



Application study of random forest method based on Sentinel-2 imagery for surface cover classification in rivers - A case of Naeseong Stream -

An, Seonggi^a · Lee, Chanjoo^{b*} · Kim, Yongmin^c · Choi, Hun^d

^aStudent Researcher, Department of Land, Water and Environment Research, Korea Institute of Civil Engineering and Building Technology, Goyang, Korea

^bSenior Researcher, International Cooperation & Public Relations Division, Korea Institute of Civil Engineering and Building Technology, Goyang, Korea

^cMaster, Geography Major of Department of Social Studies Education, Seoul National University, Seoul, Korea

^dPh.D, Social Education, Chonnam National University, Gwangju, Korea

Paper number: 24-018

Received: 18 March 2024; Revised: 3 May 2024; Accepted: 3 May 2024

Abstract

Understanding the status of surface cover in riparian zones is essential for river management and flood disaster prevention. Traditional survey methods rely on expert interpretation of vegetation through vegetation mapping or indices. However, these methods are limited by their ability to accurately reflect dynamically changing river environments. Against this backdrop, this study utilized satellite imagery to apply the Random Forest method to assess the distribution of vegetation in rivers over multiple years, focusing on the Naeseong Stream as a case study. Remote sensing data from Sentinel-2 imagery were combined with ground truth data from the Naeseong Stream surface cover in 2016. The Random Forest machine learning algorithm was used to extract and train 1,000 samples per surface cover from ten predetermined sampling areas, followed by validation. A sensitivity analysis, annual surface cover analysis, and accuracy assessment were conducted to evaluate their applicability. The results showed an accuracy of 85.1% based on the validation data. Sensitivity analysis indicated the highest efficiency in 30 trees, 800 samples, and the downstream river section. Surface cover analysis accurately reflects the actual river environment. The accuracy analysis identified 14.9% boundary and internal errors, with high accuracy observed in six categories, excluding scattered and herbaceous vegetation. Although this study focused on a single river, applying the surface cover classification method to multiple rivers is necessary to obtain more accurate and comprehensive data.

Keywords: Random Forest, Vegetation Classification, Remote Sensing, Surface Cover, Riparian zone

하천 내 지표 피복 분류를 위한 Sentinel-2 영상 기반 랜덤 포레스트 기법의 적용성 연구 - 내성천을 사례로 -

안성기^a · 이찬주^{b*} · 김용민^c · 최훈^d

^a한국건설기술연구원 수자원하천연구본부 학생연구원, ^b한국건설기술연구원 대외협력홍보본부 수석연구원,

^c서울대학교 사회교육학과 지리전공 석사, ^d전남대학교 사회교육학과 박사

요 지

하천 공간의 지표 피복 현황 파악은 하천 관리 및 홍수 재해 예방에 필수적이다. 기존 조사 방법은 전문가에 의한 식생 판독을 통한 식생도 작도 방법이나 식생지수를 활용하는 방법이 활용되어 왔으나, 역동적으로 변화하는 하천 환경을 반영하기에 한계가 있다. 이러한 배경에서 본 연구는 내성천을 대상으로 위성영상 자료를 활용한 랜덤 포레스트 기법을 활용하여 다수 연도의 하천 내 식생 분포를 파악하고, 적용성을 검토하였다. 원격탐사 자료 Sentinel-2 위성 영상을 사용하였으며, 지상 참값(ground truth)은 2016년 내성천 지표 피복 자료를 활용하였다. 랜덤 포레스트 머신러닝 알고리즘을 활용하여 미리 선정된 10개 샘플링 영역으로부터 분류군 별로 1,000개의 표본을 추출하여 훈련 및 검증하였으며, 민감도 분석, 연도별 지표 피복 분석, 정확도 분석을 통하여 적용성을 평가하였다. 연구 결과, 검증 자료 기반의 정확도는 85.1%로 나타났다. 트리 수, 샘플 수, 하천 구역에 대한 민감도 분석 결과, 각각 30개, 800개, 하류에서 효율성이 높았다. 지표 분류 유형은 6개 항목에서 높은 정확도를 보여 지표 피복 분류 결과가 실제 하천 환경을 잘 반영하는 것으로 나타났다. 정확도 분석 결과, 전체 샘플 중 14.9%의 경계오류와 내부오류를 확인하였으며, 지표 피복 분류 중 산발 식생과 초본 식생을 제외한 항목들은 높은 정확도를 보였다. 본 연구에서는 단일 하천을 대상으로 적용하였지만, 보다 정확하고 많은 자료의 구축을 위해서는 다수의 하천에 대해 지표 피복 분류 기법의 적용이 요구된다.

핵심용어: 랜덤 포레스트, 식생 분류, 원격 탐사, 지표 피복, 수변

*Corresponding Author. Tel: +82-31-910-0036

E-mail: c0gnitum@kict.re.kr (Lee, Chanjoo)

1. 서론

최근 중위도 몬순 지역의 여러 하천에서 식생 활착이 발생하여 수변의 경관이 변화되고 있다. 식생은 홍수터뿐만 아니라 사주와 저수시에 노출되는 하상에 활착하였고 그 결과 과거의 소위 백색 하천이 소위 녹색 하천으로 변화되고 있다. 하천에서 식생 활착은 주로 댐이나 가뭄에 의한 홍수 유량의 감소가 원인이 되어 발생하지만(Williams, 1978; Johnson, 1994; Choi *et al.*, 2005; Lee and Kim, 2018), 인류세에 이루어져 온 유역의 시스템적인 변화에도 기인한다(Lee *et al.*, 2023).

식생 활착은 하천 경관을 변화시킬 뿐만 아니라 하천의 지형과 형성 작용을 변화시킨다. 그것은 잠재적인 홍수 위험을 증가시킬 뿐만 아니라 하천 경관과 서식처의 다양성을 증가시키는 생태적인 기능을 갖게 된다(Gurnell, 2014; Zanetti *et al.*, 2016, Rood *et al.*, 2019). 그러므로 지형학적 홍수터의 집약적인 이용으로 인해 제방으로 하천 공간이 제약되어 있는 현대의 하천을 관리하기 위해서는 상반되는 이 두 목표 - 홍수 관리와 생태계 보전을 조화시키는 것이 중요한 이슈가 되어왔다.

하천 관리의 목적을 달성하기 위해서는 초본 및 목본 식생을 포함하여 하천 공간 안에 있는 지표 피복의 분포 현황에 대한 지식이 우선적으로 필요하다. 그런데 기존의 하천 식생 분포 조사는 식생 전문가가 항공사진 등에서 현생 식생을 판독하고 식생도를 작도하는 작업에 의존하거나(Woo *et al.*, 2010; KICT, 2013; Lee *et al.*, 2022) 식생지수를 활용하는 방법을 사용해 왔다(An *et al.*, 2018; Lee and Cho, 2023). 최근 10여년간 장기 모니터링이 수행되어 온 내성천의 경우에도 수작업으로 작도한 지표피복도가 식생 면적 변화 분석에 활용된 바 있다(Lee *et al.*, 2019a). 하지만 이러한 방법은 활발하게 변화하는 하천 식생 분포 변화를 추적하는 데 한계가 있다. 산지에서는 동일 유형의 식생이 비교적 넓은 범위에 분포하고 장기간 식생의 분포 상태가 유지되는데 비해 하천은 홍수에 의한 교란과 나지에서의 식생 활착이 반복된다. 또한 지표 피복의 복잡성이 크고 공간 범위가 넓은 하천에 대한 수동 판독 작업은 비용과 시간이 많이 소요된다.

머신러닝에 의한 분류 기법은 고해상도 위성 영상의 보급과 하천 분야에서 지표 피복 분류의 활용성을 높이고 있다. 특히 페타 바이트의 공간 데이터를 보유한 클라우드 기반 GIS 플랫폼인 구글어스엔진의 출시로 인해 그 활용성이 더욱 높아지고 있다. 이와 관련하여 Geerling *et al.* (2021)은 구글어스엔진과 Sentinel-2 영상을 기반으로 네덜란드 Rhine강 수변 지역에 대한 지표 피복을 분류하는 운영 시스템을 개발한 바 있으며, Harezlak *et al.* (2020)은 35년의 TM 위성영상 자료를

기반으로 하천 구역에서의 연도별 지표 피복 변화의 코호트 분석을 실시한 바 있다. 국내에서는 Son *et al.* (2022)이 식생지수 추출에 활용한 바 있다.

