

머신러닝 기반 초분광 이미지를 이용한 생전분의 비파괴 분류: 입도 및 수분함량의 영향

이다은¹ · 정경아² · 이창주² · 이수용^{1,*}

¹세종대학교 식품생명공학과, ²원광대학교 식품생명공학과

Nondestructive classification of native starches by using machine learning-assisted hyperspectral imaging: effects of particle size and moisture content

Da Eun Lee¹, Gyeong A Jeong², Chang Joo Lee² and Suyong Lee^{1,*}

¹Department of Food Science and Biotechnology, Sejong University, Seoul 05006, Korea

²Department of Food Science and Biotechnology, Wonkwang University, Iksan 54538, Korea

Abstract

In this study, a framework has been presented for classifying native starches by using hyperspectral imaging in combination with machine learning, to improve the classification performance of machine learning models under controlled preprocessing conditions such as particle size and moisture content. Hyperspectral analysis revealed a substantial overlap among the different types of native starches, indicating limitations in the classification on the basis of raw hyperspectral data. The starch classification performance of the machine learning model varied depending on the particle size and moisture content of the starch. The highest classification accuracy (~97%) was achieved at a particle size of 100 mesh and moisture content of 50%, and particle aggregation was effectively suppressed to maintain a uniform dispersion state. These results were confirmed using microstructural analysis of the starches on the basis of their scanning microscopic images and particle size distributions. Overall, the proposed preprocessing approach demonstrates the strong potential of machine learning-combined hyperspectral technology as a rapid and nondestructive analytical method for identifying native starches in the food industry.

Keywords: Artificial intelligence, Spectroscopy, Cereal, Non-invasive, Food

서론

전분은 식품 산업을 비롯한 제약, 화학 등 다양한 바이오 산업 전반에서 널리 활용되는 핵심 소재로서, 원료 전분의 이화학적 특성은 최종 제품의 품질에 직접적인 영향을 미친다. 특히, 생전분(native starch)은 식물학적 유래(botanical origin)에 따라 입도, 수분 함량, 결정 구조 등의 특성이 달라지며, 이러한 특성은 이후 가공

공정 및 제품 품질을 결정짓는 중요한 인자로 작용한다(Comejo-Ramírez et al., 2018). 따라서 생전분을 신속하고 정확하게 판별하는 기술은 원료 관리, 공정 제어 및 품질 관리 측면에서 매우 중요한 요소이다.

현재 생전분의 품질 및 특성 평가는 주로 화학적 특성 규명, 열적 특성 평가, 물성 및 분광학적 분석 등 다양한 방법을 통해 이루어지고 있다(Govindaraju et al., 2021). 그러나 이러한 기존 분석 방

Received: May 06, 2026 / Revised: May 29, 2026 / Accepted: Jun 08, 2026 / Published: Jul 01, 2026

Corresponding author: Suyong Lee, Department of Food Science and Biotechnology, Sejong University, Seoul 05006, Korea

E-mail: suyonglee@sejong.ac.kr

Copyright © 2026 Korean Society for Food Engineering.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://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.

법들은 시료의 전처리 과정이 비교적 복잡하고 분석 시간이 길며, 측정 과정에서 시료가 파괴되거나 본래의 이화학적 특성이 변질될 수 있는 한계가 존재한다. 특히, 대량의 시료를 실시간으로 평가해야 하는 산업 현장에서는 이러한 방법들을 신속하게 적용하기에는 한계가 있다.

최근 이러한 한계를 극복하기 위한 대안으로 초분광 분석(hyperspectral analysis)이 주목받고 있다. 초분광 분석은 수십에서 수백 개 이상의 연속적인 파장 대역에서 시료의 분광 정보를 동시에 획득함으로써, 공간 정보와 분광 정보가 결합된 고차원 데이터 큐브를 생성하는 기술이다(Liu et al., 2017). 초분광 분석을 통해 시료의 화학적 조성, 구조적 특성 및 물리·화학적 변화를 비파괴적으로 분석할 수 있으며, 육안으로 구별하기 어려운 미세한 차이까지 탐지가 가능하다. 그러나 초분광 데이터는 높은 차원성과 복잡성을 가져서 기존의 전통적인 분석 기법만으로는 처리에 한계가 있다. 따라서, 초분광 데이터의 패턴을 학습하고 특징을 추출할 수 있는 머신러닝 기법과의 결합이 활발히 이루어지고 있으며, 머신러닝 기반 분석은 대규모 분광 데이터로부터 비선형적 특성과 잠재적 패턴을 효과적으로 학습할 수 있다. 특히, 초분광 분석 및 머신러닝 기술의 결합은 항공, 환경, 농업 등 다양한 분야에서 활용되고 있으며, 최근 식품 분야에서도 적용되고 있다(Yoon, 2022). 예를 들어, 고추와 같은 농산물에서 이물질을 식별하거나(Shu et al., 2023), 절단 채소에서 이물질을 검출하기 위한 연구가 수행된 바 있다(Tunny et al., 2023). 이와 같이 현재까지 초분광 기술을 식품에 적용한 사례는 주로 이물질 검출이나 품질 이상 판별 등 제한적 영역에 국한되어 있다. 곡류 식품 분야에서 초분광을 전분에 적용한 몇몇 연구가 보고되고 있지만, 감자(Wang et al., 2021), 쌀(Lu et al., 2017), 옥수수(Qiao et al., 2024)의 전분 함량을 정량화하는 제한적 사례만 보고되고 있다. 초분광 기술은 비파괴비접촉 분석이 가능하다는 점에서 공정 모니터링 및 실시간 품질 관리에 적합한 기술로 평가되지만, 초분광 분석을 생전분에 적용하여 신속하고 정확하게 전분의 종류를 판별하는 연구는 아직 보고된 바가 거의 없는 실정이다. 이는 생전분이 미세한 분말 형태로 존재하여 입자 간 광 산란 및 분광 신호의 중첩이 발생할 수 있으며, 이로 인해 시료 간 분광 특성의 차이가 감소하여 정확한 판별이 어렵기 때문이다(Fu et al., 2021). 이러한 특성은 초분광 데이터의 변별력을 저하시켜 전분의 정확한 판별을 어렵게 만드는 주요 요인으로 작용할 수 있다.

따라서, 본 연구에서는 전분의 입도 및 수분 함량을 제어하는 전처리 과정을 통해 생전분 시료 간 분광 특성의 차이를 증폭시키고, 이로 인해 초분광 데이터의 변별력을 향상시키고자 하였다. 그 후, 인공지능 모델을 적용하여 전처리 조건에 따른 생전분의 판별 정확도를 평가하였다.

