

Research Article

무인기 기반 초분광영상을 이용한 배나무 엽록소 함량 추정

강예성¹⁾ · 박기수²⁾ · 김은리²⁾ · 정종찬²⁾ · 유찬석^{3)*} · 조정건⁴⁾

Estimation of Chlorophyll Contents in Pear Tree Using Unmanned Aerial Vehicle-Based-Hyperspectral Imagery

Ye Seong Kang¹⁾ · Ki Su Park²⁾ · Eun Li Kim²⁾ · Jong Chan Jeong²⁾ ·
Chan Seok Ryu^{3)*} · Jung Gun Cho⁴⁾

Abstract: Studies have tried to apply remote sensing technology, a non-destructive survey method, instead of the existing destructive survey, which requires relatively large labor input and a long time to estimate chlorophyll content, which is an important indicator for evaluating the growth of fruit trees. This study was conducted to non-destructively evaluate the chlorophyll content of pear tree leaves using unmanned aerial vehicle-based hyperspectral imagery for two years (2021, 2022). The reflectance of the single bands of the pear tree canopy extracted through image processing was band rationed to minimize unstable radiation effects depending on time changes. The estimation (calibration and validation) models were developed using machine learning algorithms of elastic-net, k-nearest neighbors (KNN), and support vector machine with band ratios as input variables. By comparing the performance of estimation models based on full band ratios, key band ratios that are advantageous for reducing computational costs and improving reproducibility were selected. As a result, for all machine learning models, when calibration of coefficient of determination (R^2) ≥ 0.67 , root mean squared error (RMSE) $\leq 1.22 \mu\text{g}/\text{cm}^2$, relative error (RE) $\leq 17.9\%$ and validation of R^2 ≥ 0.56 , RMSE $\leq 1.41 \mu\text{g}/\text{cm}^2$, RE $\leq 20.7\%$ using full band ratios were compared, four key band ratios were selected. There was relatively no significant difference in validation performance between machine learning models. Therefore, the KNN model with the highest calibration performance was used as the standard, and its key band ratios were 710/714, 718/722, 754/758, and

Received August 8, 2023; Revised September 5, 2023; Accepted September 18, 2023; Published online October 31, 2023

¹⁾ 경상국립대학교 농업생명과학연구원 바이오시스템공학과 연구교수(Research Professor, Department of Biosystem Engineering, Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, Republic of Korea)

²⁾ 경상국립대학교 농업생명과학연구원 바이오시스템공학과 석사과정생(Master Student, Department of Biosystem Engineering, Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, Republic of Korea)

³⁾ 경상국립대학교 농업생명과학연구원 생물산업기계공학과 교수(Professor, Department of Bio-Industrial Machinery Engineering, Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, Republic of Korea)

⁴⁾ 국립원예특작과학원 과수과 농업연구관(Senior Researcher, Fruit Research Division, National Institute of Horticultural & Herbal Science, Wanju, Republic of Korea)

* Corresponding author: Chan Seok Ryu (ryucs@gnu.ac.kr)

Copyright © 2023 by The Korean Society of Remote Sensing. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

758/762 nm. The performance of calibration showed $R^2=0.80$, $RMSE=0.94 \mu\text{g}/\text{cm}^2$, $RE=13.9\%$, and validation showed $R^2=0.57$, $RMSE=1.40 \mu\text{g}/\text{cm}^2$, $RE=20.5\%$. Although the performance results based on validation were not sufficient to estimate the chlorophyll content of pear tree leaves, it is meaningful that key band ratios were selected as a standard for future research. To improve estimation performance, it is necessary to continuously secure additional datasets and improve the estimation model by reproducing it in actual orchards. In future research, it is necessary to continuously secure additional datasets to improve estimation performance, verify the reliability of the selected key band ratios, and upgrade the estimation model to be reproducible in actual orchards.

Keywords: Band ratio, Computational cost, K-nearest neighbors, Machine learning, Reproducibility

요약: 과일 나무의 생육을 평가하는 중요한 지표인 엽록소 함량을 추정하는데 비교적 많은 노동력의 투입이 요구되고 오랜 시간이 소요되는 기존의 파괴 조사 대신 비파괴적 조사 방식인 원격탐사기술을 적용하기 위한 연구가 시도되고 있다. 이 연구에서는 2년(2021, 2022) 간 무인기 기반의 초분광 영상을 이용하여 배나무 잎의 엽록소 함량을 비파괴적으로 추정하는 연구를 수행하였다. 영상 처리로 추출된 배나무 캐노피(canopy)의 단일 band 반사율은 시간 변화에 따라 불안정한 복사 효과를 최소화하기 위해 밴드비화(band rationing) 되었다. 밴드 비(band ratios)를 입력 변수로 머신러닝 알고리즘인 elastic-net, k-nearest neighbors (KNN)과 support vector machine을 사용하여 추정(calibration, validation) 모델들을 개발하였다. Full band ratios 기반 추정 모델들의 성과와 비교하여 계산 비용 절감과 재현성 향상에 유리한 key band ratios를 선정하였다. 결과적으로 모든 머신러닝 모델에서 full band ratios를 이용한 calibration에 coefficient of determination (R^2) ≥ 0.67 , root mean squared error (RMSE) $\leq 1.22 \mu\text{g}/\text{cm}^2$, relative error (RE) $\leq 17.9\%$ 와 validation에 $R^2\geq 0.56$, $RMSE\leq 1.41 \mu\text{g}/\text{cm}^2$, $RE\leq 20.7\%$ 성능을 비교하였을 때, key band ratios 네 개가 선정되었다. 머신러닝 모델들 사이에 validation 성능에는 비교적 큰 차이가 없어 calibration 성능이 가장 높았던 KNN 모델을 기준으로 삼았으며, 그 key band ratios는 710/714, 718/722, 754/758, 758/762 nm가 선정되었다. Calibration에서 $R^2=0.80$, $RMSE=0.94 \mu\text{g}/\text{cm}^2$, $RE=13.9\%$ 와 validation에서 $R^2=0.57$, $RMSE=1.40 \mu\text{g}/\text{cm}^2$, $RE=20.5\%$ 를 나타내었다. Validation의 기준으로 한 성능 결과는 배나무 잎 엽록소 함량을 추정하기에 충분하지 않았지만, 앞으로의 연구에 기준이 될 key band ratios를 선정했다는 것에 의미가 있다. 추후 연구에서는 추정 성능을 향상하기 위해 지속적으로 추가 데이터셋을 확보하여 선정된 key band ratios의 신뢰성 검증과 함께 실제 과원에 재현 가능한 추정 모델로 고도화할 필요가 있다.

주요어: 밴드비, 계산 비용, K-최근접 이웃, 머신러닝, 재현성

1. 서론

잎의 엽록소 함량은 식생의 광합성 활동, 질소 함량 및 건강 상태를 나타낸다. 더 나아가 식생의 질병 식별, 간접적인 영양상태와 직접적인 노화 정도를 나타내는 지표이다(Gitelson and Merzlyak, 1997). 과원의 경우 수체의 영양상태가 과실의 수확량 및 품질을 결정하는데, 영양상태를 파악하기 위해서는 인력이 투입되어 잎을 파괴함으로써 복잡한 과정을 거쳐 오랜 시간 뒤에 상태를 확인할 수 있다(Paul et al., 2020). 이를 극복하기 위한 정밀농업 기술로 잎의 엽록소 함량을 비파괴적으로 추정하는 것은 앞으로 해결해야 나가야 할 중요한 문제로 대두되고 있다(Ta et al., 2021).

