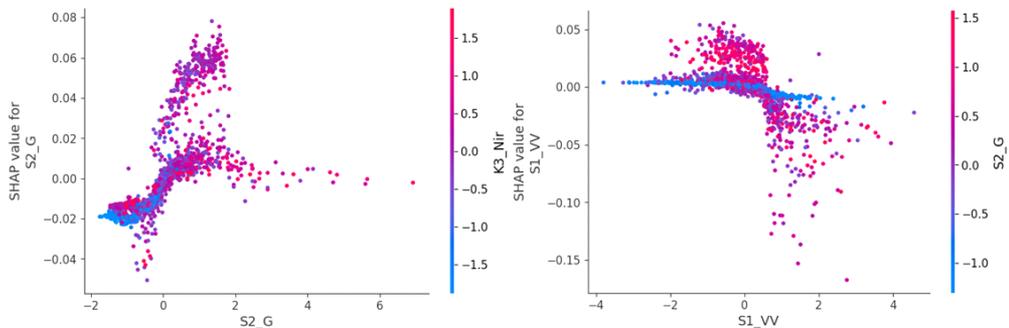
[그림 4-16] Dependency Plot(Veg. Zone): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



[그림 4-17] Dependency Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



[그림 4-18] Dependency Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)

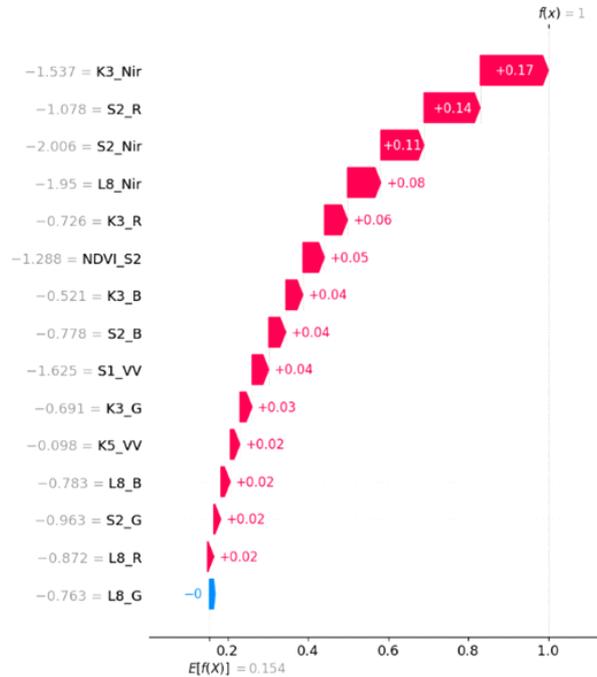


[그림 4-19] Dependency Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8)

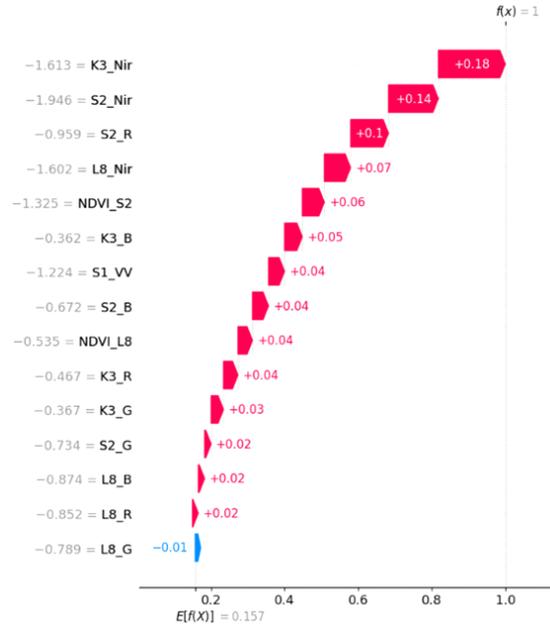
## 제 5 절 Waterfall Plot 결과

[그림 4-20], [그림 4-21], [그림 4-22], [그림 4-23], [그림 4-24], [그림 4-25] Water와 Road에 대해 세 가지 경우의 Waterfall Plot을 시각화 하였다. 위에서부터 아래의 순서대로 어떤 입력 자료가 양의 기여 또는 음의 기여를 했는지 확인하였다. Waterfall Plot은  $E[f(x)]$ 로 표시되는 모델의 평균 예측값에서 개별 각 입력 자료의 SHAP 값이 누적되며 최종 예측값  $f(x)=1$ 에 도달하는 구조를 가진다.