재료 및 방법

재료

본 실험에서는 옥수수 전분(Touevan, Gyeonggi-do, Korea), 감자 전분(Touevan), 고구마 전분(Touevan), 타피오카 전분(Dong-A Foods, Gyeonggi-do, Korea), 찹옥수수 전분(Samyang Co., Ulsan, Korea), 밀 전분(Daemyung FM, Gyeonggi-do, Korea) 등 총 6가지 생전분을 시장에서 구매하여 사용하였다.

전분 전처리 방법

생전분 시료의 전처리 조건으로써 입도와 수분함량을 조절하였다. 전분 시료를 체분리(Chung Gye Sieve, Seoul, Korea)하여 각각 80, 100, 120, 140, 270 mesh 크기로 분리하였다. 수분함량의 경우 30%, 40%, 50%, 60%가 되도록 전처리하였다. 이를 위해, 각 생전분 시료의 초기 수분함량을 열풍 건조 방법(105°C)으로 측정 한 후, 목표 수분함량에 도달하도록 시료에 일정량의 증류수를 첨가하고, 균일한 수분 분포를 확보하기 위해 믹서기(KitchenAid, St. Joseph, MI, USA)를 이용하여 5분간 혼합하였다. 특히, 수분 증발이 일어나지 않도록 믹서기 주변을 밀봉한 상태로 혼합을 진행하였고, 혼합된 시료는 측정 전까지 완전 밀봉 상태로 보관하였다.

초분광 측정

생전분 시료의 초분광 이미지 분석을 위한 실험 세팅은 Fig. 1에서 보여주고 있다. 초분광 분석을 위해 900~1,700 nm 파장 영역을 포함하는 초분광 카메라(Specim FX17, Spectral Imaging Ltd., Oulu, Finland)를 사용하였고, 해당 파장 영역은 224개의 연속적인 분광 밴드로 구성되었다. 전분 시료(20 g)를 납작한 플라스틱 용기에 담고 표면을 평평하게 정리한 후, 초분광 시스템의 이동식 스테이지 위에 올려 놓았다. 광원으로 할로겐 램프(120 W) 6개를 사용하여, 암실 환경에서 초분광 측정을 수행하였다. 이동식 스테이지

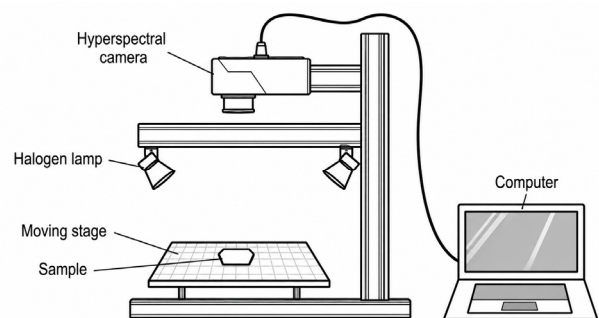


Fig. 1. Schematic diagram of hyperspectral imaging.

는 40 mm/s의 속도로 이동하였으며, 카메라의 노출 시간은 3 ms, 카메라와 시료 사이의 거리는 330 mm로 설정하였다. 시료 고유의 반사 특성을 반영하기 위해, 다음 식을 이용하여 원본 이미지(original image, OI)를 상대 반사율 이미지(relative reflectance image, R)로 보정하여 사용하였다.

$$R = \frac{OI - BI}{WI - BI} \times 100\%$$

BI: 렌즈를 차단한 상태에서 획득한 암 이미지(dark image)

WI: 표준 백색 기준판을 이용하여 획득한 백색 이미지(white reference image)

측정된 초분광 이미지는 중심부에서 50×50 픽셀 크기의 관심 영역(ROI)을 설정하여 픽셀 단위의 분광 데이터를 추출하였으며, 이를 평균화하여 시료별 대표 스펙트럼을 생성하였다.

초분광 데이터의 머신러닝 분석

초분광 데이터의 머신러닝 분석은 주피터 노트북(Jupyter Notebook) 오픈소스 프로그램을 사용하여 파이썬 프로그래밍 언어 (Version 3.12.7)로 분석하였다. 사용된 데이터셋에서 입력 데이터 (input data, X)는 224개의 분광 밴드이고, 출력 데이터 (output data, Y)는 6가지 전분 레이블로 구성하였다. 우선, 초분광 데이터의 관심 영역(ROI)으로부터 추출된 스펙트럼 데이터를 Standard Normal Variate와 Savitzky-Golay 알고리즘을 적용하였다. 이후 초분광 데이터는 k-겹 교차 검증(k-fold cross validation, k=5)을 적용하여 분류 성능을 평가하였다. 전체 데이터셋은 무작위로 5개의 부분 집합으로 분할되었으며, 이 중 4개 집합(약 80%)을 학습 데이터로 사용하고, 나머지 1개 집합(약 20%)을 검증 데이터로 사용하였다. 해당 과정은 총 5회 반복 수행하여 데이터 분할에 따른 편차를 최소화하였고, 최종 결과는 교차 검증을 통해 얻어진 평균값과 표준편차로 제시하였다. 초분광 기반 생전분 판별을 위하여 초분광 데이터를 4개 머신러닝 알고리즘(Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine)에 적용하여 시료의 분류 성능을 평가하였다. 아울러, 데이터 분산을 설명하는 주요 요인을 파악하고 데이터의 시각화를 위해 주성분 분석(Principal Component Analysis, PCA)를 수행하였다. 구축된 머신러닝 모델의 분류 성능을 평가하기 위한 지표로 정확도(accuracy)뿐만 아니라 정밀도(precision), 재현율(recall) 및 F1 점수(F1-score)를 함께 고려하였다. 또한, 보다 정량적인 비교를 위해 혼동 행렬(confusion matrix)을 이용하여 각 전분 시료 간 정분류 및 오분류 양상을 분석하였다.

주사전자현미경 이미지 분석

전처리에 따른 전분 시료의 미세 구조 특성 분석을 위해 주사전자현미경(SEM, JSM-6390, JEOL, Tokyo, Japan)을 이용한 이미지 분석을 수행하였다. SEM 분석을 위해 동결건조 된 전분 시료를 백금(Platinum)으로 코팅 처리한 후, 가속 전압 10 kV (mode3) 조건 하에 500배 배율에서 측정하였다.