최근 다양한 영상 센서가 무인기에 탑재되어 보다 넓은 면적의 농지를 모니터링 하는 것에 활용되고 있으며, 기존의 인공위성 및 항공기에 비해 상대적으로 운용이 간단하고 비용이 저렴하며 낮은 고도에서 영상을 수집하기 때문에 더욱 정밀한 생육 추정 모델을 개발하는데 유리하다(Hunt Jr and Daughtry, 2018). 특히 영상 센서 중 가시광선에서 비가시광선 영역까지 넓은 범위의 중심 파장(band)들을 포함하는 초분광 영상은 다양한 식생의 생육상태를 보다 정밀히 비파괴적으로 모니터링 하는 수단으로 활용되고 있다. 그러나 초분광 데이터의 방대한 데이터양은 느린 계산 시간은 물론 과적합과 같은 문제로 재현성을 방해하거나 성능을 하락시키는 단점도 있다(Pu et al., 2015). 이러한 한계를 극복하기 위해

불필요한 변수들을 제거하고 특정 추정 목표에 민감한 변수들만을 선정하는 과정을 수행해오고 있다(Kim et al., 2011; Acosta et al., 2023). 이를 위해 머신러닝 분석법은 식생 상태를 모니터링 하기 위한 고정밀 모델을 개발하고 주요 변수를 선정하는 것에 이용되어 왔다(Feng et al., 2022). Decision tree, support vector machine (SVM) 분류분석법을 초분광 영상에 적용하여 주요 변수 선정과 함께 배나무(*Pyrus pyrifolia* var. *culta*) 화상병 감염 유무를 판별하는 연구(Bagheri, 2020; Kang et al., 2021)와 주요 변수 선정을 고려하지 않고 Gaussian process와 SVM 회귀분석법으로 배나무 잎의 엽록소 함량을 추정하는 연구가 수행되어 왔다(Paul et al., 2020).

본 연구의 목적은 무인기 기반 초분광 영상을 이용해 시계열에 따라 변하는 배 잎의 엽록소 함량 추정모델 개발과 주요 입력 변수를 선정하는 것이다. 입력 변수로는 단일 band가 아닌 band rationing된 반사율을 기반으로 최소한의 모델 개발 가능성을 보였던 elastic-net (EN), k-nearest neighbors (KNN)와 SVM 머신러닝 회귀분석법들을 이용하였다. Full band ratios를 이용한 모델의 성능과 band ratios를 10개에서 2개까지 줄여나가면서 개발한 모델들의 성능을 비교하는 것을 통해 엽록소 함량 추정에 유리한 key band ratios를 선정하였다.

2. 연구자료 및 방법

2.1. 연구 대상지

본 실험은 전라북도 완주군의 농촌진흥청 국립원예특작과학원(35°82'8377" N, 127°03'1448" W) 과수 실험 필드에서 2021년 7월 중순에서 10월 중순까지와 2022년 5월 하순에서 9월 하순까지 수행되었다(Fig. 1). 배나무는 2년에서 3년생인 원황으로 상토, 마사토와 펄라이트(perlite)가 5:4:1로 구성된 토양이 담겨있는 포트에 각각 심겨졌다. 그림과 같이 총 76 그루가 지중 매립되었으며, 질소시비처리에 따른 엽록소의 변이를 관찰하기 위해 구역을 나누어 한 해 총량 171 g, 43 g과 0 g으로 각각 과다시비, 적정시비와 무처리 되었다. 질소시비는 총량을 나누어 5월 24일부터 8월 14일까지 매주 처리되었다.

2.2. 영상 취득

초분광 영상센서(MicroHSI 410 Shark, Corning Inc., USA)는 쿼드콥터 무인기(Matrice 300 RTK, DJI Technology Inc., China)에 탑재되어 2021년에 7회(7월 13일, 7월 28일, 8월 11일, 8월 31일, 9월 15일, 9월 30일, 10월 15일), 2022년에 9회(5월 23일, 6월 3일, 6월 17일, 7월 4일, 7월 19일, 7월 28일, 8월 16일, 9월 7일, 9월 21일)로 약 2주 간격으로 시

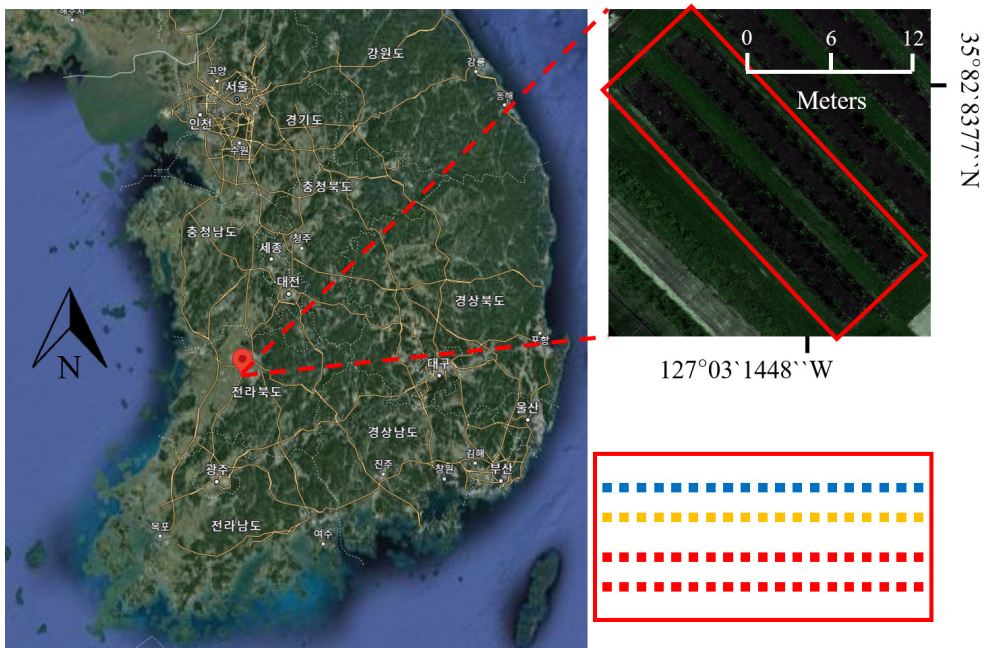


Fig. 1. Pear orchard with excessive (blue), moderate (yellow), and untreated (red) nitrogen fertilization.

계열 영상을 취득하였다. 초분광 영상센서의 field of view는 29.5°로 파장 영역 400-1000 nm에서 총 150개의 중심파장을 포함한 영상이 취득된다. 쿼드콥터 무인기는 크기 96×103×43 cm, 최대 하중 무게 2.7 kg과 최대비행시간 45분의 사양을 가지고 있다. 비행속도 6 m/s, 중첩률 70%와 고도 60 m 설정으로 DJI pilot을 이용하여 무인기를 자동 비행하였고, 이때 취득된 초분광영상 공간해상도는 4.4 cm였다.