이러한 배경에서 본 연구에서는 국내 하천에 적합한 지표 피복 분류 기법을 개발하기 위한 노력의 일환으로 고해상도 위성영상 자료를 기반으로 머신러닝 분류 기법의 하나인 랜덤 포레스트 기법을 내성천에 적용하고 그 적용성을 평가하고자 한다.

2. 랜덤 포레스트 기법

랜덤 포레스트 알고리즘(이하 RF)은 Breiman (2001)이 제안한 기계 학습(machine learning) 기법의 하나로써 특정 사건 발생에 대한 분류, 회귀 분석을 수행함에 있어 랜덤 샘플링한 서로 다른 자료를 학습한 다수의 의사결정나무를 구성하는 앙상블 학습 방법의 일종이다.

RF의 알고리즘은 다음 네 단계로 요약될 수 있다.

- 1) 전체 자료로부터 n 개의 임의의 부트스트랩 표본을 추출한다. 전체 자료는 변수(feature)의 수(본 연구에서는 위성영상의 개별 밴드가 모두 변수가 된다)와 훈련 표본수의 곱이다.
- 2) 부트스트랩 샘플을 이용하여 개별 의사결정 나무를 학습하는데, 각 노드에서는 중복을 허용하지 않고 랜덤하게 d 개의 변수를 선택한다. 정보 획득량과 같은 목적함수를 기준으로 최선의 분할 결과를 만드는 변수를 사용해서 노드를 분할한다.
- 3) 단계 1), 2)를 k 번 반복한다
- 4) 여러 개의 의사결정나무 회귀결과 중에 투표(voting)를 통해 가장 인기 있는 등급(class)을 정하는데, 보통은 평균에 의한다.

가장 정보가 많은 변수를 기준으로 노드를 나누기 위해 최적화할 목적함수를 정의한다. RF에서 사용될 수 있는 목적함수는 각 분할 과정에서 정보 획득량을 최대화한다. 정보 획득량(I)은 Eq. (1)과 같이 정의할 수 있다.

$$I = H(S) - \sum_{i \in \{L, R\}} \frac{|S^i|}{|S|} H(S^i) \quad (1)$$

여기서, S 는 한 노드에 도달하는 데이터의 집합을, S^i 는 이 노드의 $i \in \{L, R\}$, 왼쪽 또는 오른쪽 방향 자식 노드로 들어가는 데이터 집합을 가리킨다. 또한 $|S^i|$, $|S|$ 는 각각 데이터 집합

에 속한 데이터의 개수이며, $H(S)$, $H(S^i)$ 는 새넨 엔트로피 (shannon entropy)를 가리킨다.

포레스트의 모든 나무는 독립적으로 훈련 단계를 거친다. 시험(테스트) 단계에서 데이터 포인트는 모든 트리에 동시에 입력되어 종단 노드에 도달하게 된다. 이러한 테스트 단계는 병렬적으로 진행될 수 있으며, 따라서 병렬 CPU 또는 GPU 하드웨어를 통해 높은 계산 효율성을 얻을 수 있다.

식생 분류와 관련하여 Park *et al.* (2019)은 낙엽송과 편백의 적지분석을 실시하였으며, 머신러닝을 통한 의사결정이 기존에 활용되던 회귀기법보다 더욱 적합한 지점을 선정할 수 있다고 주장하였다. Jung *et al.* (2021)의 연구에서는 RF를 남조류 발생 예측모델 개발에 활용하여 발생 주요 영향인자를 도출하고, 변수 중요도를 순위화하여 남조류 발생에 대한 보의 영향을 규명하였다. 한편 Tang *et al.* (2022)은 Landset과 Sentinel-1 위성의 영상을 기반으로 RF를 활용하여 중국 전역의 지표수 매핑을 실시하였으며, AI에 의해 도출된 결과가 실제 고해상도 이미지와 큰 차이가 없다고 보고하였다. 또한 Auda *et al.* (2023)은 Sentinel-2 위성 영상과 RF를 활용하여 스웨덴 북부 지역의 토지 피복 변화를 평가하였는데, 연구 결과 실제 토지 피복에 대비해 약 88%의 정확도를 보이는 것으로 나타났다.

3. 연구 방법

3.1 연구지역

본 연구에서는 내성천을 대상으로 하천의 지표 피복 분류를 1년 단위로 제작하였다(Fig. 1). 내성천은 낙동강의 제1지류 하천으로, 경북 봉화군 물야면 오전리에서 발원해 예천군 용궁면 향석리에서 낙동강과 합류한다. 법정 유로 연장은 108.2 km이다. 내성천 유역은 소백산맥 남쪽의 경북 북부 내륙에 위치하며, 유역면적 1814.7 km²이다. 내성천은 총적사질하천의 특성을 잘 나타내며, 하상과 사주는 모래가 주를 이룬다. 하상토의 대표 입경은 D₅₀ 기준으로 0.92~1.63 mm이며, 이는 유역의 대보화강암의 풍화에 의해 주로 공급되었다(Lee, 2010). 하도경사는 0.09~0.17%로 완만하다.

내성천 주변의 골짜기의 폭은 하폭의 1.5~3배 정도에 불과하여 산사면과 직접 접해있는 곳이 많으며, 만곡부에선 그 폭이 더욱 좁다(Lee *et al.*, 2022). 반면 직선구간은 홍수터가 잘 발달하여 있어 상대적으로 넓은 하곡을 흐른다. 홍수터의 제 내지는 주로 경작지로 활용되고 있다. 반면 제외지는 주기적인 교란을 받아 준설되거나, 주민들에 의해 경작지로 이용되거나, 방치되어 있다. 2013년 이전에는 하폭이 넓고 수심이 얇은 망상(braided)하도 형태였으나, 2015년 이후 급격한 식생 활착이 이루어져 현재는 수역과 일부 사주를 제외한 상당 부분의 하천면적에 식생이 활착해 있다(Lee *et al.*, 2019b).

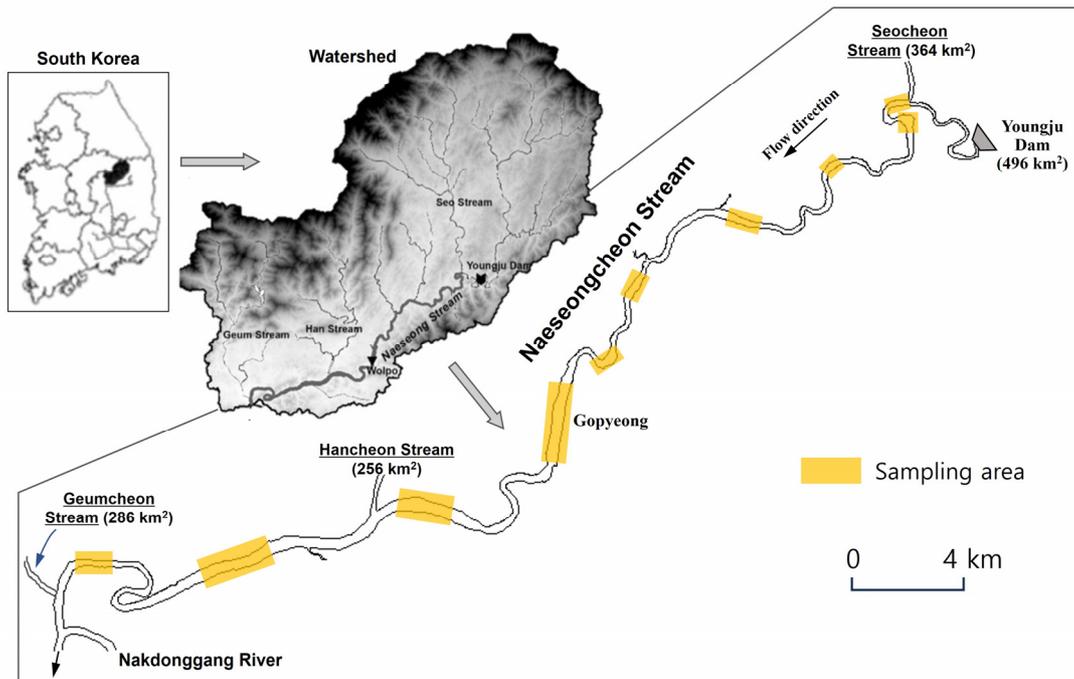


Fig. 1. Study area (Modified from Lee *et al.*, 2023)