입도 분석

각 시료의 입자 크기 분포는 레이저 회절 방식의 입도분석기 (PSA 1090 LD, Anton Paar, Graz, Australia)를 이용하여 측정하였다. 약 0.1 g의 시료를 10 mL의 ethanol과 혼합한 후, 이 현탁액을 주입하여 가루의 체적 기반 입자 크기 분포를 분석하였다. 시료의 균일한 분산과 분석 중 고형화 방지를 위해 초음파와 교반기를 사용하며 측정하였다. Span 값은 입자 크기 분포의 폭을 나타내는 지표로 사용하였으며, 전체 체적의 10% (D10), 50% (D50), 90% (D90) 이하에 해당하는 입자 크기를 각각 나타냈다.

통계 처리

각 처리군에 대해 독립적으로 3개 반복 시료를 준비하였고, 초분광 측정은 각 시료를 90도씩 회전시키면서 4번 수행하였다. 이에 따라, 처리군당 총 12회 초분광 측정이 이루어졌고 측정된 각 이미지로부터 25개의 ROI를 추출하여 머신러닝 데이터셋은 총 1,800개 (12 측정×25 ROI×6개 시료)로 구성되었다. 그 외 실험은 3회 반복 측정하였고, 측정된 결과는 평균 및 표준편차로 나타내었다.

결과 및 고찰

Fig. 2는 입도(80, 100, 120, 140, 270 mesh)에 따른 전분의 초분광 스펙트럼을 보여주고 있다. 본 연구에 사용된 전분 시료들 모두 유사한 초분광 스펙트럼 패턴을 보여주었는데 특히, 1,200 nm와 1,450 nm 부근에서 주요한 반사율 신호를 보여주었다. 1,200 nm 부근의 신호는 전분에 존재하는 CH₃ 및 CH₂에 기인한 C-H 결합과 관련된 밴드로 알려져 있으며, 1,450 nm 영역은 물 분자에 기인한 O-H 결합의 밴드에 해당한다(Schwanninger et al., 2011). 이러한 초분광 스펙트럼은 밀가루(Hwang et al., 2025) 및 흑미(Amanah et al., 2021) 등의 초분광 특성 연구에서도 유사하게 관찰되었다. 전체적으로 감자 전분의 경우 다른 전분들과는 확연히 구별되는 낮은 초분광 스펙트럼을 보여주었지만, 그 외 다른 전분들의 초분광 곡선은 서로 중첩되어 관찰되었다. 전분 입도에 따른 초분광 곡선을 관찰 시, 입도가 감소(mesh 값 증가)할수록 중첩된 초분광 곡선이 어느 정도 분리되는 현상이 관찰되었지만, 100~120 mesh 이후로는 분리 효과

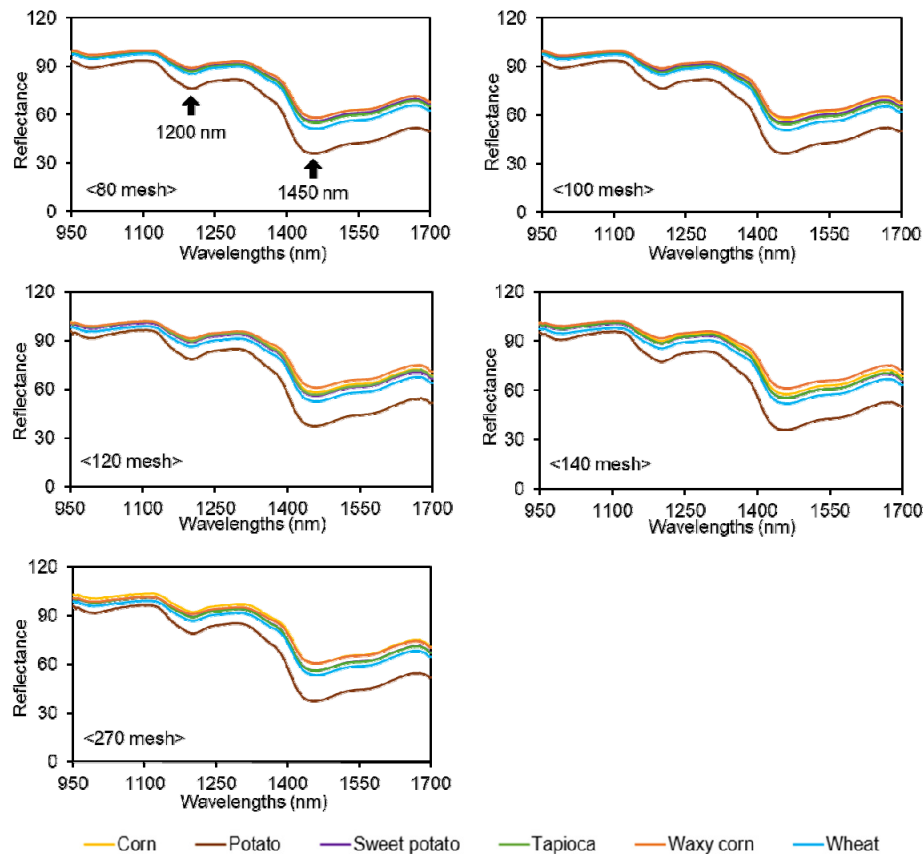


Fig. 2. Hyperspectral reflectance curves of native starches with different particle sizes.

가 미미하였다. 이는 입자의 크기와 표면 구조에 따라 빛의 산란 정도와 경로가 달라져 초분광 밴드 형태가 변하기 때문으로 해석된다 (Pasikatan et al., 2001). 그러나 일정 크기(100~120 mesh) 이하에서는 입도 감소에 따른 산란 특성 변화가 제한적이어서 스펙트럼 분리 효과가 더 이상 뚜렷하게 나타나지 않는 것으로 판단된다.

전분 입도가 초분광 판별 성능에 미치는 영향을 분석하기 위하여, Fig. 2의 초분광 스펙트럼을 4가지 머신러닝 모델 즉, Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine에 적용하여 판별 정확도를 분석하였다. Decision Tree는 입력 데이터를 특정 기준에 따라 반복적으로 분할하여 트리(tree) 구조를 형성하고, 이를 기반으로 분류 또는 회귀를 수행하는 대표적인 지도학습 알고리즘이다 (Song & Lu, 2015). Random Forest는 다수의 Decision tree를 기반으로 한 앙상블(ensemble) 방법으로, 훈련 데이터의 부트스트랩(bootstrap) 표본과 무작위로 선택된 변수 부분집합을 사용하여 각각의 Decision tree를 독립적으로 학습시킨다. 이후 각 트리의 예측 결과를 결합하여 최종 예측을 수행하며, 분류 문제에서는 다수결 투표를, 회귀 문제에서는 평균값을 이용한다(Breiman, 2001). K-Nearest Neighbor는 새로운 데이터와 훈련 데이터 사이의 거리를 기반으로 가장 가까운 K개의 이

웃을 탐색하고, 이들의 다수 클래스를 기준으로 분류를 수행하는 알고리즘이며(Cover & Hart, 1967), Support Vector Machine은 서로 다른 클래스 간의 분리를 위한 최적의 초평면을 찾고, 각 클래스에 가장 가까운 데이터 포인트와 초평면 사이의 마진을 최대화하도록 학습하는 지도학습 알고리즘이다(Cortes & Vapnik, 1995).