2.3. 영상 처리

취득된 초분광영상은 영상처리소프트웨어(ENVI 5.6, Exelis Visual Information Solutions, USA)를 이용하여 자이로, 방사 및 기하보정 되었다. 기술 발전과 함께 무인기 비행시간이 점점 늘어남에 따라 모니터링 할 수 있는 배 과원의 수와 면적이 늘어나고 있다. 이에 따라 분광영상의 방사 보정은 긴 비행 중의 시간 변화로 영상 취득 위치별로 달라지는 광 조건을 정규화하기 위해 reference target을 여러 위치에 설치하는 방법을 고려할 필요가 있다(Vanegas et al., 2018). 이를 위해서 본 연구에서는 추후 여러 target을 설치해야 한다는 점을 고려하여 다양한 흰색과 검정색 조합의 panel 중 12% reference target (Portable Fabric Target, Group 8 Technology Inc., USA)만을 실험 필드에 설치한 후 영상 처리로 추출한 target의 반사율을 이용하여 초분광영상을 방사 보정하였다(Abdulridha et al., 2019). 식생을 강조하기 위해 초분광 영상을 normalized difference vegetation index 영상으로 변환하였다. 식(1)의 near infrared (NIR)에 850 nm와

red에 677 nm의 반사율(ρ)을 이용하였다(Yu et al., 2022).

Normalized difference vegetation index =

$$\frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (1)$$

Fig. 2에서와 같이 전체 배나무 캐노피 영역에서 폴리곤(polygon)을 이용하여 추출한 배나무 개체별 각 band 반사율이 산출되었다.

시계열 영상의 단일 band 반사율은 시간 변화에 따라 달라지는 태양광 각도와 강도의 차이에 의한 불안정한 복사 효과를 최소화하기 위해 band rationing 방법으로 서로 인접해 있는 단일 band의 반사율들을 나누었다 (Mather and Koch, 2011; Kang et al., 2020).

2.4. 샘플 조사

초분광영상 취득 직후 처리구 당 3그루(총 9그루) 배나무에서 각각 성엽 10개를 채취한 후 지름 1.1 cm의 코르크로 잎의 일부분을 추출해 아세톤액이 담긴 삼각 플라스크에 4°C의 암 조건에서 12시간 이상 침지한 후 분광광도계를 이용하여 총 엽록소 함량을 측정하였다.

2.5. 통계 처리

측정한 엽록소 함량은 프로그래밍 코드(Python 3.9, Python Software Foundation, USA)로 two sample을 t-test 하여 시계열 및 처리구별 변이를 비교하였다. 농업데이터분석 플랫폼(FinePro, Hortizen Co. Ltd., Korea)을 통해 배나무 캐노피의 반사율을 이용하여 EN, KNN, SVM 회귀분석으로 배나무 잎의 엽록소 함량을 추정하기 위

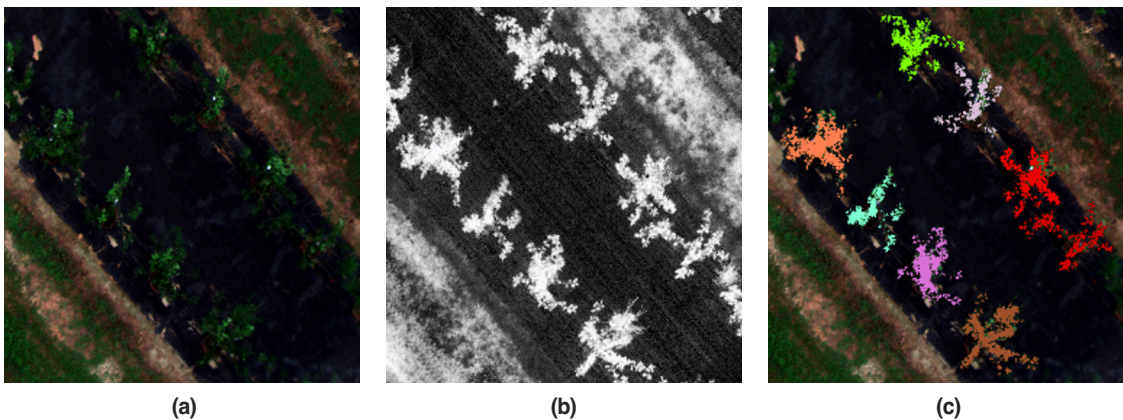


Fig. 2. Hyperspectral image processing procedure: (a) raw RGB image, (b) conversion to normalized difference vegetation index, and (c) extraction of individual canopies.

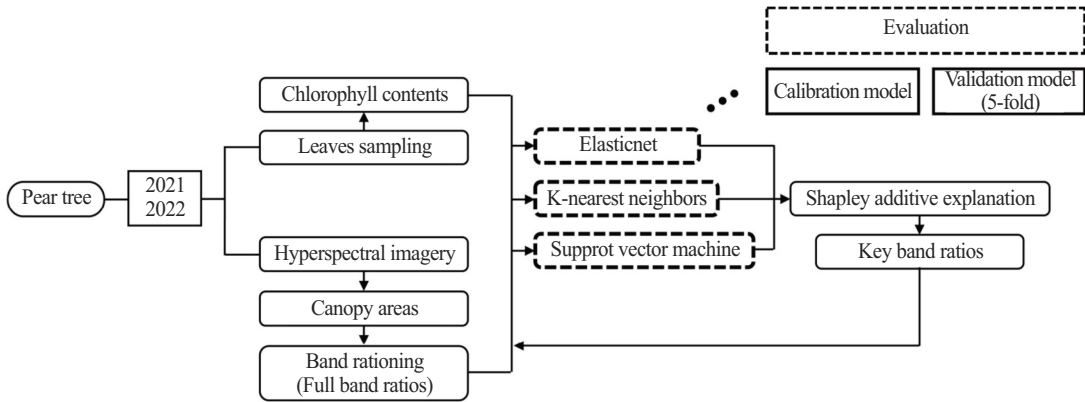


Fig. 3. Flow chart for estimating the chlorophyll contents in pear tree leaves using hyperspectral imagery.

한 calibration 모델을 개발하였다. Grid search법을 이용하여 각 회귀분석법 모델에 대한 최적의 parameters (EN의 alpha, KNN의 n_neighbors, SVM의 C와 gamma)를 식별하였다. EN의 alpha는 회귀 모델을 정규화하는 것으로 회귀 계수의 크기에 제약을 두어 설명력이 낮은 변수에 대한 가중치를 줄이는 것이다(De Mol et al., 2009). 유클리드, 맨해튼과 본 연구에 채택된 민코프스키(Minkowski) 거리를 계산하여 가장 가까운 이웃의 수 범주에 포함된 예측치를 도출하기 위해 KNN의 이웃의 수 n_neighbors를 먼저 결정한다(Ge et al., 2021). N_neighbors가 낮을 수록 과적합되기 쉬우며, 반대로 과소적합된 모델의 경우 복잡성을 주기 위해 n_neighbors를 낮추기도 한다. SVM의 C는 값이 크면 과적합 가능성이 높아지는 하드마진으로 가까워지고 작으면 소프트마진에 가까워지는 파라미터로 오류를 얼마나 허용할 것인지를 결정한다. Gamma는 커널 적용 시 결정 경계의 곡률을 결정하는 파라미터이므로 값이 커지면 과적합 될 가능성이 높아진다(Sharma et al., 2022). 이때 SVM의 kernel은 모델의 복잡도를 줄이기 위해 linear로 고정하였고, 이 연구의 모든 SVM 모델에서 gamma는 1e-07로 선정되었다. 유사성이 높은 실험 필드 데이터라는 점을 고려하여 균형 있는 편향과 분산의 추정치를 도출하기 위해 calibration 모델 개발에 사용된 데이터세트에 k-fold를 5로 설정하여 cross validation한 모델(validation)의 성능을 평가하였다(Aragon et al., 2020). 이 모든 과정은 Fig. 3에 나타내었으며 계산비용절감을 위해 full band ratios를 이용해 개발한 모델과 입력 변수를 10개에서 2개까지 축소해 나가면서 개발한 모델의 R²를 비교하여 key band

ratios를 선정하였다. Key band ratios 선정 지표로는 모든 회귀분석 모델에서 각 입력 변수의 중요도를 나타내는 shapley additive explanation (SHAP)을 이용하였다. SHAP는 shapley 값을 통해 여러 변수의 조합에서 해당 변수의 추가 또는 제거에 따른 변화를 얻을 수 있으며, 그 변화는 양의 방향과 음의 방향의 영향을 모두 계산할 수 있다. 이러한 계산방식 때문에 계산 시간이 오래 걸리는 단점이 있지만, 기존의 중요도 계산 방법보다 변수가 추정치에 미치는 영향을 더 정확하게 측정할 수 있어 중요한 변수를 선정하는 데 목적이 있는 본 연구에 사용되었다(Kannangara et al., 2022).