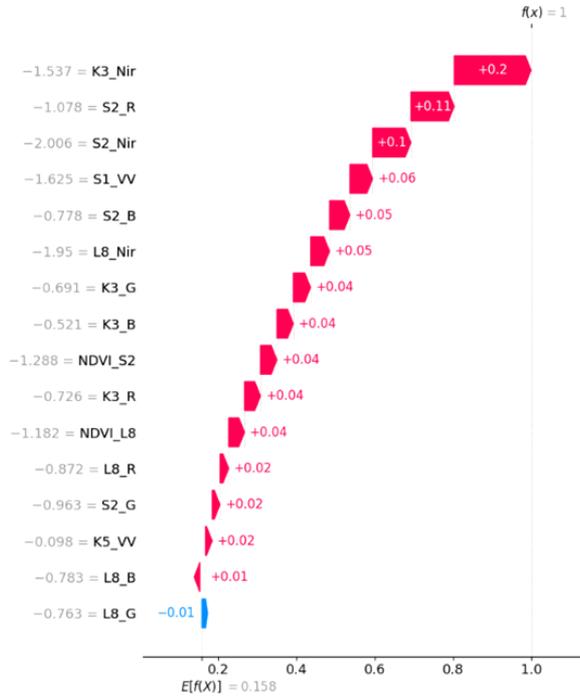
세 가지 경우에 대한 결과에서 상위 5~6개 밴드는 거의 동일하다. 이는 실제로 예측을 진행할 때 어떤 자료가 분류에 영향을 가장 많이 미치는지를 시각적으로 확인할 수 있으며, 연구를 진행한 모든 경우에서 순서가 매우 유사하게 나타났다는 것을 알 수 있다.



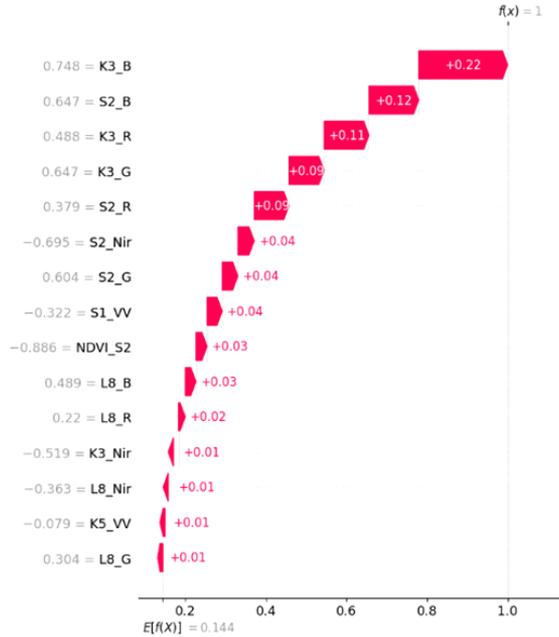
[그림 4-20] Waterfall Plot(Water): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



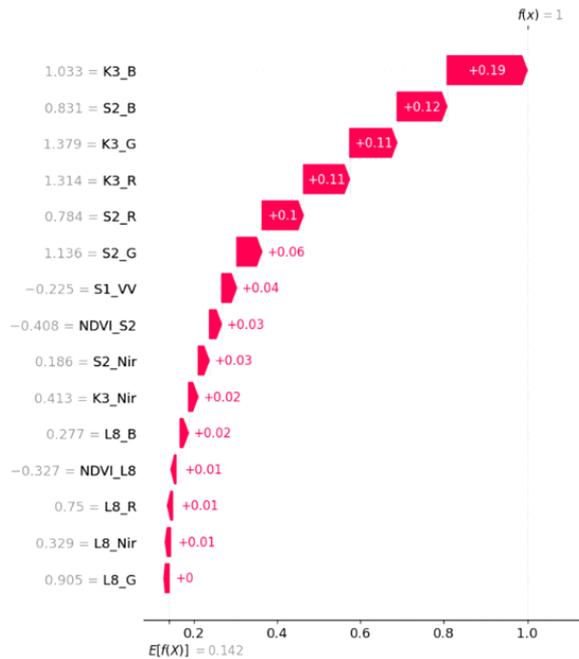
[그림 4-21] Waterfall Plot(Water): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