Table 1은 전분 입도에 따른 초분광 스펙트럼을 Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine에 각각 적용하여 얻어진 학습(train) 및 테스트(test) 데이터셋의 판별 정확도(accuracy)를 보여주고 있다. 전체적으로, Decision Tree, Random Forest, K-Nearest Neighbor 모델이 높은 정확도를 보여주었지만, 학습 및 테스트 데이터셋의 정확도 차이가 크게 관찰되었다. 따라서, 학습 데이터는 높은 성능을 보이지만, 새로운 데이터에 대해서는 예측 성능이 저하되는 과적합(overfitting) 문제가 발생한 것으로 판단되었다. 반면에, Support Vector Machine 모델의 경우 학습 및 테스트 데이터셋 간 정확도 차이가 상대적으로 작게 나타나 과적합이 비교적 적은 것으로 판단되었다. 비록 테스트 정확도는 다른 모델보다 다소 낮게 나타났으나, 새로운 데이터에 대해서도 비교적 안정적인 예측 성능을 유지하였다. 입도가 가장 큰 80 mesh 전분의 경우 모든 머신러닝 모델에서

Table 1. Performance of machine learning models to classify native starches with different particle sizes

Particle size		Decision tree	Random forest	K-nearest neighbor	Support vector machine
80 mesh	Train	0.994±0.003	0.981±0.004	0.931±0.005	0.813±0.009
	Test	0.863±0.017	0.870±0.020	0.873±0.024	0.808±0.030
100 mesh	Train	0.999±0.001	0.980±0.003	0.937±0.004	0.855±0.003
	Test	0.881±0.016	0.896±0.010	0.908±0.009	0.849±0.013
120 mesh	Train	0.997±0.002	0.971±0.003	0.935±0.001	0.885±0.002
	Test	0.861±0.010	0.884±0.008	0.900±0.007	0.882±0.011
140 mesh	Train	0.997±0.001	0.972±0.003	0.941±0.003	0.870±0.003
	Test	0.869±0.036	0.890±0.020	0.904±0.011	0.868±0.013
270 mesh	Train	0.990±0.001	0.966±0.005	0.903±0.006	0.819±0.003
	Test	0.821±0.014	0.851±0.019	0.840±0.013	0.811±0.015

테스트 데이터셋의 정확도 80.8~87.3% 정도로 가장 낮게 관찰된 반면, 100 mesh에서 가장 높은 정확도(약 90.8%)를 보여주었다. 이후, 입도가 더 작아질수록 다시 정확도가 감소하는 경향을 보여주었다.

전분 수분함량이 초분광을 통한 전분 판별 정확도에 미치는 영향을 분석하기 위해, 전분의 수분함량을 30, 40, 50, 60%로 조절한 후 초분광 분석을 실시하였다(Fig. 3). 30% 수분함량에서 감자 전분 및 밀 전분의 경우 낮은 반사신호를 보여주어 다른 전분과 확연히 구별되었다. 반면, 다른 전분들의 초분광 곡선은 상호 중첩되어 관

찰되었다. 수분함량이 50%까지 증가 시 전분 시료별 초분광 곡선 차이가 커지면서, 수분함량 30% 시료와는 다른 초분광 패턴을 보여주었다. 하지만, 옥수수과 밀 전분, 그리고 고구마와 타피오카 전분의 초분광 곡선은 많은 부분 겹쳐져 있는 경향이 관찰되었다. 수분함량을 60%까지 증가시킬 경우 높아진 수분의 영향으로 오히려 초분광 곡선의 시료 간 차이가 감소하였다.

전분 수분함량에 따른 초분광 스펙트럼(Fig. 3)을 4가지 머신러닝 모델, 즉 Decision Tree, Random Forest, K-Nearest Neighbor, Support

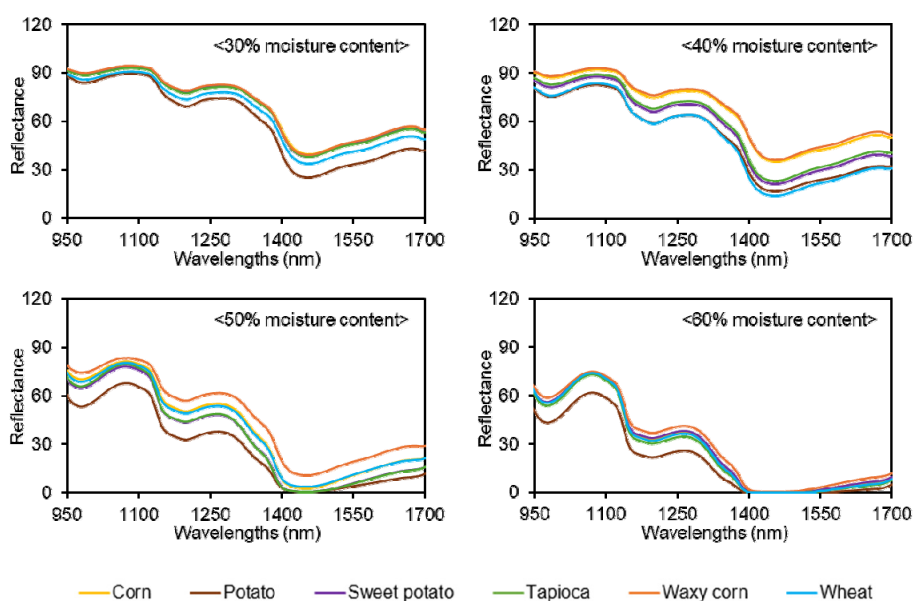


Fig. 3. Hyperspectral reflectance curves of native starches with different moisture contents.