3.2 Sentinel-2 영상 및 처리

본 연구에서는 RF를 적용하기 위해 Sentinel-2의 레벨-1C 위성영상 자료를 사용하였다. Sentinel-2는 지표 피복 및 식생 연구를 위해 설계된 위성으로 지구상의 동일한 지점을 5일마다 반복 촬영하며 swath를 고려하면 한반도는 3일 이내의 재방문 주기를 가진다고 볼 수 있다(ESA, 2023). 2015년 이후 자료를 이용 가능하며 구글어스엔진(Google Earth Engine)을 통해 무료로 제공한다. Sentinel-2는 400~2300 nm 파장대에 대해 3가지 공간해상도(10 m급 4개 밴드, 20 m급 6개 밴드, 60 m급 3개 밴드)의 13개 다중분광 이미지 위성으로 토지 피복 분류에 적합하다(Nguyen *et al.*, 2020). 주요 밴드의 최고 해상도가 10 m로서 하천의 규모가 작고 수변 식생의 범위가 좁아 저해상도 영상 활용이 곤란한 국내 하천에 적합하다(Kim and Jang, 2015).

Sentinel 영상을 이용한 처리 과정은 다음과 같다. 첫째, 각 하천별로 하천 전체가 충분히 포함되는 영역을 설정하였다. 이 영역은 각 하천에 적용할 영상을 탐색하는데 사용한다. 레벨-1C 영상은 UTM/WGS84 기준으로 110*110 km² 타일로 제공되지만(ESA, 2023) 하천별로 두 개 이상의 타일에 걸쳐 있는 경우가 있기 때문이다. 둘째, 각 영역에 속하는 Sentinel 영상에 대해 구름 보정을 수행하기 위한 필터링을 수행한다. 필터링은 QA 밴드에 존재하는 값과 bitmask 알고리즘을 이용한다. QA 밴드에서는 비트 10이 불투명(opaque) 구름을 나타내며, 11은 권운(cirrus)을 나타낸다. 본 연구에서는 구름 비율이 10%인 경우를 적용하였다. 셋째, RF 학습 및 분류에 적용할 영상을 제작한다. 여기에서는 1년의 달력 기간을 계절적 특성을 고려하여 2개월씩 묶어 총 6개 시기로 구분하고 각 시기에 속하는 여러 영상을 기반으로 중앙값 필터를 사용하여 시기별로 하나의 영상만을 추출한다. 넷째, 6개의 시기를 대표하는 6개의 영상을 연속적으로 합성하여 최대 78(13*6)개 밴드를 갖는 단일 합성영상을 제작한다. 다만 이 영상은 6개의

계절 중에 영상이 없는 계절이 있는 경우 밴드의 수가 감소할 수 있다. 본 연구에서는 내성천 지역의 2016년 합성영상을 학습 영상으로 활용했는데, 이 영상은 9-10월 영상이 없어서 총 65개 밴드만을 사용하였으며, 그에 따라 분류에 사용할 다른 연도의 영상도 65개 밴드 정보만이 사용되었다.

3.3 학습 및 검증 자료 구축

RF를 포함한 대부분의 머신러닝 알고리즘은 참값을 이용하여 학습 및 검증을 수행하도록 되어 있다. 본 연구에서는 학습을 위한 지상 참값(ground truth)으로 내성천을 대상으로 2016년 12월에 촬영한 항공영상을 기반으로 제작한 지표피복도를 사용하였다. 이는 제방과 산사면을 하천의 외곽 경계로 정하고 하천 현장에 익숙한 전문가가 수작업으로 작도한 것이다. 지표 피복은 Table 1과 같이 총 9개 유형(수역(OWN, 1), 개방 수역(OWN, 2), 사주(BAN, 3), 교란지(BAC, 4), 산발 식생(BAV, 5), 초본 식생(VGH, 6), 목본 식생(VGT, 7), 경작지(HUC, 8), 목초지(HUP, 9))로 구분되어 있으나 공사 등에 의해 일시적으로 조성한 교란된 수역(OWE)은 개방 수역(OWN)과 본질적으로 분광 특성이 같을 것으로 판단되어 실제로는 8개 유형만을 사용하였다.

내성천 구간 전체의 하천 공간 면적은 약 14백만 m²로 셀의 수는 137,982개이며, 각 분류 군별로 1,000개의 표본이 무작위로 추출되었다. 표본은 미리 선정한 10개의 샘플링 영역에서 선택하였는데(Fig. 1), 이는 내성천 전 구간의 길이가 52 km로 길고 지표 피복의 분포 상태가 균일하지 않으므로 이를 고려하여 적절히 배분하였기 때문이다. 추출된 표본의 수는 모두 7,670개로, 지표 피복의 상대적 비율이 높은 수역, 사주, 초본 식생 등은 1,000개가 추출되었으나 다른 지표 피복 유형은 1,000개에 미치지 못하는 경우도 있었다. 이들을 절반으로 나누어 각각 학습과 분류된 결과의 정확성을 검증하는데 사용하였다.

Table 1. Type of land surface cover (Modified from Lee and Kim, 2018)

| Type | | Description |
|------|-----------------------------|---|
| 1 | Open water (OWN) | Open water surface mainly in a low flow channel and minorly floodplain (ponds and wetland) without vegetation |
| 2 | Bare bar (BAN) | Emerged part of riverbed and floodplain covered with sand without vegetation |
| 3 | Disturbed land (BAC) | Land of construction work, unpaved roads, temporarily cleared area etc. |
| 4 | Incipient vegetation (BAV) | Emerged riverbed and floodplain covered with incipient herbaceous vegetation |
| 5 | Herbal vegetation (VGH) | Emerged riverbed and floodplain covered with herbaceous vegetation |
| 6 | Woody vegetation (VGT) | Emerged riverbed and floodplain covered by shrubs and trees and shrubs |
| 7 | Cropland (HUC) | Dry field cropland area |
| 8 | Plantation or pasture (HUP) | Hay field, formerly herbaceous vegetation area |

3.4 데이터 분석

본 연구의 데이터 분석은 Python 환경에서 실시되었으며 크게 클라우드 서버 분석 단계, 이후의 로컬 서버 분석 단계로 구분된다. 고해상도 Sentinel-2 영상 처리와 RF의 반복 적용은 대용량 데이터에 대한 대규모 연산이 수반되므로, 연산 효율을 제고하기 위해 클라우드 서버에서 실행되는 것이 적절하다. 이를 위하여 구글 클라우드(google cloud) 상에서 구동되는 구글어스엔진의 고성능 GIS 서비스를 Python API를 통해 접근·활용하였다.

한편 클라우드 상에서 RF 적용 시, 트리 수, 피복 분류군 별 샘플 수(이하 샘플 수), 대상 구역 파라미터를 조정하여 2016년도 모델의 민감도 분석을 함께 수행하였다. 이는 다른 파라미터가 기본값에서 불변일 때 한 파라미터의 값을 변화하였을 때 정확도의 추이를 보는 것이다. 파라미터 기본값은 트리 수의 경우 50개, 샘플 수의 경우 1,000개, 대상 구역의 경우 Fig. 1의 10개 연구 구역 전체로 주어지며, 파라미터 변화는 트리 수의 경우 10개~100개(10 단위), 샘플 수의 경우 100개~1,000개(100 단위), 대상 구역의 경우 Fig. 1의 상류 5개 구역, 하류 5개 구역, 전체 10개 구역으로 설정하였다.