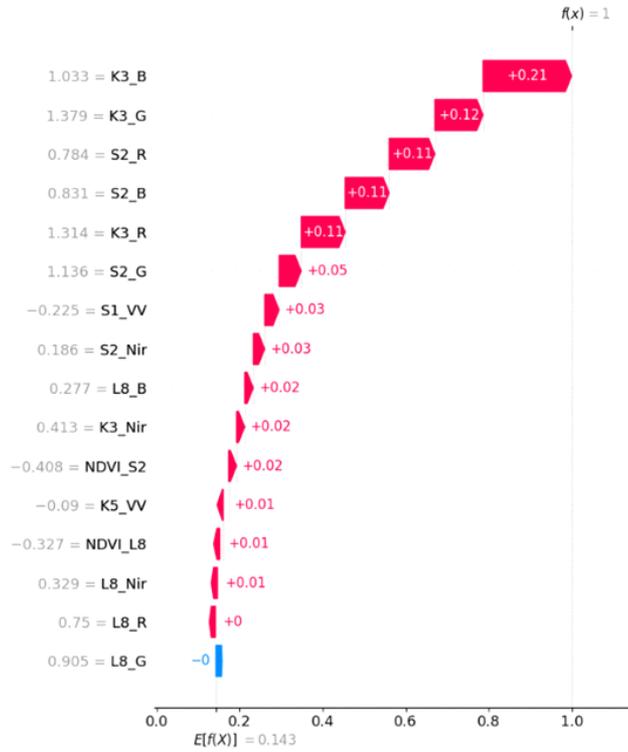
[그림 4-22] Waterfall Plot(Water): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



[그림 4-23] Waterfall Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2)



[그림 4-24] Waterfall Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, Landsat 8 and NDVI(Sentinel-2, Landsat 8)



[그림 4-25] Waterfall Plot(Road): Sentinel-1, Sentinel-2, KOMPSAT-3A, KOMPSAT-5, Landsat 8 and NDVI(Sentinel-2, Landsat 8)

## 제 5 장 결 론

토지피복 지도는 지표면의 물리적, 인위적 특성과 토지 이용 형태를 나타내는 핵심 공간정보로 환경 모니터링, 도시계획, 농업 및 수자원 관리, 생태계 관리 등 다양한 분야에서 활용되는 중요한 지표이다. 토지의 자연적 상태와 농경지, 도시지역, 산업시설 등의 인간의 사회·경제적 활동에 따른 토지 이용 형태를 나타내 그 지역의 생활 및 경제 환경 파악 및 지표 변화를 장기적으로 평가할 수 있는 기반자료로 기후변화와 재난 예측, 수문학적 분석 등 정책 및 과학적 의사결정에 이용될 수 있다.

클라우드 기반 위성영상정보 플랫폼인 GEE에서 제공하는 위성영상과 지리정보데이터의 종류와 개수가 점점 증가하고 있다. 이를 실시간으로 처리하고 가공할 수 있는 함수들이 계속 개발되며 다양한 처리 기법을 활용한 연구 사례가 점차 증가하는 추세이다.

특히 이번 연구에서는 GEE에서 제공하지 않는 고해상도 광학영상인 KOMPSAT-3A와 레이더 SAR 영상인 KOMPSAT-5 위성영상정보를 추가로 입력하여 GEE에서 제공하는 광학 및 SAR 위성영상과 함께 토지피복 분류를 진행하였다. 추가로 GEE에서 제공하는 지표 반사도 전처리를 완료한 Sentinel-2와 Landsat 8 위성영상으로 직접 NDVI 계산을 진행해 적용하였다. 총 세 가지의 경우의 토지피복 분류 결과를 확인하였고, 전체 정확도와 Kappa 계수를 확인하였다.

SHAP 결과로 Stacked Summary Bar Plot, Bee Swarm Plot, Dependency Plot, Waterfall Plot까지 총 4가지 경우를 확인했으며, 연구를 진행한 세 가지의 경우 모두 입력 자료가 달라졌음에도 불구하고 각각의 SHAP 결과가 비슷하게 나타났다. 특히 Water Zone에서는 NIR 밴드가 예측 값을 증가시키는 데 가장 영향력이 높았으며, Veg. Zone은 NDVI의 값이 높을수록 양의 SHAP 값이 감소하는 경향이 있었다.