Table 2. Performance of machine learning models to classify native starches with different moisture contents

Moisture content		Decision tree	Random forest	K-nearest neighbor	Support vector machine
30%	Train	0.942±0.007	0.890±0.009	0.796±0.008	0.549±0.008
	Test	0.658±0.023	0.697±0.023	0.681±0.020	0.539±0.040
40%	Train	1.000±0.000	0.996±0.001	0.973±0.003	0.938±0.004
	Test	0.957±0.007	0.968±0.011	0.963±0.011	0.934±0.011
50%	Train	0.999±0.001	0.996±0.001	0.983±0.003	0.943±0.003
	Test	0.953±0.009	0.970±0.007	0.971±0.007	0.942±0.011
60%	Train	0.999±0.001	0.995±0.002	0.985±0.003	0.811±0.005
	Test	0.961±0.013	0.967±0.010	0.973±0.005	0.805±0.017

Vector Machine에 각각 적용하여 판별 정확도를 비교 분석하였다 (Table 2). 전체적으로 전분의 수분함량을 증가시키기에 따라 판별 정확도가 점진적으로 향상되어, 수분함량이 50%일 때 가장 높은 판별 정확도(약 97%)를 보여주었다. 반면, 수분함량을 60%로 증가시킨 경우 모든 4가지 머신러닝 모델에서 판별 정확도가 오히려 감소하는 현상이 관찰되었다. 이는 수분함량이 증가할수록 수분 신호 밴드의 강도가 크게 증가하여 전분 고유의 결합에 기인한 분광 정보를 방해하는 것으로 판단된다(Reeves, 1993). 또한, 과잉 수분 조건에서는 전분과 수분 간의 수소 결합 네트워크가 형성되면서 스펙트럼의 흡수 강도 변화와 파장 이동이 동시에 발생하여 전분 고유의 스펙트럼 패턴이 변형될 수 있다고 보고 되었다(Wesley & Blakeney, 2001). 아울러, 높은 수분함량은 시료 표면에 수막을 형성하여 산란 특성 변화가 발생한 것으로 판단된다(Bogomolov et al., 2018). 머신러닝 모델의 성능을 비교한 결과, Random Forest와 K-Nearest Neighbor 모델이 전반적으로 가장 높은 판별 정확도를 나타내었다. Decision Tree 모델 역시 높은 정확도를 나타내었으나, 학습 및 테스트 데이터셋 간 정확도 차이가 상대적으로 크게 나타나 일부 과적합 경향이 관찰되었다. 반면, Support Vector Machine 모델은 다른 모델에 비해 전반적으로 낮은 정확도를 나타내었다. 이는 수분함량 변화에 따른 초분광 스펙트럼의 차이가 크지 않고 클래스 간 중첩이 발생함에 따라, 클래스 간 최대 마진을 기반으로 결정 경계(decision boundary)를 설정하기 어렵기 때문에 Support Vector Machine 모델에서 분류 성능이 저하된 것으로 판단된다(Mahmood et al., 2025).

수분 처리에 따른 전분 판별 개선 효과를 시각적으로 나타내기 위해 초분광 스펙트럼 데이터를 PCA로 차원축소하여 3차원으로 나타내었다(Fig. 4). 전처리를 하지 않은 생전분의 경우 감자 전분만이 다른 전분들과 뚜렷하게 분리되는 경향을 보였으며, 그 외 전

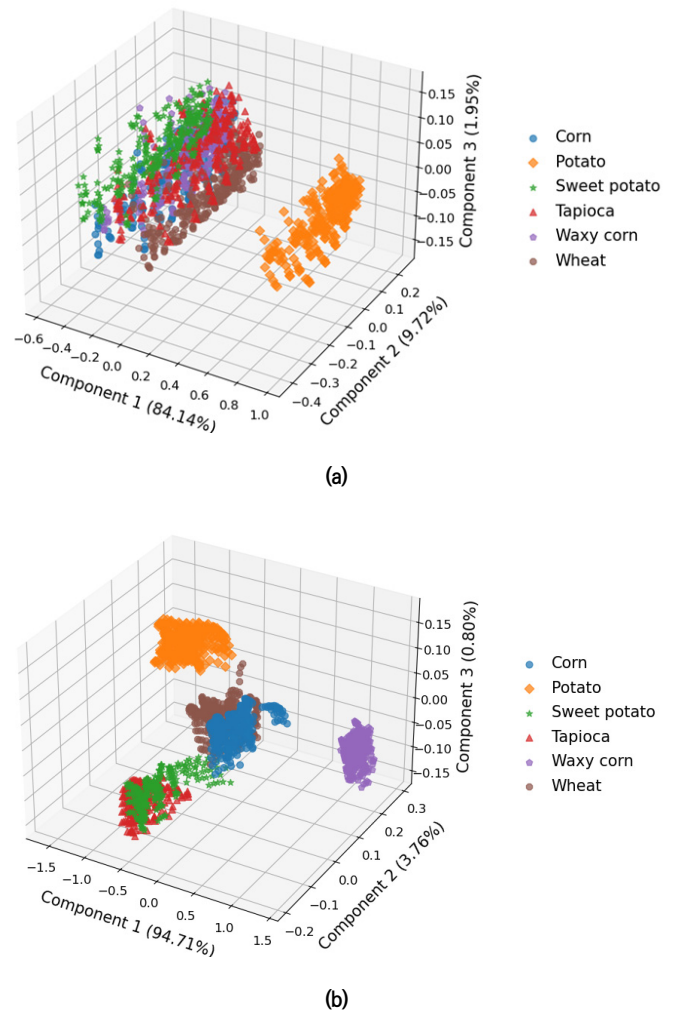


Fig. 4. Hyperspectral image-based PCA score plot of native starches (a) and starches pretreated at 50% moisture content (b).

분은 서로 중첩되는 분포를 보인다. 반면, 50% 수분함량으로 전처리한 경우, 감자 전분과 찹옥수수 전분은 완벽하게 분리된 군집을 형성하였다. 그 외 전분의 경우에도 서로 가깝게 위치하고 있지만, 뚜렷한 군집을 형성하는 것으로 관찰되었다.

Table 2에서 보여주고 있는 4가지 머신러닝 모델의 전분 판별 성능을 혼동 행렬(confusion matrix)을 사용하여 좀 더 세부적으로 분석하였다(Fig. 5). 혼동행렬은 지도 학습 기반 분류 알고리즘의 성능을 시각화하기 위한 표 형태의 구조로 되어 있다. 분류 모델이 각 클래스를 어떻게 예측하고 어떤 클래스를 혼동하는지 직관적으로 파악할 수 있다. 일반적으로 각 행은 실제 클래스(actual class)를 나타내고 각 열은 예측 클래스(predicted class)를 나타낸다(Haghighi et al., 2018). Fig. 5(a)의 혼동 행렬을 살펴보면, 수분 전처리 하지 않은 생전분의 경우 전분 종류 간 오분류가 현저히 많이 발생하였다. 세부적으로는 옥수수 전분과 찹옥수수 전분 간 상호 혼동을 보이며, 타피오카 전분은 고구마 전분과 밀 전분으로 혼동되어 분류되었다. 즉, 초분광 곡선이 서로 겹치는 전분에서 오분류가 많이 발생