모든 모델들은 R², RMSE와 RE로 성능이 평가되었다. RMSE와 RE는 다음과 같이 계산된다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$RE = \frac{100}{\bar{y}} \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \quad (3)$$

이 때 y_i 와 \hat{y}_i 는 배나무 잎의 엽록소 함량의 측정값과 추정값이다. \bar{y} 는 측정된 엽록소 함량의 평균값이며, n 은 샘플 수이다. RE는 측정된 엽록소 함량 평균값에 대한 RMSE의 비율을 나타낸다.

3. 연구결과 및 토의

3.1. 배나무 잎 엽록소 함량 기초 분석

Table 1은 엽록소 함량을 시비처리구별 및 시계열별

Table 1. Two-sample t-test of the chlorophyll content in pear tree leaves depending on the growing season and fertilization

	2021								
	Jul. 13	Jul. 28	Aug. 11	Aug. 31	Sep. 15	Sep. 30	Oct. 14		
E ^{a)}	3.84 ^{b)} ±0.63 ^a	4.99±0.81 ^a	3.73±1.67 ^a	3.74±1.68 ^a	4.33±1.59 ^a	5.84±0.85 ^a	5.84±0.81 ^a		
M	4.17±0.67 ^a	3.52±0.55 ^a	4.63±1.28 ^a	4.63±1.28 ^a	6.68±1.55 ^a	6.90±0.93 ^a	7.58±0.90 ^a		
U	4.77±0.86 ^a	3.93±0.25 ^a	5.08±1.21 ^a	5.08±1.21 ^a	6.63±1.79 ^a	7.41±0.59 ^a	7.33±0.66 ^a		
All	4.26±0.75 ^A	4.14±0.83 ^{AB}	4.48±1.35 ^{AB}	4.48±1.35 ^B	5.88±1.84 ^{BC}	6.72±0.98 ^C	6.92±1.07 ^C		
	2022								
	May 23	Jun. 03	Jun. 17	Jul. 04	Jul. 19	Jul. 28	Aug. 16	Sep. 7	Sep. 21
E	7.56±0.58 ^a	6.84±0.27 ^{ab}	6.74±1.05 ^a	8.23±1.27 ^a	8.57±0.47 ^a	9.13±0.91 ^a	9.19±1.29 ^a	9.86±2.47 ^a	8.46±0.63 ^a
M	7.47±0.82 ^{ab}	5.93±0.60 ^a	7.20±0.34 ^a	7.53±0.93 ^a	7.51±1.09 ^a	9.60±1.11 ^a	9.61±0.59 ^a	10.4±1.35 ^a	8.95±0.55 ^a
U	5.93±1.39 ^b	7.54±0.39 ^b	6.22±1.41 ^a	7.18±1.77 ^a	6.25±1.60 ^a	10.6±0.58 ^a	9.11±0.77 ^a	9.70±2.80 ^a	8.01±1.33 ^a
All	6.99±1.20 ^A	6.77±0.80 ^A	6.72±0.90 ^A	7.65±1.27 ^{AB}	7.44±1.42 ^A	9.78±1.02 ^C	9.30±0.84 ^{CD}	9.97±2.01 ^{CD}	8.47±0.89 ^{BD}

^{a)} E: excessive fertilization, M: moderate fertilization, U: untreated fertilization.

^{b)} Two-sample t-test at significance level (p-value < 0.05) with mean ± standard deviation: uppercase letters indicate significant differences between dates, and lowercase letters indicate significant differences between different nitrogen fertilizations.

로 two sample t-test한 결과를 나타낸다. 시비처리구별의 경우에 2022년 5월 23일과 6월 3일을 제외하고 2021년과 2022년의 6월 중순 이후의 날짜에서는 모두 유의한 차이를 나타내지 않았다. 6월 중순부터 꽃눈 분화의 시작과 6월 하순부터 7월 상순까지 신초 생장이 이루어지는 시기에 흡수된 질소가 잎뿐만 아니라 다른 부위로 활발히 이동되어 시비구별로 잎의 엽록소 함량에 큰 차이를 나타내지 않았을 것으로 판단된다(Kim et al., 2006; Kim et al., 2018). 시계열별 분석에서는 2021년과 2022년 모두 7월 말 이전보다 이후의 엽록소 함량이 유의하게 증가하는 경향을 나타내었다. 7월 중순 이후부터 배나무가 본격적으로 생식생장으로 접어들면서 엽록소의 주원료인 마그네슘의 흡수가 많아졌기 때문이라고 여겨진다.

3.2. 배나무 캐노피 반사율 곡선

Fig. 4는 2021년과 2022년의 시계열별 초분광 영상에서 band rationing된 배나무 캐노피 반사율 곡선을 나타낸다. 가시광 영역 중 가장 높은 green (510-520 nm), red edge의 시작점(680-700 nm), 끝점(750-770 nm) 및 900 nm 이후의 수분 흡수 파장대인 NIR 영역(960-970 nm)의 반사율이 급격히 떨어지는 경향을 나타내었다. 반대로 red edge의 끝점(750-760 nm)과 약 900 nm 시작점인 NIR 영역의 반사율은 상승하는 경향을 나타내었다. 이 곡선은 식생의 주요 분광 특성에 따라 하강과 상승이 정상적으로 나타났다는 것을 확인할 수 있었다.

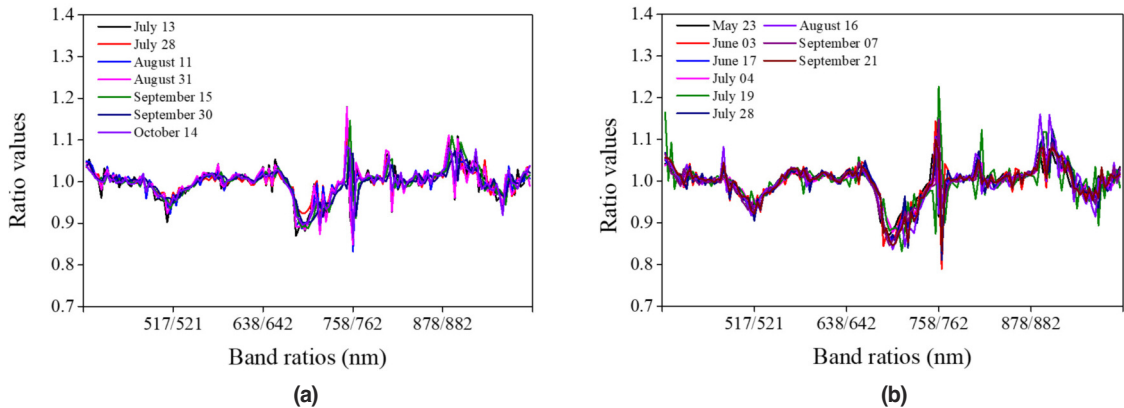


Fig. 4. Reflectance curves of pear tree canopy area obtained by hyperspectral imagery in 2021 (a) and 2022 (b).