클라우드 서버상의 분석 결과 연도별 내성천 유역 토지 피복 유형 래스터 데이터가 산출된다. 래스터 데이터의 시각화를 위해 지도를 제작하였다. 이 래스터 데이터를 다시 로컬 서버 Python에서 재구동하여 연도별 토지 피복 예측치에 대한 기본적인 통계, 2016년 모델의 오류 분석을 실시하였다. Rasterio와 Geopandas 패키지를 이용하여 래스터 데이터를 각 셀을 행으로, 셀별 경위도 좌표와 토지 피복 예측치를 열로 가지는 테이블 데이터로 재구조화한 후 통상적인 기술 통계로 연도별 토지 피복 면적 변화의 추이를 도출할 수 있다.

같은 방법으로 지상 참값과 샘플 포인트를 테이블 데이터로 재구조화하고 2016년의 테이블 데이터와 연계하면 샘플 포인트를 행으로, 경위도 좌표, 지상 참값 및 토지 피복 예측치를 열로 가지는 테이블을 얻는다. 마지막으로 지상 참값을 입력값(input), 토지 피복 예측치를 출력값(output)으로 한 에러 매트릭스를 도출하여 오류 유형을 개관하고, 에러 매트릭스에 나타난 주요 오류 유형에 대한 지도화 및 분석을 실시하였다.

4. 결과 및 고찰

4.1 민감도 분석

트리 수, 샘플 수, 하천 구역에 따른 검증 정확도의 변화를 분석한 결과는 Fig. 2와 같다. 먼저 트리 수의 경우, 30개의 트

리 수를 가지는 환경에서 가장 높은 검증 정확도(85.2%)를 보였다(Fig. 2(a)). 30개 이하의 트리 수에서는 그 수가 증가할수록 검증 정확도도 함께 증가하는 경향이 나타났다. 트리 수 30개에서부터 100개까지 큰 변화는 발생하지 않았다. 따라서 트리 수의 경우 30개에서 수렴하므로, 가장 효율적인 지점으로 판단된다.

샘플 수의 경우 모든 경우에서 검증 정확도 80% 이상의 값을 보였으며, 800개 이하에서는 그 수가 증가할수록 검증 정확도도 함께 증가하는 경향이 나타났다(Fig. 2(b)). 가장 높은 검증 정확도가 나타난 샘플 수는 800개(86.4%)였으며, 그 뒤로

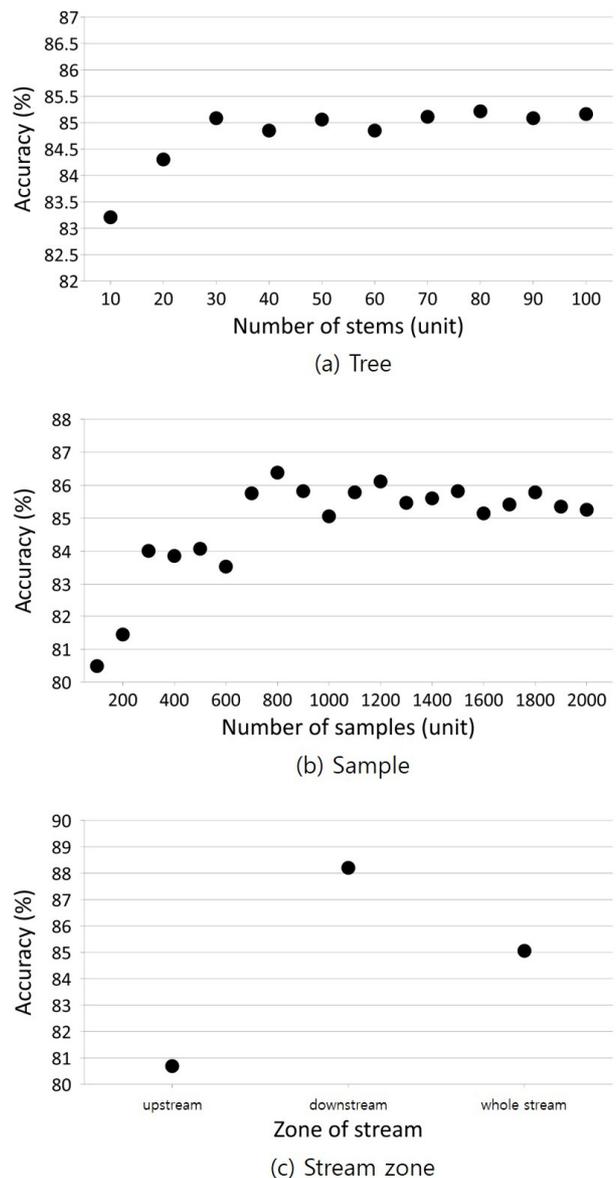


Fig. 2. Validation accuracy

1,200개(86.1%), 900개(85.8%) 순이었다. 반면 가장 낮은 검증 정확도의 샘플 수는 100개(80.5%)로 분석되었다. 따라서 샘플수의 경우 800개 이상에서 검증 정확도의 큰 차이가 발생하지 않으므로, 800개가 가장 효율적인 지점이라고 판단된다.

하천 구역 구분에 따른 검증 정확도는 하류에서 가장 높은 정확도를 보였으며(88.2%), 상류에서 가장 낮았다(80.7%). 전체 구역에서는 그 사이의 값인 85.1%로 나타났다. 이와 같은 경향이 나타난 이유는 분류 유형의 가짓수가 하류에서 더 많기 때문이다. 내성천의 경우, 상류는 하폭이 좁고 하천 주변의 토지가 적어 초지 및 경작지가 나타나지 않는다. 그러나 하류에서는 상류에서 나타나지 않는 분류군이 분포한다. 따라서 상류는 분류 유형이 6가지로 제한되고, 하류에 비해 적은 분류군을 갖게 되므로, 상대적으로 낮은 정확도를 보인다.

4.2 연도별 지표 피복 변화 분석(2016~2023)

내성천 전구간의 연도별 지표 피복 변화는 부록에 제시하였다. 연도별 지표 피복 변화를 부분적으로 살펴보면, Fig. 3과 같다. 예시의 위치는 내성천의 고평교 상류 부근이다. 2016년(Fig. 3(a))의 지표 피복은 우안 상류에 교란지가 위치하며, 좌안을 따라 목본 식생이 줄지어 분포한다. 그러나 2017년(b)에서는 교란지가 사라지고, 목본이 감소한 것을 확인할 수 있다. 2018년(c)에는 그 면적이 더욱 감소하며, 경작지와 목초지가 간헐적으로 확인된다. 반면 2019년(d)에는 좌안의 목본이 증가하고, 우안에 목초지가 크게 형성된다. 또한 하도 중류 부근에서 하류 방향으로 하안침식이 발생하여 유로의 이동이 관찰된다. 2020년(e)에는 수역의 범위가 줄어들었으며, 전년도에 증가하였던 목본이 다시 감소하여 매우 적게 분포한다. 한편, 2021년(f)에는 상류 부근에서 하천분기(river bifurcation)가 발생하여 우안으로 새로운 물줄기가 형성되었다. 2022년(g)에는 교란지가 다수 분포하는 것이 확인되었으며, 경작지의 분포가 증가하였다. 2023년(h)에는 2016년 교란지가 위치하던 지역에 목초지가 형성되었으며, 중류에 사주가 크게 발달한 것을 확인할 수 있다.

이러한 결과는 실제 하천 환경의 변화를 잘 반영한다. 2016년에 나타나는 우안의 교란지는 현재는 사라졌다. 또한 다수의 연도에서 확인할 수 있는 목본 식생의 감소는 실제 하천에서도 확인된다. 내성천은 홍수 예방 및 하천 관리 측면에서 목본 식생이 지속적으로 제거되어 왔다. 따라서 2016년 지도에서 확인되는 띠 형태의 목본이 해가 지날수록 사라지는 현상은 인간 개입에 의한 결과로 판단된다. 또한 2023년 지도에서 사주의 면적이 증가한 것을 관찰할 수 있다. 이는 2023년 여름철 집중호우에 의한 결과인 것으로 추정된다. 실제 2023년 여

름철 강우는 기록적인 강수량(1018.5 mm)을 보였는데, 이는 여름철 전국 평균 강수량(727.3 mm)보다 많은 수준이다(KIPA, 2023). 따라서 하천 내 유사의 공급이 많아져 사주가 증가한 것으로 추측된다.