하였다. 반면, 50% 수분함량으로 전처리한 경우(Fig. 5(b)), 대부분의 전분 시료에 대해 대각선 부분의 수치가 증가하여 높은 정확도로 전분이 판별되었다. 특히, 오분류 판별이 많이 관찰되었던, 옥수수 전분, 타피오카 전분과 찹옥수수 전분에서 정분류 빈도가 증가하고 오분류 사례가 현저히 감소하였다. 이는 적정 수준의 수분 처리를 통해 생전분 시료의 분광적 특성이 보다 명확히 드러나고, 그 결과 머신러닝 모델이 전분 간 차이를 효과적으로 학습할 수 있었기 때문으로 판단된다. 아울러, Fig. 5의 Precision, recall, F1-score 값의 경우 수분함량을 50%로 조절 시 약 10% 정도 향상된 0.95 이상의 매우 높은 값을 보여주었다.

입도 및 수분 처리에 따른 초분광 기반 머신러닝 모델의 판별 성능 개선 효과를 분석하기 위해 주사전자현미경(scanning electron microscopy, SEM)을 이용하여 전분 시료의 미세구조적 특성 및 입자 분산 상태를 분석하였다(Fig. 6). 분석 결과, 수분 처리를 수행하지 않은 생전분 시료에서는 전분 입자들이 국부적으로 응집되어 불균일하게 뭉쳐 있는 구조가 관찰되었다. 이는 입자 간 물리적 결합

<Decision Tree>							<Random Forest>							<K-Nearest Neighbor>							<Support Vector Machine>						
	Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat		Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat		Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat		Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat
Corn	256	0	0	8	36	0	Corn	261	0	0	4	35	0	Corn	252	0	0	2	46	0	Corn	172	0	0	0	128	0
Potato	0	300	0	0	0	0	Potato	0	300	0	0	0	0	Potato	0	300	0	0	0	0	Potato	0	300	0	0	0	0
Sweet potato	0	0	283	16	0	1	Sweet potato	0	0	292	8	0	0	Sweet potato	0	0	296	4	0	0	Sweet potato	0	0	277	23	0	0
Tapioca	3	0	29	221	14	33	Tapioca	7	0	28	224	10	31	Tapioca	3	0	35	229	8	25	Tapioca	19	0	29	214	4	34
Waxy corn	55	0	1	12	232	0	Waxy corn	66	0	3	2	229	0	Waxy corn	57	0	0	6	237	0	Waxy corn	63	0	0	7	230	0
Wheat	0	0	2	37	0	261	Wheat	0	0	5	35	0	260	Wheat	0	0	4	38	0	258	Wheat	0	0	5	33	0	262
Precision							Precision							Precision							Precision						
	0.863±0.018							0.872±0.021							0.874±0.024							0.811±0.030					
Recall	0.863±0.017						Recall	0.870±0.020						Recall	0.873±0.024						Recall	0.808±0.030					
F1 score	0.862±0.017						F1 score	0.868±0.021						F1 score	0.872±0.025						F1 score	0.807±0.029					

(a)

<Decision Tree>							<Random Forest>							<K-Nearest Neighbor>							<Support Vector Machine>						
	Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat		Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat		Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat		Corn	Potato	Sweet potato	Tapioca	Waxy corn	Wheat
Corn	288	0	0	0	0	12	Corn	300	0	0	0	0	0	Corn	300	0	0	0	0	0	Corn	300	0	0	0	0	0
Potato	0	300	0	0	0	0	Potato	0	300	0	0	0	0	Potato	0	300	0	0	0	0	Potato	0	300	0	0	0	0
Sweet potato	1	0	277	22	0	0	Sweet potato	1	0	282	17	0	0	Sweet potato	0	0	289	11	0	0	Sweet potato	0	0	265	35	0	0
Tapioca	0	0	31	269	0	0	Tapioca	0	0	29	271	0	0	Tapioca	0	0	36	264	0	0	Tapioca	0	0	64	236	0	0
Waxy corn	0	0	0	0	300	0	Waxy corn	0	0	0	0	300	0	Waxy corn	0	0	0	0	300	0	Waxy corn	0	0	0	0	300	0
Wheat	19	0	0	0	0	281	Wheat	7	0	0	0	0	293	Wheat	5	0	0	0	0	295	Wheat	5	0	0	0	0	295
Precision							Precision							Precision							Precision						
	0.955±0.014							0.971±0.007							0.965±0.005							0.969±0.007					
Recall	0.954±0.014						Recall	0.971±0.007						Recall	0.965±0.006						Recall	0.969±0.006					
F1 score	0.954±0.014						F1 score	0.971±0.007						F1 score	0.965±0.006						F1 score	0.969±0.006					

(b)

Fig. 5. Confusion matrix of machine learning models for native starches (a) and starches pretreated at 50% moisture content (b).