Table 2. Estimation performance of models using full band ratios depending on each year and all years

		EN ^{a)}	KNN	SVM	
2021	n ^{b)}	69			
	Mean±standard deviation (µg/cm ²)	5.25±1.57			
	Parameters ^{c)}	0.001	1	10	
	Calibration	R ²	0.69	N.A.	0.54
		RMSE (µg/cm ²)	0.87	N.A.	1.05
		RE (%)	16.6	N.A.	20.1
	Validation	R ²	0.42	0.33	0.39
RMSE (µg/cm ²)		1.19	1.27	1.22	
RE (%)		22.6	24.2	23.2	
2022	n	87			
	Mean±standard deviation (µg/cm ²)	8.04±1.68			
	Parameters	0.001	3	9	
	Calibration	R ²	0.55	0.59	0.43
		RMSE (µg/cm ²)	1.12	1.07	1.26
		RE (%)	13.9	13.3	15.7
	Validation	R ²	0.39	0.24	0.35
RMSE (µg/cm ²)		1.31	1.45	1.35	
RE (%)		16.3	18.1	16.8	
All	n	156			
	Mean±Standard deviation (µg/cm ²)	6.81±1.84			
	Parameters**	0.001	4	10	
	Calibration	R ²	0.70	0.75	0.67
		RMSE (µg/cm ²)	1.17	1.08	1.22
		RE (%)	17.1	15.8	17.9
	Validation	R ²	0.62	0.56	0.62
RMSE (µg/cm ²)		1.32	1.41	1.32	
RE (%)		19.4	20.7	19.4	

a) EN: elastic-net, KNN: k-nearest neighbors, SVM: support vector machine.
 b) n: number of samples.
 c) Alpha in elastic-net, n_neighbors in k-nearest neighbors, C in support vector machine.

3.3. Full band ratios를 이용한 배 잎 엽록소 함량 추정모델

Table 2는 연도별 배나무 캐노피의 full band ratios를 이용하여 개발한 각 머신러닝 모델의 엽록소 추정 성능을 나타내었다. 2021년의 경우 EN의 calibration 모델의 추정 성능이 R²=0.69, RMSE=0.87 µg/cm²와 RE=16.6%로 가장 높았지만, validation 모델에서는 R²=0.42의 선형성을 나타내었다. KNN과 SVM의 경우 calibration 모델은 R²≤0.54, RMSE≥1.05 µg/cm²와 RE≥20.1% 성과 과 그보다 더 낮은 validation 모델의 추정 성능을 나타내었다. 2022년의 경우 모든 머신러닝 모델에서 R²≤0.59,

RMSE≥1.07 µg/cm²와 RE≥22.6%의 calibration 모델과 R²≤0.39, RMSE≥1.31 µg/cm²와 RE≥16.3%의 validation 모델 결과로 정밀한 엽록소 함량 추정에 어려움이 있었다. 대조적으로 두 연도를 합친 엽록소 추정에서는 모든 머신러닝 모델이 R²≥0.67, RMSE≤1.22 µg/cm²와 RE≤17.9%의 calibration 결과와 함께 선형성이 R²≥0.56의 validation 결과를 나타내었다. 그 원인은 2021년과 2022년의 엽록소 함량 평균값에서 확인할 수 있듯이 낮은 2021년의 엽록소 함량과 높은 2022년의 함량이 보완되면서 추정 모델을 개발하기에 유리해진 것으로 여겨진다(Kang et al., 2018). 하지만 calibration의 성능에 비해

validation의 성능이 낮은 경향을 보여 모델의 과적합 여부를 확인하였다. KNN 모델의 경우 유사한 다른 연구 결과를 기준으로 판단하였을 때 파라미터 $n_neighbors=4$, calibration과 validation 사이 R^2 의 0.19, RE의 4.90% 차이로 극적인 과적합 영향이 있다고 판단되지 않았다(Defernez and Kemsley, 1999; Chen et al., 2020; Cedric et al., 2022). EN 모델과 SVM 모델에서도 다른 연구와 비교하여 calibration과 validation 사이에 R^2 와 RE의 더 작은 차이와 선정된 파라미터 값을 고려하였을 때 과적합 되었다고 보기는 어려웠다(Wang et al., 2021). 결론적으로 미미한 과적합 영향과 함께 나무 연생에 상관없이 엽록소 함량 추정 모델 개발이 가능함에 따라 두 연도를 혼합한 데이터셋을 이용하였다.

3.4. Key Band Ratios를 이용한 배 잎 엽록소 함량 추정모델

Table 3는 배나무 잎의 엽록소 함량 추정에 SHAP를 이용하여 선정된 band ratios 10개부터 2개까지를 이용한 머신러닝 모델별 결정계수를 나타내었다. Full band ratios를 이용하였을 때보다 머신러닝 모델별 각각 다른 설명력을 가졌던 입력 변수 band ratios가 동일하게 10개 이하로 줄어들면서 모델들 사이에 추정 성능 차이가 확연히 드러났다. Band ratios 10개 기준으로 EN과 SVM에서 calibration과 validation 모델은 각각 0.62와 0.59 이하로 Table 2의 full band ratios 결과와 비교하여 낮아진 결정계수를 나타내었다. 마찬가지로 Table 2의 결과와 동일하게 KNN의 calibration 결정계수가 0.76으로 가장 높았지만 validation 모델의 결정계수는 0.49로 가장 낮았다. KNN에서 그 경향은 5개 ratio를 이용하였을 때까지 동일하였지만 4개의 band ratios를 이용하였을 때 EN, SVM 뿐만 아니라 Table 2의 full band ratios를 이용한 KNN과 비교해도 가장 높은 calibration에 0.80과 validation에 0.57의 결정계수를 나타내었다. 이는 264 bands로 SVM을 이용해 품종이 다른 배나무 잎의 엽록소 함량을 $R^2 \geq 0.80$ 으로 추정하던 이전 연구(Paul et al., 2020)의 결과와 유사하였다. Validation 모델 성능은 모델 검증을 통해 추후 모델 재현성을 방해하는 과소적합이나 과적합없이 적절하게 잘 설계되었는지를 나타내는 중요한 지표로 설명되기 때문에 key band ratios 선정 기준에 우선순위로 두었다(Zhang et al., 2019). 3개의 band ratios를 이용하였

Table 3. Coefficient of determination (R^2) depending on the number of band ratios selected by each model

No. of band ratios	Model	EN	KNN	SVM
10	Calibration	0.62	0.76	0.59
	Validation	0.59	0.49	0.58
9	Calibration	0.62	0.71	0.59
	Validation	0.59	0.51	0.57
8	Calibration	0.60	0.63	0.59
	Validation	0.58	0.50	0.58
7	Calibration	0.60	0.76	0.58
	Validation	0.57	0.54	0.57
6	Calibration	0.59	0.76	0.57
	Validation	0.58	0.54	0.55
5	Calibration	0.59	0.79	0.57
	Validation	0.57	0.54	0.56
4	Calibration	0.58	0.80	0.56
	Validation	0.57	0.57	0.55
3	Calibration	0.53	0.85	0.51
	Validation	0.52	0.50	0.51
2	Calibration	0.54	0.41	0.46
	Validation	0.53	0.26	0.45

EN: elastic-net, KNN: k-nearest neighbors, SVM: support vector machine.