연도별 지표 피복 비율은 Fig. 4와 같다. 분류군 중 초본 식생이 모든 연도에서 가장 높은 비율을 차지했다. 특히 2020년에는 53.1%, 2022년에는 50.3%로, 양 연도 모두 절반이 넘는 비율을 기록했다. 다음으로 높은 비율을 차지한 항목은 모든 연도에서 10% 이상의 값이 나타난 수역이었다. 2020년에는 12.9%로 가장 낮은 비율을 기록했는데, 해당 연도는 Fig. 3(e)에서 확인 가능한 것처럼 하천의 범위가 크게 감소한 시기이다. 이어서 목본 식생, 사주, 산밭 식생 등이 비율 순위에서 뒤를 이었다. 반면 교란지, 경작지, 목초지는 상대적으로 낮은 비율을 보였는데, 각각 2017년에는 9.2%, 2018년에는 11.8%, 2019년에는 5.3%가 가장 높은 비율을 기록했다.

초본 식생이 가장 높은 비율을 차지하는 이유는 넓은 분포와 다른 분류군과의 혼재 때문이다. 초본은 다양한 하천 환경에서 성장하여 전체적으로 분포가 가능하며, 수역, 사주, 산밭 식생, 목본과 함께 위치할 수 있어서 분포 비율이 높게 나타난다. 또한, 이는 내성천의 전체적인 환경 변화와도 관련이 깊다. Lee *et al.* (2023)에 따르면 내성천은 2013~2014년부터 녹색 하천의 형태를 띠게 되어, 초본의 비율이 꾸준히 높게 나타나는 것으로 판단된다. 한편 목본의 비율은 2018년과 2021년, 2023년에 크게 감소하였는데, 이는 상위에 기술하였듯 인간 개입에 의한 결과이다. 하천 공간의 수목은 홍수 시 수위를 상승시키고 범람을 야기할 수 있으므로, 이를 감소시키기 위해 지속적으로 제거되어 왔다. 사주의 비율은 2023년에 11.0%로 가장 높은 비율을 기록했다. 이 또한 앞서 기술하였듯 해당 연도에 발생한 기록적 호우 때문이다. 이에 따라 하천 내 유사의 공급이 증가하여 초본의 비율이 감소하고 사주의 비율이 증가하였다.

4.3 정확도 분석

4.3.1 에러 매트릭스

본 연구에서 적용한 RF를 이용한 지표 피복 분류의 오류를 확인하기 위해 지상 참값을 활용해 작성한 에러 매트릭스는 Table 2와 같다. 가장 높은 정확도를 보인 항목은 목초지(99.0%)이며, 그 뒤로 교란지(97.1%), 경작지(93.8%), 목본 식생(87.4%) 순으로 높은 정확도를 보였다. 산밭 및 초본 식생을 제외한 6가지 항목에서는 전체 정확도(85.1%)와 유사하거나 그보다 높은 값이 나타난 반면에, 산밭 및 초본 식생은 각각 72.7%, 64.2%로 전체 정확도보다 크게 낮은 값을 보였다.

목초지의 정확도가 가장 높게 나타난 이유는 목초지가 내성천 하류에 넓고 집약적으로 분포하여 일정한 분광특성이 유지되기 때문이다. 교란지 및 경작지의 정확도가 높은 것 역시 이와 같은 이유이다. 반면 산밭 및 초본 식생의 정확도가 낮은 이유는 산밭 및 초본 식생이 목초지와 반대되는 특성을 보이기 때문이다. 해당 항목들은 다른 요소들과 혼재되어 분

포하며, 산밭적으로 위치한다. 특히, 초본의 경우 타 항목과의 관계에서도 많은 오류가 발생하는데, 이는 초본이 오류가 발생하는 분류군들과 유사한 분광특성을 지녔기 때문이다. 또한 초본과 목본은 108개로 가장 많은 오류가 나타나는데, 이는 둘의 분광특성이 유사하며 서로 근접하여 위치하기 때문이다.