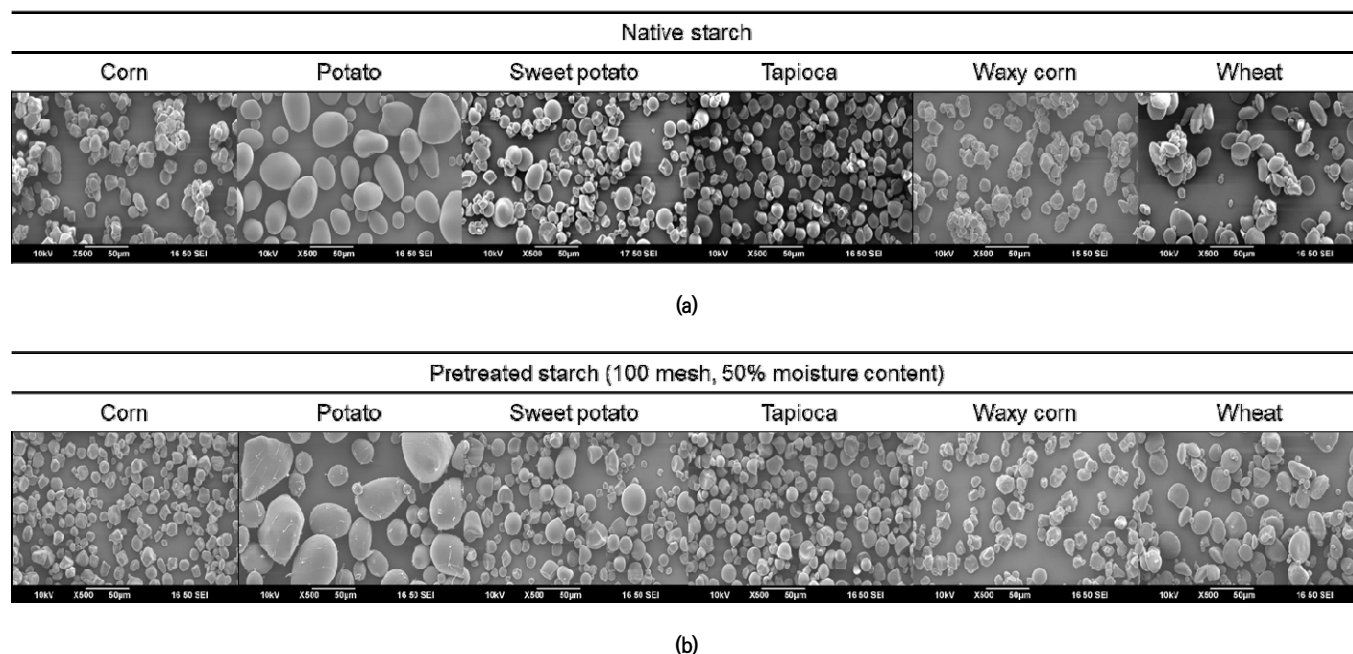


Fig. 6. Scanning microscopic images of native starches (a) and starches pretreated at 50% moisture content (b).

및 정전기적 상호작용 등에 의해 입자 분산성이 저하된 상태로 해석된다. 반면, 약 50%의 수분함량으로 처리된 전분 시료에서는 입자간 응집 현상이 현저히 감소하였으며, 개별 전분 입자들이 비교적 균일하게 분산된 형태를 나타내었다.

Table 3은 전분 시료의 입도 분포를 정량적으로 분석한 결과를 보여주고 있다. 입도는 입자 체적의 평균 직경과 시료 부피 10%, 50%, 90%에 해당하는 입자크기, 입자들의 분포 폭(span)으로 나타내었다. 생전분 시료의 경우 평균 입자 크기(mean size volume)가 약 23~48 μm 범위로 나타났으며, 상대적으로 큰 span 값을 보여 입자 크기 분포가 넓고 불균

일한 특성을 보였다. 반면, 50%의 수분함량으로 처리된 전분 시료에서는 평균 입자 크기는 생전분보다 작게 관찰되었다. 그리고 전체적으로 표준편차와 span 또한 감소하여 보다 균일한 입도 분포를 나타내었다. 이러한 결과는 SEM 분석에서 확인된 바와 같이 수분 처리를 통해 전분 입자의 응집이 완화되고 입자 분산성이 향상되었음을 의미한다. 특히, 옥수수 전분과 찹옥수수 전분의 경우 전처리 전에는 평균 입자 크기가 모두 약 39 μm 으로 유사하였지만 수분처리 후 각각 19.4 μm , 21.2 μm 로 감소하여 확연한 입도 차이가 관찰되었는데, 혼동행렬에서도 정분류가 높아진 결과와 일치하였다. 전처리하지 않은 생전분 시료에 비해, 일

Table 3. Particle size distribution of native starches and starches pretreated at 50% moisture content

Sample	D ₁₀ (μm)	D ₅₀ (μm)	D ₉₀ (μm)	Mean size (μm)	Span
Corn	14.7 $\pm$ 0.01	33.2 $\pm$ 0.05	68.1 $\pm$ 0.14	39.0 $\pm$ 0.06	1.61 $\pm$ 0.00
Corn (pretreated)	7.70 $\pm$ 0.01	16.6 $\pm$ 0.04	33.7 $\pm$ 0.04	19.4 $\pm$ 0.03	1.56 $\pm$ 0.00
Potato	19.3 $\pm$ 0.01	41.8 $\pm$ 0.03	82.7 $\pm$ 0.07	48.5 $\pm$ 0.03	1.52 $\pm$ 0.00
Potato (pretreated)	21.1 $\pm$ 0.02	44.1 $\pm$ 0.04	85.6 $\pm$ 0.09	50.9 $\pm$ 0.04	1.46 $\pm$ 0.00
Sweet potato	12.2 $\pm$ 0.01	26.8 $\pm$ 0.03	54.0 $\pm$ 0.06	31.2 $\pm$ 0.03	1.56 $\pm$ 0.00
Sweet potato (pretreated)	7.41 $\pm$ 0.02	17.0 $\pm$ 0.04	34.6 $\pm$ 0.07	19.8 $\pm$ 0.04	1.60 $\pm$ 0.00
Tapioca	8.86 $\pm$ 0.00	19.3 $\pm$ 0.01	41.0 $\pm$ 0.07	23.1 $\pm$ 0.02	1.67 $\pm$ 0.00
Tapioca (pretreated)	7.73 $\pm$ 0.02	16.2 $\pm$ 0.03	31.7 $\pm$ 0.04	18.7 $\pm$ 0.03	1.48 $\pm$ 0.00
Wheat	14.1 $\pm$ 0.16	36.3 $\pm$ 0.34	83.4 $\pm$ 0.56	44.8 $\pm$ 0.36	1.91 $\pm$ 0.01
Wheat (pretreated)	11.4 $\pm$ 0.02	25.3 $\pm$ 0.05	52.5 $\pm$ 0.06	29.9 $\pm$ 0.04	1.63 $\pm$ 0.00
Waxy corn	14.9 $\pm$ 0.01	33.7 $\pm$ 0.04	69.9 $\pm$ 0.09	39.7 $\pm$ 0.04	1.63 $\pm$ 0.00
Waxy corn (pretreated)	8.98 $\pm$ 0.02	18.4 $\pm$ 0.05	36.0 $\pm$ 0.07	21.2 $\pm$ 0.04	1.47 $\pm$ 0.00

정 수준(약 50%)의 수분함량으로 전처리한 전분 시료에서 판별 정확도가 향상되는 결과는, 미세구조의 균질화에 기인한 데이터 재현성 및 특성 추출의 안정성 증가와 밀접한 관련이 있다고 판단된다. 즉, 입도 및 수분함량 조절에 따른 전분의 미세구조적 차이는 초분광 특성을 조금 더 부각시켜 머신러닝 모델 학습 시 전분 판별 성능을 높이는데 기여한 것으로 판단된다.