을 때 0.85로 가장 높은 calibration R^2 를 보여주었지만 validation에서 전체 Table 3의 R^2 들과 비교하여 0.50으로 낮아 key band ratios는 4개로 선정하였다. EN과 SVM에서도 동일하게 calibration과 validation 모두 0.55 이상의 R^2 를 유지하는 4개의 key band ratios로 선정하였다. 그러나 배나무 잎의 엽록소 함량을 추정하기에 충분한 성능이라고 보기 어려워 추후 데이터셋 확보를 통한 성능 향상이 필수적이라고 판단된다. 이 연구에서는 Table 3와 full bad ratios 결과인 Table 2를 비교하였을 때 calibration과는 대조적으로 validation은 $R^2=0.62$ 에서 KNN 모델에 $R^2=0.57$, EN과 SVM 모델에 $R^2=0.55$ 로 변화가 상대적으로 크지 않았다. 이 결과는 선정된 key band ratios가 엽록소 함량 추정에 민감하게 반응한다는 것을 방증한다.

Table 4는 머신러닝 모델별 배나무 잎의 엽록소 함량 추정에 유리한 key band ratios를 나타낸다. EN과 SVM은 Table 3에서 비슷한 추정 성능을 보인 것과 마찬가지로 식생의 생화학적 또는 생물리적 변화에 관련이 깊은 red edge 영역의 706/710, 714/718, 718/722와 NIR 영역의 894/898로 동일한 band ratios가 선정되었다. 반면 KNN에서는 red edge영역에서 Fig. 4와 같이 급격히 하

Table 4. Selected key band ratios for different models

	EN	KNN	SVM
Band ratio 1	706/710	710/714	706/710
Band ratio 2	714/718	718/722	714/718
Band ratio 3	718/722	754/758	718/722
Band ratio 4	894/898	758/762	894/898

EN: elastic-net, KNN: k-nearest neighbors, SVM: support vector machine.

강하는 영역인 710/714와 718/722로 동일하게 선정되었지만 나머지 2개의 band ratios는 급격히 상승하는 red edge 끝점인 754/758과 급격히 하강하는 NIR의 시작점인 758/762가 선정되었다. 유사한 연구에서는 본 연구와 마찬가지로 초분광 영상을 이용해 같은 장미과 낙엽교목인 사과나무 잎의 엽록소 함량 추정을 위해 red edge 영역(680–760 nm)의 band들이 선정되었다(Ta et al., 2021). 반면 상록 소교목인 감귤나무의 경우 red edge를 제외한 NIR, green, red와 blue 영역에서 유의한 band들이 선정되었다(Naqvi et al., 2021). 이처럼 수체의 유전적인 영향이 잎의 엽록소 함량 추정에 유의한 band들을 선정하는 데 영향을 미쳤을 것으로 판단된다. 결과적으로 710, 714, 718, 722, 754, 758, 762 nm로 7개의 band들이 선정되었다.

Table 5는 최종적으로 선정된 4개의 key band ratios를 이용해 calibration에서 가장 높은 성능을 보였던 KNN으로 개발한 배나무 잎 엽록소 함량 추정 모델의 성능을 나타내었다. Calibration 모델에서는 RMSE=0.94 $\mu\text{g}/\text{cm}^2$, RE=13.9%를 나타내었고 validation 모델에서는 RMSE=1.40 $\mu\text{g}/\text{cm}^2$, RE=20.5%로 Table 2의 full band ratios를 이용한 오차와 큰 차이를 보이지 않았다. 추가 데이터셋 확보를 통해 실제 배나무 과원에 재현될 때 full band ratios와 같은 고차원 데이터에서 발생하는 차원의 저주에 의한 과적합으로 인해 오차가 커지기 쉽기 때문에 입력 변수들의 배제를 통해 선정된 key band ratios만으로 구성된 모델이 복잡한 모델보다 재현성 향상에 더 도움이 될 것이다(Defernez and Kemsley, 1999; Kang et al., 2020; Nalepa, 2021). 또한 KNN 회귀분석은 사과나무의 수확량을 예측하기 위해 이용된 연구를 제외하고 배나무의 생육 지표를 추정하는데 KNN 회귀분석이 사용된 경우는 본 연구가 처음이다(Chen et al., 2022).

Table 5. Estimation performance using a KNN-based model with key band ratios

		KNN ^{a)}
No. of samples		156
Mean±Standard deviation		6.81±1.84
Parameter ^{b)}		3
Calibration	R ²	0.80
	RMSE ($\mu\text{g}/\text{cm}^2$)	0.94
	RE (%)	13.9
Validation	R ²	0.57
	RMSE ($\mu\text{g}/\text{cm}^2$)	1.40
	RE (%)	20.5

^{a)} KNN: k-nearest neighbors

^{b)} N_neighbors in k-nearest neighbors

3.5. 배 잎 엽록소 함량 추정 선형 관계 및 Key Band Ratios의 중요도

Fig. 5에 각 EN과 KNN의 full band ratios와 key band ratios를 이용해 개발한 모델에서 추정된 배나무 잎 엽록소 함량과 실제 측정된 엽록소 함량 사이에 선형 관계를 나타내었다. Fig. 5(a)의 full band ratios를 이용한 EN 모델은 Fig. 5(c)의 KNN 모델과 비교하였을 때 약 10 $\mu\text{g}/\text{cm}^2$ 이상의 엽록소 함량을 추정하는 것에 낮은 반응을 보였다. 마찬가지로 key ratio를 이용하였을 때 Fig. 5(b)의 EN에 비해 Fig. 5(d)의 KNN 모델의 추정치가 약 10 $\mu\text{g}/\text{cm}^2$ 이상 엽록소 함량에 더 잘 반응하는 것을 확인할 수 있었다.

Fig. 6에 배나무 잎 엽록소 추정에 이용된 5에서 2개의 band ratios를 이용한 KNN 모델에 SHAP 값을 나타내었다. Fig. 6(a)의 5개의 band ratios를 이용한 모델에서는 Table 4의 key band ratios에 762/766이 추가되었다. Key band ratios에서 배제되었다는 사실에서 알 수 있듯이 나머지 band ratios이 비슷한 SHAP 값을 보인 것과 달리 762/766의 SHAP 값이 가장 낮았고, 결과적으로 Fig. 6(b)의 key band ratios에서 배제되었다. Key band ratios 중에서는 718/722의 SHAP 값이 가장 높아 Fig. 6(c)에서 3개의 band ratios 조합에 의해 718/722가 배제되었을 때 validation 모델 성능 저하에 결정적인 영향을 미쳤다고 판단된다. 또한 Fig. 6(d)에서 710/714의 배제가 calibration과 validation 모델에 전체적인 하락의 원인이 된 것으로 보아 추정 성능에 많은 영향을 미쳤다는 것을 알 수 있었다.