지표 피복 분류의 오류를 구분하여 살펴보면 경계오류

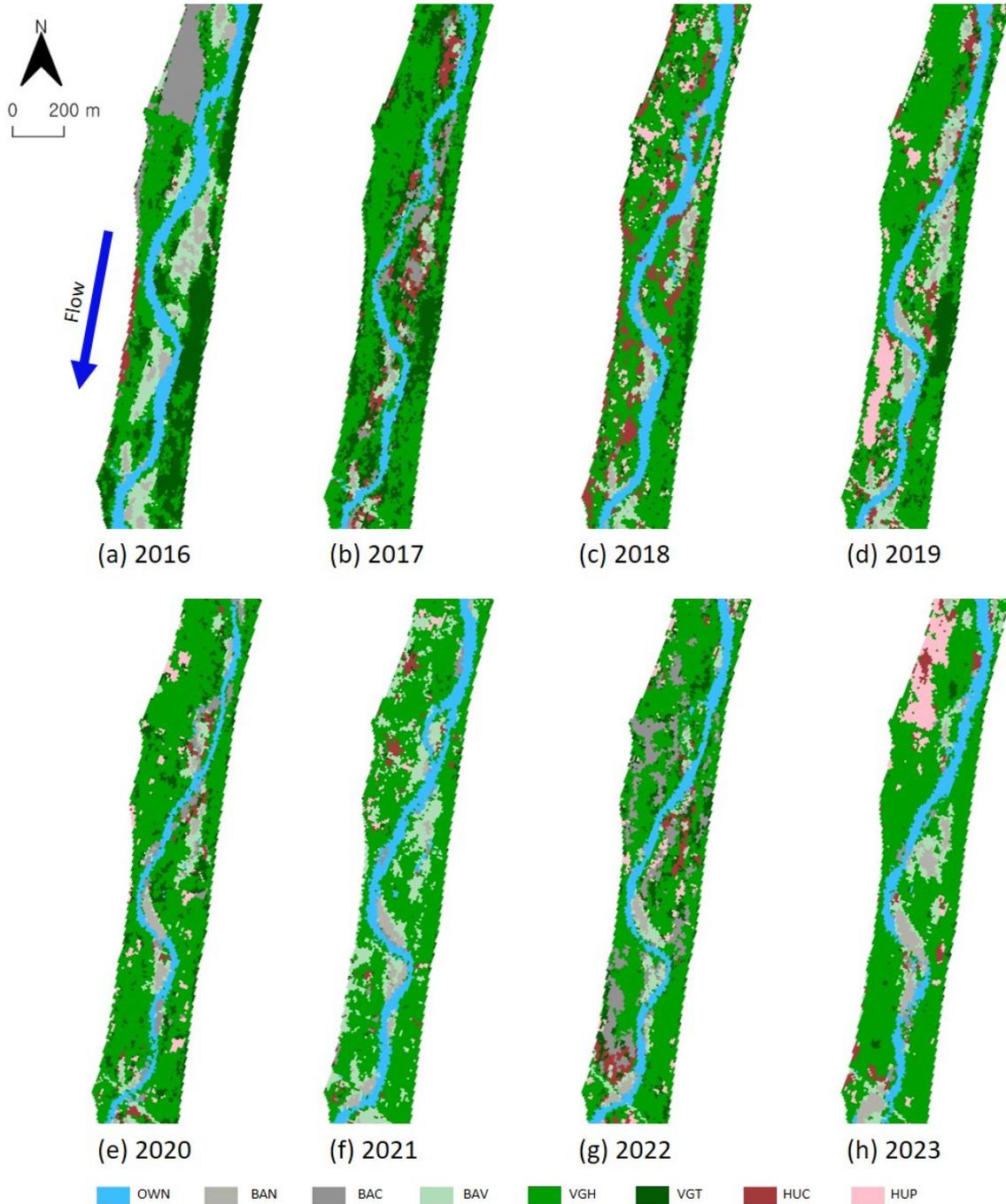


Fig. 3. Changes in surface cover for 2016-2023

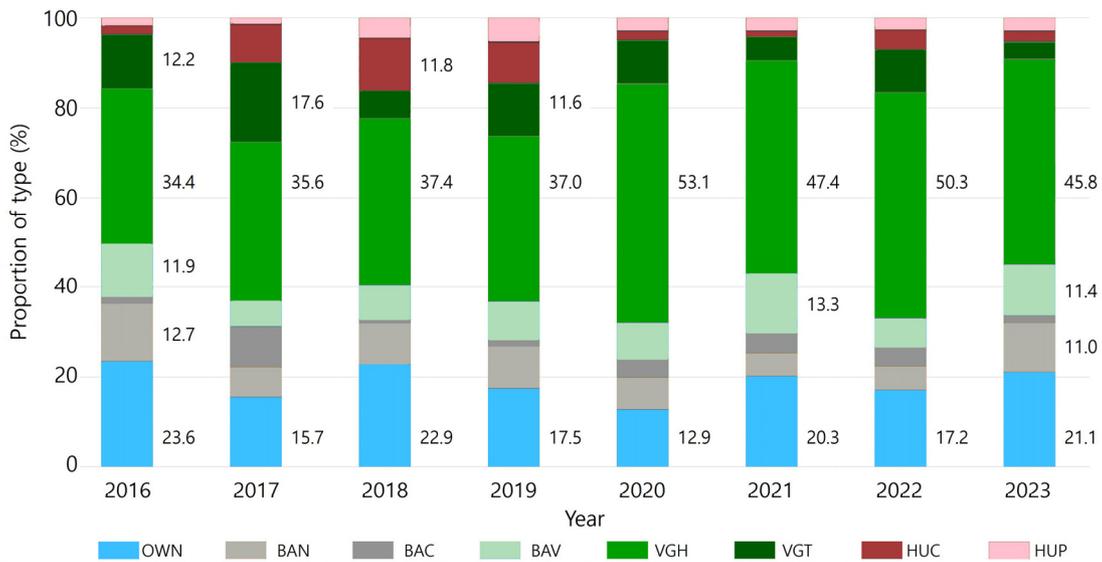


Fig. 4. Changes in surface cover by year

Table 2. Error matrix

| Actual classes | | Predicted classes | | | | | | | |
|----------------|-----|-------------------|------|------|------|------|------|------|------|
| | | OWN | BAN | BAC | BAV | VGH | VGT | HUC | HUP |
| Total (unit) | | | | | | | | | |
| 563 | OWN | 478 | 8 | 9 | 4 | 24 | 2 | 0 | 0 |
| 515 | BAN | 17 | 435 | 0 | 54 | 7 | 0 | 0 | 0 |
| 344 | BAC | 7 | 1 | 334 | 6 | 1 | 0 | 1 | 0 |
| 545 | BAV | 1 | 52 | 0 | 396 | 28 | 0 | 4 | 0 |
| 436 | VGH | 53 | 16 | 1 | 77 | 280 | 46 | 11 | 5 |
| 451 | VGT | 7 | 3 | 0 | 8 | 62 | 394 | 15 | 0 |
| 500 | HUC | 0 | 0 | 0 | 0 | 18 | 7 | 469 | 0 |
| 481 | HUP | 0 | 0 | 0 | 0 | 16 | 2 | 0 | 476 |
| Accuracy (%) | | 84.9 | 84.5 | 97.1 | 72.7 | 64.2 | 87.4 | 93.8 | 99.0 |

(boundary error)와 내부오류(internal error)로 구분된다. 경계오류는 인접한 셀에 영향을 받아 발생하는 오류이다. 경계 오류는 Sentinel-2 위성영상의 셀 크기(10 m²)에 기인하며, 셀 안에 두 개 이상의 분류가 공존하는 경우에 발생한다. 예를 들어, 초본 식생과 목본 식생은 서로 인접하여 분포하는 경우가 많은데 이에 따라 같은 셀 안에 위치할 가능성이 높으며, 이에 따라 경계오류가 발생할 확률 또한 상승한다. 이와 달리 내부 오류는 셀 내부에서 발생한 오류이다. 내부오류는 유사한 분광특성 및 낮은 밀도 등에 의해 발생한다. 초본과 목본 식생의 경우를 살펴보면, 이들은 유사한 분광특성을 지니고 있기 때문에 내부오류가 발생하기 쉽다. 또한 목본의 경우 나무의 수관이 밀집하여 분포하는 지역은 목본이 밀도 높게 나타나지만, 한 그루만 있는 경우나 산발적으로 분포하는 경우에도 인

공지능이 셀을 목본으로 판독할 수 있다.

가장 많은 오류를 보인 항목은 초본과 목본 식생으로, 전체 오류 중 18.8%의 비율을 보였다(Table 3). 이 둘의 오류 중 54.6%는 경계오류, 45.4%는 내부오류 가장 높은 내부오류 비율이 나타났다. 나지 사주와 산발 식생, 산발과 초본 식생, 수역과 초본 식생 순으로 많은 오류가 발생하였으며, 전체 오류 중 18.5%, 18.3%, 13.4%를 차지하였다. 이 중 경계오류의 비율이 가장 높은 항목은 수역과 초본 식생이지만(97.4%), 경계오류의 개수가 가장 많은 항목은 산발 식생과 초본 식생이다(49개). 이들 네 분류군의 오류 비율은 전체 중 69%를 차지했다. 따라서 위성영상의 셀 크기에서 기인하는 경계오류를 오류에 포함하지 않고 전이지역을 고려하면 순수한 오차는 2.8%이므로 매우 정확한 성능을 얻을 수 있다고 판단된다.

Table 3. Number and proportion of main, boundary, and internal error

| Actual classes | Predicted classes | Misclassification count (Overall error, %) | Boundary error (%) | Internal error (%) |
|----------------|-------------------|---|-----------------------|-----------------------|
| VGT | VGH | 108 (18.8) | 54.6 | 45.4 |
| VGH | VGT | | | |
| BAN | BAV | 106 (18.5) | 79.2 | 20.8 |
| BAV | BAN | | | |
| VGH | BAV | 105 (18.3) | 93.3 | 6.7 |
| BAV | VGH | | | |
| VGH | OWN | 77 (13.4) | 97.4 | 2.6 |
| OWN | VGH | | | |
| Other classes | | 177 (30.9) | 85.3 | 14.7 |
| Total | | 573 (100) | 81.5 | 18.5 |

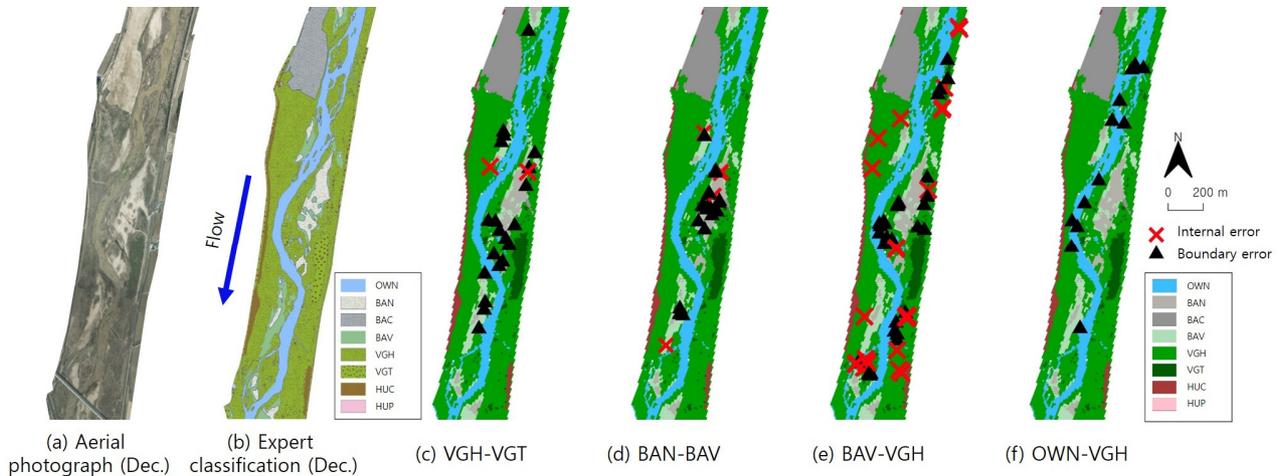


Fig. 5. Aerial photograph, Expert classification and main error in Gopyeong, Naeseong Stream (2016)