요약

본 연구에서는 초분광 및 머신러닝 분석 기술을 활용하여 생전분의 판별 가능성을 평가하고, 입도 및 수분함량 전처리를 통해 판별 성능을 개선하고자 하였다. 초분광 스펙트럼 분석 결과, 생전분 종류 간 스펙트럼이 상당 부분 중첩되어 단순 스펙트럼 기반 판별에는 한계가 있는 것으로 나타났다. 전분의 입도 변화에 따라 일부 스펙트럼 분리 효과가 확인되었으며, 수분함량 조절 시 스펙트럼 변별성에 큰 영향을 미치는 것으로 나타났다. 그 결과, 입도 100 mesh 및 수분함량 50% 조건에서 가장 높은 판별 정확도(약 97%)를 나타내었다. 이러한 결과는 입도 및 수분함량 전처리가 전분 입자의 응집을 완화하고 분산성을 향상시켜 보다 균일한 미세구조를 형성하기 때문으로 판단되며, SEM 및 입도분석을 통해 이를 확인하였다. 본 연구에서는 입도 및 수분함량을 제어한 전처리 전략을 통해 초분광 기반 생전분 판별 성능을 효과적으로 향상시킬 수 있음을 확인하였으며, 이와 같은 시도는 식품 산업에서 원료 품질 평가 및 공정 제어를 위한 신속하고 비파괴적인 분석 기술로 활용될 수 있을 것으로 기대된다.

ORCID

Da Eun Lee <https://orcid.org/0009-0003-3612-7062>
 Gyeong A Jeong <https://orcid.org/0000-0003-0391-1427>
 Chang Joo Lee <https://orcid.org/0000-0003-2212-252X>
 Suyong Lee <https://orcid.org/0000-0003-1284-211X>

Conflict of interests

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

Acknowledgements

This work was supported by the Rural Development Administration, Republic of Korea (RS-2025-02213186).

Data availability

Upon reasonable request, the datasets of this study can be available from the corresponding author.

Authorship contribution statement

Conceptualization: Lee S.
 Data curation: Lee DE, Lee S.
 Formal analysis: Lee DE, Jeong GA.
 Methodology: Lee DE, Lee S.
 Software: Lee DE, Lee S.
 Validation: Lee DE, Lee CJ, Lee S.
 Investigation: Lee DE, Lee CJ, Lee S.
 Writing - original draft: Lee DE, Lee CJ, Lee S.
 Writing - review & editing: Lee DE, Jeong GA, Lee CJ, Lee S.

Ethics approval

Not applicable.

References

- Amanah HZ, Wakholi C, Perez M, Faqeerzada MA, Tunny SS, Masithoh RE, Choung MG, Kim KH, Lee WH, Cho BK. 2021. Near-infrared hyperspectral imaging (NIR-HSI) for nondestructive prediction of anthocyanins content in black rice seeds. *Appl. Sci.* 11: 4841.
- Bogomolov A, Mannhardt J, Heinzerling O. 2018. Accuracy improvement of in-line near-infrared spectroscopic moisture monitoring in a fluidized bed drying process. *Front. Chem.* 6: 1-10.
- Breiman L. 2001. Random forests. *Mach. Learn.* 45: 5-32.
- Cornejo-Ramírez YI, Martínez-Cruz O, Del Toro-Sánchez CL, Wong-Corral FJ, Borboa-Flores J, Cinco-Moroyoqui FJ. 2018. The structural characteristics of starches and their functional properties. *CyTA J. Food.* 16: 1003-1017.
- Cortes C, Vapnik V. 1995. Support-vector networks. *Mach. Learn.* 20: 273-297.
- Cover T, Hart P. 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory.* 13: 21-27.
- Fu X, Chen J, Zhang J, Fu F, Wu C. 2021. Effect of penetration depth and particle size on detection of wheat flour adulterant using hyperspectral imaging. *Biosyst. Eng.* 204: 64-78.
- Govindaraju I, Chakraborty I, Baruah VJ, Sarmah B, Mahato KK, Mazumder N. 2021. Structure and morphological properties of starch macromolecule using biophysical techniques. *Starch-Stärke.* 73: 2000030.
- Haghighi M, Varmazyar M, Karevan Z, Ghasemigol M. 2018. PyCM: multiclass confusion matrix library in Python. *J. Open*

- Source Softw. 3: 729.
- Hwang JH, Jeong S, Lee S. 2025. Machine learning-combined hyperspectral imaging analysis for the non-destructive identification of wheat flours with varying gluten strengths. *Spectrochim. Acta A*. 338: 126118.
- Liu Y, Pu H, Sun DW. 2017. Hyperspectral imaging technique for evaluating food quality and safety during various processes: a review of recent applications. *Trends Food Sci. Technol.* 69: 25-35.
- Lu X, Sun J, Mao H, Wu X, Gao H. 2017. Quantitative determination of rice starch based on hyperspectral imaging technology. *Int. J. Food Prop.* 20: S1037-S1044.
- Mahmood Z, Jamel L, Salem DA, Ashraf I. 2025. Improving learning from the complex multi-class imbalanced and overlapped data by mapping into higher dimension using SVM++. *Sci Rep.* 15: 31245.
- Pasikatan MC, Steele JL, Spillman CK, Haque E. 2001. Near infrared reflectance spectroscopy for online particle size analysis of powders and ground materials. *J. Near Infrared Spectrosc.* 9: 153-164.
- Qiao M, Cui T, Xia G, Xu Y, Li Y, Fan C, Han S, Dong J. 2024. Integration of spectral and image features of hyperspectral imaging for quantitative determination of protein and starch contents in maize kernels. *Comput. Electron. Agric.* 218: 108718.
- Reeves JB. 1993. Influence of water on the near infrared spectra of model compounds. *J. AOAC Int.* 76: 741-748.
- Schwanninger M, Rodrigues JC, Fackler K. 2011. A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectrosc.* 19: 287-308.
- Shu Z, Li X, Liu Y. 2023. Detection of chili foreign objects using hyperspectral imaging combined with chemometric and target detection algorithms. *Foods.* 12: 2618.
- Song YY, Lu Y. 2015. Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiatry.* 27: 130-135.
- Tunny SS, Kurniawan H, Amanah HZ, Baek I, Kim MS, Chan D, Faqeerzada MA, Wakholi C, Cho BK. 2023. Hyperspectral imaging techniques for detection of foreign materials from fresh-cut vegetables. *Postharvest Biol. Technol.* 201: 112373.
- Wang F, Wang C, Song S, Xie S, Kang F. 2021. Study on starch content detection and visualization of potato based on hyperspectral imaging. *Food Sci. Nutr.* 9: 4420-4430.
- Wesley IJ, Blakeney AB. 2001. Investigation of starch-protein-water mixtures using dynamic near infrared spectroscopy. *J. Near Infrared Spectrosc.* 9: 211-220.
- Yoon J. 2022. Hyperspectral imaging for clinical applications. *BioChip J.* 16: 1-12.