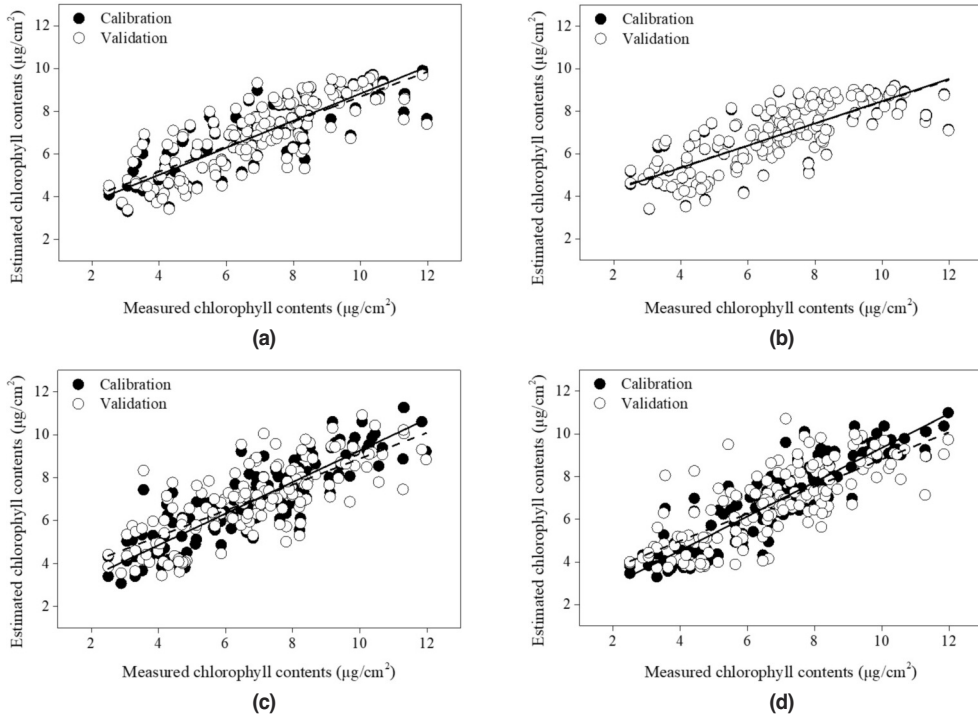


Fig. 5. Linear relationship between measured chlorophyll contents and estimation by EN-based model with (a) full band ratios, (b) key band ratios, (c) KNN-based model with full band ratios, and (d) key band ratios.

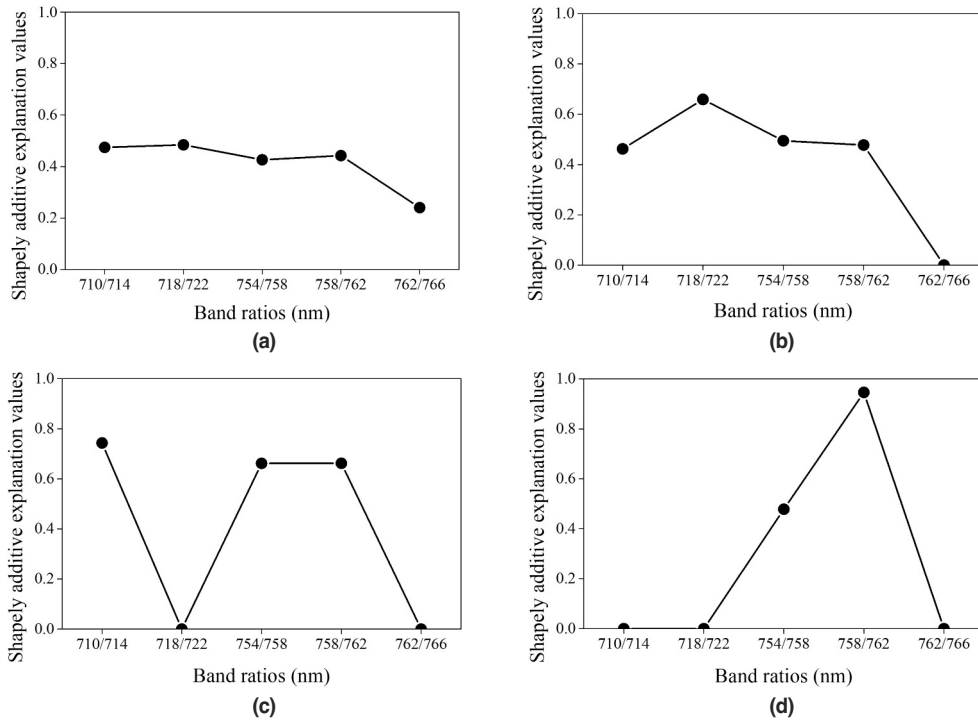


Fig. 6. Shapely additive explanation values depending on the number of key band ratios in the KNN-based model to estimate chlorophyll contents: (a) 5, (b) 4, (c) 3, and (d) 2.

4. 결론

무인기 기반의 초분광 영상을 이용하여 시계열에 따라 변하는 배나무 잎 엽록소 함량을 비파괴적으로 추정하기 위한 연구를 수행하였다. 결론적으로 4개의 key band ratios를 이용한 머신러닝 모델들의 추가 데이터셋이 확보된다면 충분한 엽록소 함량 추정 가능성을 보여줄 것이라고 기대된다. 모든 머신러닝 모델에서 full band ratios와 key band ratios의 validation 사이에서는 상대적으로 크지 않은 성능 저하 결과를 보였다. Calibration 성능이 가장 높았던 KNN 모델에서는 온전한 2개의 red edge 영역의 band ratios (710/714와 718/722)와 그 red edge 끝점에 2개의 band ratios (754/758과 758/762)들이 엽록소 함량 추정에 민감하게 반응한다는 것을 알 수 있었다. 그러나 본 연구 결과는 엽록소 함량 추정에 충분한 성능을 보이지 않았으며, 미미한 과적합의 영향이 있다는 한계점이 남아있다. 이를 극복하기 위한 해결책으로 실험 필드에서 추가 데이터셋을 확보하는 것과 동시에 다양한 환경 조건의 과원에 추정 모델을 재현해보는 것을 통해 입력 변수와 파라미터들을 조정해 나가는 과정이 필요하다. 이 연구는 배나무 잎의 엽록소 함량 추정을 통해 간접적으로 영양상태 및 직접적으로 노화 상태를 모니터링하여 안정적인 재배환경 구축을 통해 과실의 수확량과 품질을 유지하고 향상하는 연구에 도움을 줄 수 있을 것으로 기대한다.

사사

본 연구는 농촌진흥청 국립식량과학원 “사과, 배 생산성 향상을 위한 영상기반 정밀 생리·생태 진단기술 개발(과제번호: PJ0156572023)”의 지원으로 수행되었습니다.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- Abdulridha, J., Batuman, O., and Ampatzidis, Y., 2019. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sensing*, 11(11), 1373. <https://doi.org/10.3390/rs11111373>
- Acosta, M., Rodríguez-Carretero, I., Blasco, J., de Paz, J. M., and Quiñones, A., 2023. Non-Destructive appraisal of macro- and micronutrients in persimmon leaves using Vis/NIR hyperspectral imaging. *Agriculture*, 13(4), 916. <https://doi.org/10.3390/agriculture13040916>
- Aragon, B., Johansen, K., Parkes, S., Malbeteau, Y., Al-Mashharawi, S., Al-Amoudi, T. et al., 2020. A calibration procedure for field and UAV-based uncooled thermal infrared instruments. *Sensors*, 20(11), 3316. <https://doi.org/10.3390/s20113316>
- Bagheri, N., 2020. Application of aerial remote sensing technology for detection of fire blight infected pear trees. *Computers and Electronics in Agriculture*, 168, 105147. <https://doi.org/10.1016/j.compag.2019.105147>
- Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M. et al., 2022. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*, 2, 100049. <https://doi.org/10.1016/j.atech.2022.100049>
- Chen, S., Hu, T., Luo, L., He, Q., Zhang, S., Li, M. et al., 2020. Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods. *Infrared Physics & Technology*, 111, 103542. <https://doi.org/10.1016/j.infrared.2020.103542>
- Chen, R., Zhang, C., Xu, B., Zhu, Y., Zhao, F., Han, S. et al., 2022. Predicting individual apple tree yield using UAV multi-source remote sensing data and ensemble learning. *Computers and Electronics in Agriculture*, 201, 107275. <https://doi.org/10.1016>