4.3.2 오류의 공간 분포 분석

지표 피복 분류 상태의 오류는 분류된 지표피복도에서도 가시적으로 확인할 수 있다. 경계오류는 다른 분류를 지니고 있는 셀이 인접한 경우에 잘 관찰되며, 내부오류는 위치에 관련 없이 나타난다. 오류 발생 위치의 예시는 Fig. 5과 같으며, Fig. 5(a)는 2016년 12월 고평 부근의 항공사진이고, (b)는 같은 시기의 지표 피복을 전문가가 분류한 지도이다. 가장 많은 오류가 발생한 관계는 초본 식생과 목본 식생이다(Fig. 5(c)). 그 뒤로 사주-산밭 식생(d), 산밭 식생-초본 식생(e), 수역-초본 식생(f) 순으로 나타난다. Figs. 5(c)-5(e)에 나타난 삼각형은 경계오류이다. 앞서 기술하였듯 경계오류는 Sentinel 위성의 셀 크기에 기인한다. X는 내부오류로, 산밭-초본 식생에서 잘 확인된다.

Fig. 5(d)에서 경계오류가 주로 발생하는 이유는 두 분류군이 인접하여 분포하기 때문이다. 반면 (e)에서는 내부오류의 수가 많은데, 이는 산밭 식생과 초본 식생의 분광특성이 비슷하기 때문으로 판단된다. 한편 (f)에서는 수역과 초본 식생은 분광특성 상 혼동될 수 없는 항목임에도 불구하고 경계오차가 발생한다. 이러한 이유는 수역과 초본 식생이 근접하여 있기 때문으로 생각된다.

5. 요약 및 결론

본 연구에서는 국내 하천에 적합한 지표 피복 분류 기법을 개발하기 위해 고해상도 위성 영상 자료를 활용하여 RF를 내

성천에 적용하고 그 적용성을 평가하였다. 이를 위해 4단계 과정으로 영상을 처리하였으며, 2016년 합성 영상을 학습에 사용하였다. 학습 및 검증 자료 구축에는 2016년 12월에 촬영한 항공 영상을 기반으로 제작된 전문가 분류의 지표피복도를 활용하였다. 8개의 유형으로 분류된 지표 피복은 총 14백만 m²의 하천 공간을 대상으로 137,982개의 셀로 나누어 무작위로 추출되었다. 이러한 자료를 활용하여 민감도 분석, 연도별 피복 변화, 정확도 분석을 실시하였다.

연구 결과는 다음과 같다. 먼저, 민감도 분석 결과, 트리 수는 30개, 샘플 수는 800개, 하천 구역은 하류에서 상대적으로 높은 효율성을 보이는 것으로 나타났다. 트리 수의 정확도는 30개 이하에서 증가하다가 그 이상에서 유사한 값을 보였으며, 샘플 수 또한 800개 이상에서 수렴하였다. 샘플 영역은 상류보다는 하류를 포함하였을 때 높은 정확도를 보였으며, 이는 분류군의 다양성이 더 높기 때문으로 판단된다. 두 번째, 연도별 피복 변화를 살펴본 결과, 실제 하천 환경을 잘 반영하는 것으로 확인되었다. 하도 면적의 변화 및 목본의 유무가 잘 판독되었으며, 실제 수문 조건 변화에 의한 하천 환경 변화도 반영하는 것으로 판단된다. 이러한 변화는 셀의 비율 변화에서도 확인할 수 있다. 마지막으로 정확도 분석 결과, 산발 식생과 초본 식생을 제외한 6가지 항목에서 상대적으로 높은 정확도를 보였다. 산발 및 초본 식생은 다수의 오류가 발생하였는데, 특히 초본 식생은 목본 식생과 내부오류를, 산발 식생과 경계오류가 많았다. 이는 초본과 목본이 유사한 분광특성을 지니고 있으며, 초본과 산발 식생이 인접하여 분포하기 때문으로 판단된다.

하천 식생을 포함하는 하천 내 공간의 지표 피복 상태에 대한 정량화와 정기적인 모니터링은 하천 관리를 위한 기본적인 업무이다. 원격 탐사 자료를 기반으로 정량화된 지표 피복 정보는 하천의 흐름 저항과 통수능의 공간적인 상태를 평가하는 기본 자료로 활용될 수 있으며, 이를 기반으로 홍수위를 예측하거나 생태계를 유지하면서 적절한 홍수 안전을 도모하는 방안을 수립하기 위한 의사결정 과정에 활용될 수 있다. 따라서 지속적인 데이터베이스 구축을 위해서는 보다 많은 하천에 원격 탐사 및 지표 피복 분류 기법의 적용이 필요하다.

감사의 글

이 논문은 한국건설기술연구원 주요사업(디지털뉴딜 기반 통합물관리 기술 융합 플랫폼(IWRM-K) 개발(202401-66))의 연구비 지원으로 수행되었습니다.

Conflicts of Interest

The authors declare no conflict of interest.

References

- An, H.Y., Na, S.I., Park, C.W., So, K.H., and Lee, K.D. (2018). "Atmospheric correction effectiveness analysis of reflectance and NDVI using multispectral satellite image." *Korean Journal of Remote Sensing*, Vol. 34, No. 6-1, pp. 981-996.
- Auda, Y., Lundin, E.J., Gustafsson, J., Pokrovsky, O.S., Cazaurang, S., and Orgogozo, L. (2023). "A new land cover map of two watersheds under long-term environmental monitoring in the Swedish arctic using Sentinel-2 Data." *Water*, Vol. 15, No. 18, 3311.
- Breiman, L. (2001). "Random forests." *Machine Learning*, Vol. 45, pp. 5-32. doi: 10.1023/A:1010933404324.
- Choi, S.U., Yoon, B.M., and Woo, H.S. (2005). "Effects of dam-induced flow regime change on downstream river morphology and vegetation cover in the Hwang River, Korea." *River Research and Applications*, Vol. 21, No. 2-3, pp. 313-325.
- European Space Agency (ESA) (2023). Sentinel-2 MSI Revisit and Coverage, accessed 12 October 2023, <<https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage>>.
- Geerling, G., Penning, E., Rogers, C., Nic, C.V.D.V.S., Harezlak, V., Wilson, S., and Donchyts, G. (2021). *Operational monitoring of floodplain vegetation using remote sensing*. Authorea, Hoboken, NJ, U.S., pp. 1-16.
- Gurnell, A. (2014). "Plants as river system engineers." *Earth Surface Processes and Landforms*, Vol. 39, No. 1, pp. 4-25.
- Harezlak, V., Geerling, G.W., Rogers, C.K., Penning, W.E., Augustijn, D.C., and Hulscher, S.J.M.H. (2020). "Revealing 35 years of landcover dynamics in floodplains of trained lowland rivers using satellite data." *River Research and Applications*, Vol. 36, No. 7, pp. 1213-1221.
- Johnson, W.C. (1994). "Woodland expansions in the Platte River, Nebraska: Patterns and causes." *Ecological Monographs*, Vol. 64, No. 1, pp. 45-84.
- Jung, W.S., Kim, S.E., and Kim, Y.D. (2021). "Analysis of influential factors of cyanobacteria in the mainstream of Nakdong River using random forest." *Journal of Wetlands Research*, Vol. 23, No. 1, pp. 27-34.
- Kim, J.Y., and Jang, D.H. (2015). "An applications of remote sensing data for the land surface characteristics analysis of ecological stream restoration area in Daejeon Stream." *Journal of the Association of Korean Geographers*, Vol. 4, No. 2, pp. 231-340.
- Korea Institute of Construction Technology (KICT) (2013). *Development of floodplain maintenance technology for enhancement of waterfront value*, No. 2013-128, p. 197.
- Lee, C.H., and Cho, K.H. (2023). "Development of a method for