이렇듯 입력 데이터가 어떻게 영향을 미치는지 블랙박스로 다루어지는 문제가 있는 기계학습 및 딥러닝에 대해 대표적인 XAI 방법 중 하나인 SHAP 분석을 진행하였다. 그 결과 분류에 영향을 미치는 입력 자료의 기여도를 정

량적·시각적으로 확인함으로써 블랙박스 모델을 해석할 수 있는 근거를 제시할 수 있다는 점에서 중요한 의의가 있다.

# 참 고 문 헌

## 1. 국내문헌

- Molnar, C. (2022). 『Interpretable Machine Learning: A Guide for Making Black Box Models Explainable』. Durham, NC: Lulu Press.
- Shapley, L. S. (1953). A value for n-person games. 수록처: 『Contributions to the Theory of Games, Vol. II』 (pp. 307–317). Princeton, NJ: Princeton University Press.
- Google Earth Engine Team. (2025). 『Supervised Classification on Earth Engine: Developers Guide』. Mountain View, CA: Google LLC.

## 2. 국외문헌

- Aji, M. A. P., Kamal, M., & Farda, N. M. (2023). Mangrove species mapping through phenological analysis using random forest algorithm on Google Earth Engine. *Remote Sensing Applications: Society and Environment*, 30, 100978. <https://doi.org/10.1016/j.rsase.2023.100978>
- Ahmed, R., Zafor, M. A., & Trachte, K. (2024). Land-use and land-cover changes in Cottbus City and Spree-Neisse District, Germany, in the last two decades: A study using remote sensing data and Google Earth Engine. *Remote Sensing*, 16(15), 2773. <https://doi.org/10.3390/rs16152773>
- Amani, M., Ghorbanian, A., Ahmadi, S. A., Kakooei, M., Moghimi, A., Mirmazloumi, S. M., et al. (2020). Google Earth Engine cloud computing platform for remote sensing big data applications: A

- comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5326–5350. <https://doi.org/10.1109/JSTARS.2020.3021052>
- Amindin, A., Siamian, N., Kariminejad, N., Clague, J. J., & Pourghasemi, H. R. (2024). An integrated GEE and machine learning framework for detecting ecological stability under land use/land cover changes. *Global Ecology and Conservation*, 53, e03010. <https://doi.org/10.1016/j.gecco.2024.e03010>
- Alshehri, B., Zhang, Z., & Liu, X. (2025). A review of Google Earth Engine for land use and land cover change analysis: Trends, applications, and challenges. *ISPRS International Journal of Geo-Information*, 14(11), 416. <https://doi.org/10.3390/ijgi14110416>
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, 10(7), e0130140. <https://doi.org/10.1371/journal.pone.0130140>
- Cai, G., Ren, H., Yang, L., Zhang, N., Du, M., & Wu, C. (2019). Detailed urban land use land cover classification at the metropolitan scale using a three-layer classification scheme. *Sensors*, 19(14), 3120. <https://doi.org/10.3390/s19143120>
- Fentaw, A. E., & Abegaz, A. (2024). Analyzing Land Use/Land Cover Changes Using Google Earth Engine and Random Forest Algorithm and Their Implications to the Management of Land Degradation in the Upper Tekeze Basin, Ethiopia. *The Scientific World Journal*, 2024, 3937558. <https://doi.org/10.1155/2024/3937558>
- Gevaert, C. M. (2022). Explainable AI for Earth observation: A review

- including societal and regulatory perspectives. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102869. <https://doi.org/10.1016/j.jag.2022.102869>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 93. <https://doi.org/10.1145/3236009>
- Höhn, A., Mirus, B. B., Dedecker, A. P., Bennett, G. L., Handwerger, A. L., & Kirschbaum, D. B. (2025). A spatio-temporal dataset for satellite-based landslide inventory automation. *Scientific Data*, 12, Article 22. <https://doi.org/10.1038/s41597-025-06167-2>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. <https://doi.org/10.48550/arXiv.1705.07874>
- Mutanga, O., & Kumar, L. (2019). Google Earth Engine applications in the mapping of forest degradation and biomass. *Remote Sensing*, 11(7), 754. <https://doi.org/10.3390/rs11070754>
- Räuber, T., Ho, A., Casper, S., & Hadfield-Menell, D. (2022). Toward transparent AI: A survey on interpreting the inner structures of deep neural networks. *Artificial Intelligence*, 307, 103704. <https://doi.org/10.1016/j.artint.2022.103704>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining, 1135–1144. <https://doi.org/10.1145/2939672.2939778>