j.compag.2022.107275

- Defernez, M., and Kemsley, E. K., 1999. Avoiding overfitting in the analysis of high-dimensional data with artificial neural networks (ANNs). *Analyst*, 124(11), 1675–1681.
- De Mol, C., De Vito, E., and Rosasco, L., 2009. Elastic-net regularization in learning theory. *Journal of Complexity*, 25(2), 201–230. <https://doi.org/10.1016/j.jco.2009.01.002>
- Feng, Z. H., Wang, L. Y., Yang, Z. Q., Zhang, Y. Y., Li, X., Song, L., He, L. et al., 2022. Hyperspectral monitoring of powdery mildew disease severity in wheat based on machine learning. *Frontiers in Plant Science*, 13, 828454. <https://doi.org/10.3389/fpls.2022.828454>
- Ge, H., Ma, F., Li, Z., Tan, Z., and Du, C., 2021. Improved accuracy of phenological detection in rice breeding by using ensemble models of machine learning based on UAV-RGB imagery. *Remote Sensing*, 13(14), 2678. <https://doi.org/10.3390/rs13142678>
- Gitelson, A. A., and Merzlyak, M. N., 1997. Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing*, 18(12), 2691–2697. <https://doi.org/10.1080/014311697217558>
- Hunt Jr, E. R., and Daughtry, C. S., 2018. What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture?. *International Journal of Remote Sensing*, 39(15–16), 5345–5376. <https://doi.org/10.1080/01431161.2017.1410300>
- Kannangara, K. P. M., Zhou, W., Ding, Z., and Hong, Z., 2022. Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method. *Journal of Rock Mechanics and Geotechnical Engineering*, 14(4), 1052–1063. <https://doi.org/10.1016/j.jrmge.2022.01.002>
- Kang, Y. S., Ryu, C. S., Kim, S. H., Jun, S. R., Jang, S. H., Park, J. W. et al., 2018. Yield prediction of Chinese cabbage (Brassicaceae) using broadband multispectral imagery mounted unmanned aerial system in the air and narrowband hyperspectral imagery on the ground. *Journal of Biosystems Engineering*, 43(2), 138–147. <https://doi.org/10.5307/JBE.2018.43.2.138>
- Kang, Y. S., Jang, S. H., Park, J. W., Song, H. Y., Ryu, C. S., Jun, S. R. et al., 2020. Yield prediction and validation of onion (*Allium cepa* L.) using key variables in narrowband hyperspectral imagery and effective accumulated temperature. *Computers and Electronics in Agriculture*, 178, 105667. <https://doi.org/10.1016/j.compag.2020.105667>
- Kang, Y. S., Park, J. W., Jang, S. H., Song, H. Y., Kang, K. S., Ryu, C. S. et al., 2021. Spectral band selection for detecting fire blight disease in pear trees by narrowband hyperspectral imagery. *Korean Journal of Agricultural and Forest Meteorology*, 23(1), 15–33. <https://doi.org/10.5532/KJAFM.2021.23.1.15>
- Kim, J. K., Kim, S. H., Lim, Y. J., and Seo, H. H., 2006. Anatomical feature, sugar, carbohydrate, and hormone content of normal and abnormal floral buds in ‘Niitaka’ pear trees. *Journals of the Korean Society for Horticultural Science*, 24(3), 354–358.
- Kim, H. C., Bae, K. S., Yoon, S. K., and Kim, T. C., 2018. Shoot growth and fruit characteristics of ‘Yumi’ peach in response to side dressing with fused phosphate fertilizer. *Korean Journal of Horticultural Science and Technology*, 36(5), 630–639. <https://doi.org/10.12972/kjhst.20180063>
- Kim, Y., Glenn, D. M., Park, J., Ngugi, H. K., and Lehman, B. L., 2011. Hyperspectral image analysis for water stress detection of apple trees. *Computers and Electronics in Agriculture*, 77(2), 155–160. <https://doi.org/10.1016/j.compag.2011.04.008>
- Mather, P. M., and Koch, M., 2011. *Computer processing of remotely-sensed images: An introduction*. John Wiley & Sons.

- Nalepa, J., 2021. Recent advances in multi-and hyperspectral image analysis. *Sensors*, 21(18), 6002. <https://doi.org/10.3390/s21186002>
- Naqvi, S. M. Z. A., Awais, M., Khan, F. S., Afzal, U., Naz, N., and Khan, M. I., 2021. Unmanned air vehicle based high resolution imagery for chlorophyll estimation using spectrally modified vegetation indices in vertical hierarchy of citrus grove. *Remote Sensing Applications: Society and Environment*, 23, 100596. <https://doi.org/10.1016/j.rsase.2021.100596>
- Paul, S., Poliyapram, V., İmamoğlu, N., Uto, K., Nakamura, R., and Kumar, D. N., 2020. Canopy averaged chlorophyll content prediction of pear trees using convolutional autoencoder on hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1426–1437. <https://doi.org/10.1109/JSTARS.2020.2983000>
- Pu, H., Kamruzzaman, M., and Sun, D. W., 2015. Selection of feature wavelengths for developing multispectral imaging systems for quality, safety and authenticity of muscle foods-a review. *Trends in Food Science & Technology*, 45(1), 86–104. <https://doi.org/10.1016/j.tifs.2015.05.006>
- Sharma, P., Leigh, L., Chang, J., Maimaitijiang, M., and Caffé, M., 2022. Above-ground biomass estimation in oats using UAV remote sensing and machine learning. *Sensors*, 22(2), 601. <https://doi.org/10.3390/s22020601>
- Ta, N., Chang, Q., and Zhang, Y., 2021. Estimation of apple tree leaf chlorophyll content based on machine learning methods. *Remote Sensing*, 13(19), 3902. <https://doi.org/10.3390/rs13193902>
- Vanegas, F., Bratanov, D., Powell, K., Weiss, J., and Gonzalez, F., 2018. A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data. *Sensors*, 18(1), 260. <https://doi.org/10.3390/s18010260>
- Wang, J., Zhou, Q., Shang, J., Liu, C., Zhuang, T., Ding, J. et al., 2021. UAV- and machine learning-based retrieval of wheat SPAD values at the overwintering stage for variety screening. *Remote Sensing*, 13(24), 5166. <https://doi.org/10.3390/rs13245166>
- Yu, Q., Mickler, R. A., Liang, T., Liu, Y., Jiang, J., Song, K. et al., 2022. Hyperspectral differences between sunlit and shaded leaves in a Manchurian ash canopy in Northeast China. *Remote Sensing Letters*, 13(8), 800–811. <https://doi.org/10.1080/2150704X.2022.2088255>
- Zhang, Y., Gao, J., Cen, H., Lu, Y., Yu, X., He, Y. et al., 2019. Automated spectral feature extraction from hyperspectral images to differentiate weedy rice and barnyard grass from a rice crop. *Computers and Electronics in Agriculture*, 159, 42–49. <https://doi.org/10.1016/j.compag.2019.02.018>