- tracking sandbar formation by weir-gate opening using multi-spectral satellite imagery in the Geumgang River, South Korea." *Ecology and Resilient Infrastructure*, Vol. 10, No. 4, pp. 135-142.
- Lee, C.H., Kim, H.R., and Cho, K.H. (2022). "Spatial distribution and successional changes of riparian vegetation on sandbars exposed after watergate-opening of weirs in the Geumgang River, South Korea." *Ecology and Resilient Infrastructure*, Vol. 9, No. 3, pp. 194-205.
- Lee, C.J., and Kim, D.G. (2018). "Analysis of the changes of the vegetated area in an unregulated river and their underlying causes: A case study on the Naeseong Stream." *Ecology and Resilient Infrastructure*, Vol. 5, No. 4, pp. 229-245.
- Lee, C.J., Choi, H., Kim, D.G., Van Oorschot, M., Penning, E., Geerling, G.W. (2023). "Bio-geomorphic alteration through shifting flow regime in a modified monsoonal river system in Korea." *River Research and Applications*. Vol. 39, pp. 1639-1651.
- Lee, C.J., Go, D.H., Kim, D.G., and Choi, H. (2022). "Quantitative analysis on the lateral migration of an alluvial river - Case study on Gopyeong site of Naeseong Stream." *Journal of the Korean Geomorphological Association*, Vol. 29, No. 2, pp. 25-39.
- Lee, C.J., Kim, D.G., Hwang, S.Y., Kim, Y.J., Jeong, S.J., Kim, S.N., and Cho, H.J. (2019a). "Dataset of long-term monitoring on the change in hydrology, channel morphology, landscape and vegetation along the Naeseong Stream (II)." *Ecology and Resilient Infrastructure*, Vol. 6, No. 1, pp. 34-48.
- Lee, C.J., Kim, D.G., Ji, U., and Kim, J.S. (2019b). "Dataset of long-term monitoring on the change in hydrology, channel morphology, landscape and vegetation along the Naeseong Stream (I)." *Ecology and Resilient Infrastructure*, Vol. 6, No. 1, pp. 23-33.
- Lee, G.R. (2010). "A comparative analysis on channel forms and landscapes at Naeseongcheon River and Wicheon River in Gyeongpook Province." *Journal of the Korean Geomorphological Association*, Vol. 17, No. 1, pp. 1-16.
- Nguyen, T.T.H., Doan, M.T., Tomppo, E., and McRoberts, R.E. (2020). "Land use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods - A case study from Dak Nong, Vietnam." *Remote Sensing*. Vol. 12, No. 9, 1367.
- Park, E.J., Park, J.H., and Kim, H.H. (2019). "Mapping species-specific optimal plantation sites using random forest in Gyeongsangnam-do Province, South Korea." *Journal of Agriculture & Life Science*, Vol. 53, No. 6, pp. 65-74.
- Rood, S.B., Kaluthota, S., Philipsen, L.J., Slaney, J., Jones, E., Chasmer, L., and Hopkinson, C. (2019). "Camo-maps: An efficient method to assess and project riparian vegetation colonization after a major river flood." *Ecological Engineering*, Vol. 141, 105610.
- Son, S.Y., Kim, J.H., and Kim, Y.J. (2022). "Google Earth Engine-based satellite image processing and ground vegetation index extraction technique." *Journal of Korea Water Resources Association*, Vol. 55, No. 9, pp. 46-53.
- Tang, H., Lu, S., Ali Baig, M. H., Li, M., Fang, C., and Wang, Y. (2022). "Large-scale surface water mapping based on landsat and sentinel-1 images." *Water*, Vol. 14, No. 9, 1454.
- The Korea Institute of Public Administration (KIPA) (2023). *ISSUE PAPER-2023 analysis of disaster management issues in concentrated rainfall: Focusing on the on-site situation propagation system*, No. 2023-132.
- Williams, G. P. (1978). *The case of the shrinking channels the North Platte and Platte Rivers in Nebraska*. U.S. Geological Survey Circular, 781, U.S. Geological Survey, Reston, VA, U.S. pp.1-48.
- Woo, H.S., Park, M.H., Cho, K.H., Cho, H.J., and Jeong, S.J. (2010). "Recruitment and succession of riparian vegetation in alluvial river regulated by upstream dams - Focused on the Nakdong River downstream Andong and Imha Dams -." *Journal of Korea Water Resources Association*, Vol. 43, No. 5, pp. 455-469.
- Zanetti, C., Macia, J., Liency, N., Vennetier, M., Mériaux, P., and Provencal, M. (2016). "Roles of the riparian vegetation: the antagonism between flooding risk and the protection of environment." *3rd European Conference on Flood Risk Management (FLOODrisk 2016)*, Lyon, France, p. 13015.

부록. 내성천 전체 연도별 피복 면적(2016~2023)

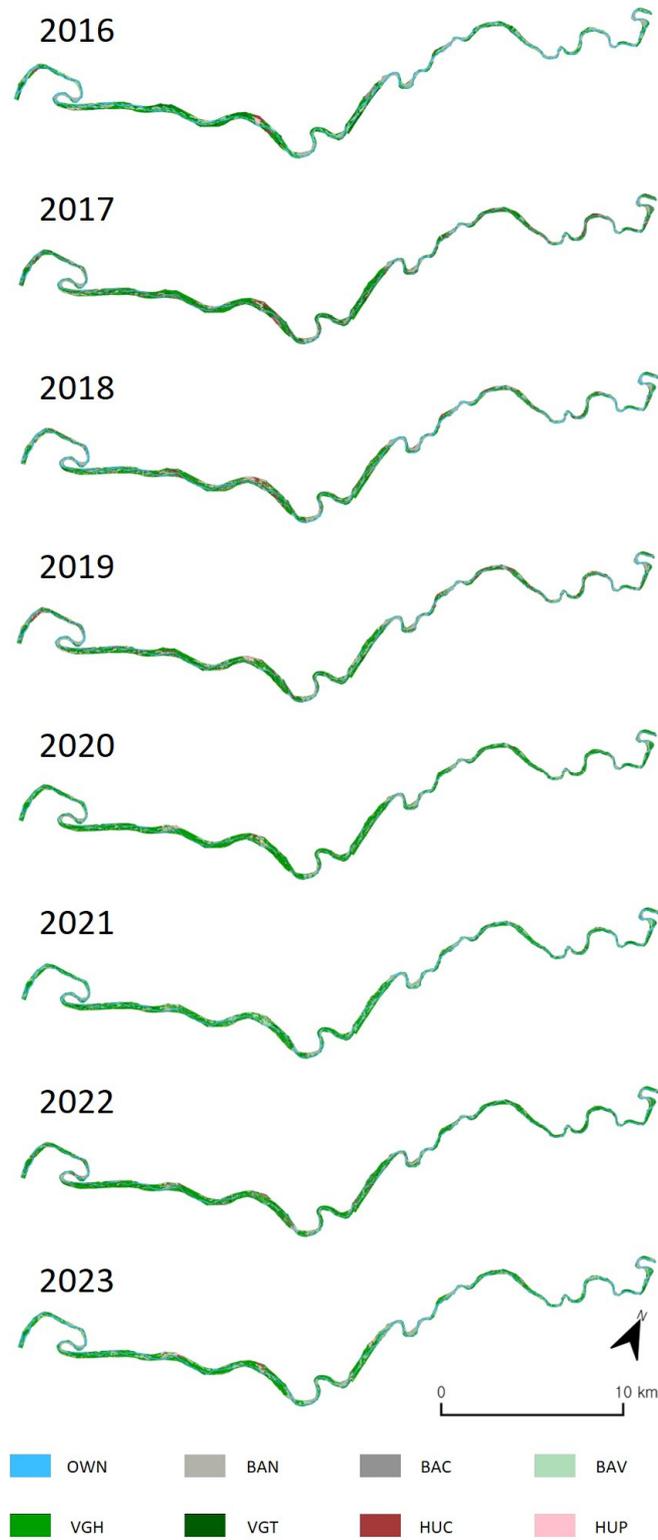


Fig. 6. Surface cover by year at Naeseong Stream