- Rodríguez-Pérez, R., & Bajorath, J. (2020). Interpretation of machine learning models using Shapley values: Application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34, 1013–1026. <https://doi.org/10.1007/s10822-020-00314-0>
- Seitkazy, M., Beisekenov, N., Taukebayev, O., Zulpykharov, K., Tokbergenova, A., Duisenbayev, S., et al. (2024). Forecasting land use dynamics in Talas District, Kazakhstan, using Landsat data and the Google Earth Engine (GEE) platform. *Sustainability*, 16(14), 6144. <https://doi.org/10.3390/su16146144>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626. <https://doi.org/10.1109/ICCV.2017.74>
- Temenos, A., Temenos, N., Kaselimi, M., Doulamis, A., & Doulamis, N. (2023). Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. *IEEE Geoscience and Remote Sensing Letters*, 20, 1–5. <https://doi.org/10.1109/LGRS.2023.3251652>
- Velastegui-Montoya, A. E., Herrera-Yunga, G. R., Lapo-Villacís, N. C., Beltrán-Aguirre, J. L., & Berrezueta-Monsalve, J. A. (2023). Google Earth Engine (GEE): A global analysis and future trends. *Remote Sensing*, 15(14), 3675. <https://doi.org/10.3390/rs15143675>
- Vizzari, M. (2022). PlanetScope, Sentinel-2, and Sentinel-1 data integration for object-based land cover classification in Google Earth Engine. *Remote Sensing*, 14(11), 2628. <https://doi.org/>

10.3390/rs14112628

- Wang, Y., Sun, Y., Cao, X., Wang, Y., Zhang, W., & Cheng, X. (2023). A review of regional and global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206, 311–334. <https://doi.org/10.1016/j.isprsjprs.2023.11.014>
- Xu, Z., Ma, L., Zhang, X., Li, M., Wang, Y., & Liu, D. (2022). Cloud-based storage and computing for remote sensing big data: A review. *International Journal of Digital Earth*, 15(1), 78–106. <https://doi.org/10.1080/17538947.2022.2115567>

## ABSTRACT

### Land Cover Classification by Machine Learning in Google Earth Engine with KOMPSAT-3/5 Satellite Imagery and SHAP-Based Results Analysis

Lee, Jeong-Hee

Major in Convergence Security

Dept. of Convergence Security

The Graduate School

Hansung University

Recently, research utilizing machine learning algorithms provided within the Google Earth Engine(GEE) platform has increased substantially. In this study, low to medium satellite resolution satellite data provided by GEE through the application programming interface (API) and high-resolution satellite data not directly provided by GEE. Sentinel-2, KOMPSAT-3A, and Landsat 8, which are optical images that provide spectral information, and Sentinel-1 and KOMPSAT-5 data, which are radar-based SAR images that are not affected by clouds or atmospheric conditions, were used together. In addition, NDVI (Normalized Difference Vegetation Index) calculated using surface reflectance data Sentinel-2 and Landsat-8 was added, and land use/land cover (LULC) classification was

performed using the Random Forest algorithm. The resulting LULC classification was evaluated by comparing the overall accuracy and Kappa coefficient, while the F1-Score was employed as a complementary performance metric. The SHAP method was then applied to interpret the classification model. As an explainable machine learning technique, SHAP is particularly suitable for analyzing complex models that incorporate numerous input variables and multiple target classes, and it can be more effective than deep learning approaches that primarily focus on single-object extraction. Based on the classification outcomes, SHAP analysis was conducted to identify the most influential features for each land-cover type, demonstrating its usefulness in understanding the contribution of multi-sensor inputs. This study highlights that explainable machine learning techniques, such as SHAP, can serve as powerful tools for interpreting and monitoring surface environmental conditions, particularly when working with multi-sensor high-resolution satellite imagery.

**【Key words】** KOMPSAT-3A, KOMPSAT-5, Google Earth Engine, SHAP, Image classification, Random Forest