Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable hyperspectral authentication of Pinellia ternata against Pinellia pedatisecta across tuber, slice, and powder forms with GC–TOF–MS corroboration

Yea-Jin Park
Kyung Hee University
Yu-Jin Jeon
Kyung Hee University
So Jin Park
Kyung Hee University
Da Hye Ryu
Korea Institute of Science and Technology (KIST)
Hye Min Kim
Korea Institute of Science and Technology (KIST)
Ho-Youn Kim
Korea Institute of Science and Technology (KIST)
Soo Hyun Park
Korea Institute of Science and Technology (KIST)
Dae-Hyun Jung
daehyun@khu.ac.kr

Kyung Hee University

Research Article

Keywords: Hyperspectral imaging, Deep learning, Explainable artificial intelligence, Multi-form authentication, Chemical corroboration, Chinese medicinal materials, Pinellia ternata

Posted Date: August 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10493138/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Explainable hyperspectral authentication of *Pinellia ternata* against *Pinellia pedatisecta* across tuber, slice, and powder forms with GC–TOF–MS corroboration

Yea-Jin Park[a,b], Yu-Jin Jeon[a,b], So Jin Park[a,b], Da Hye Ryu[c,d], Hye Min Kim[c], Ho-Youn Kim[c], Soo Hyun Park[c], Dae-Hyun Jung[a,b*]

[a] Department of Smart Farm Science, Kyung Hee University, Yongin 17104, Republic of Korea

[b] Interdisciplinary Program in IT-Bio Convergence System, Kyung Hee University, Yongin 17104, Republic of Korea

[c] Smart Farm Research Center, Korea Institute of Science and Technology (KIST), Gangneung-si 25451, Republic of Korea

[d] Department of Plant Science, The Pennsylvania State University, University Park, PA 16802, USA

*** Corresponding author: Dae-Hyun Jung(daehyun@khu.ac.kr)**

Abstract

Background *Pinellia ternata* is the botanical source of *Pinelliae Rhizoma* (Banxia), a medicinal material widely used in Chinese medicine. The morphologically similar adulterant *P. pedatisecta* reportedly contains higher levels of toxic lectins and differs in therapeutic indications and processing requirements, raising safety and efficacy concerns. We developed an explainable hyperspectral framework for nondestructive authentication across physical forms.

Methods Visible and near-infrared (VNIR; 400–1000 nm) and short-wave infrared (SWIR; 900–1700 nm) hyperspectral images were acquired from both species in tuber, slice, and powder forms. HyGD-Net integrates Savitzky–Golay-initialized learnable derivative filters, three convolutional branches, and position-wise gated fusion. It was compared with six classifiers across seven multi-form scenarios. Integrated Gradients identified recurrent wavebands, and GC–TOF–MS profiling assessed their chemical plausibility.

Results HyGD-Net achieved VNIR accuracies of 91.55–100.00%, with a mean of 94.46% and the lowest cross-scenario variability. In SWIR, HyGD-Net and ResNet achieved 100.00% accuracy in all scenarios, while other models also approached ceiling performance. Ablation analysis showed a mean VNIR gain of 6.30 percentage points from the triple-branch architecture and gains of 0.23–4.09 percentage points from gated fusion. Integrated Gradients identified recurrent regions within 400–500, 900–1000, and 1100–1200 nm. GC–TOF–MS revealed distinct volatile profiles and provided functional-group-level context for these regions.

Conclusions HyGD-Net enables nondestructive authentication across physical forms without manually selected derivative preprocessing. By combining learnable spectral transformation with chemically corroborated attribution, it provides accurate and explainable discrimination of *P. ternata* from *P. pedatisecta*. These findings demonstrate its potential as a practical quality-control tool for *Pinelliae Rhizoma* authentication.

Keywords: Hyperspectral imaging; Deep learning; Explainable artificial intelligence; Multi-form authentication; Chemical corroboration; Chinese medicinal materials; *Pinellia ternata*

1. Introduction

Plant-derived medicinal products are a core component of healthcare systems worldwide. Their widespread use has made authenticity increasingly important in post-harvest quality control. The global herbal medicine market grew to approximately USD 251.24 billion in 2025 [1]. During the same period, the WHO adopted the Traditional Medicine Strategy 2025–2034, reinforcing international attention to the safety and quality of herbal medicines [2]. In this context, reliable authentication of medicinal materials is not only a regulatory concern under frameworks such as the Nagoya Protocol on access and benefit-sharing [3] but also a practical challenge for quality control systems.

However, rising demand for herbal medicines has been accompanied by increased adulteration and misidentification. Reported adulteration rates in commercial products range from 27% to 41% [4, 5]. One representative example is *Pinellia ternata* (Thunb.) Breit, a herb widely used in East Asian traditional medicine, and its primary adulterant, *Pinellia pedatisecta* Schott [6]. The two species are morphologically similar; however, *P. pedatisecta* has a shorter cultivation period and a higher yield, making it substantially less expensive and a frequent adulterant. In the Chinese market, approximately 66% of *Pinellia* products have been reported to be adulterated [7], with *P. pedatisecta* alone accounting for 57% [8]. The risk extends beyond commercial fraud. *P. pedatisecta* contains higher levels of toxic lectin proteins and differs in its pharmacological indications and processing conditions[6, 9]. Such substitution may therefore reduce efficacy and cause adverse effects, underscoring the need for accurate species discrimination [8].

Conventional analytical methods have important limitations. Sensory evaluation is subjective and requires costly expert assessment [10]. Chromatographic techniques and DNA barcoding are accurate but time-consuming, expensive, and destructive because they require dedicated sample preparation [11, 12]. Rapid, nondestructive computer vision approaches have therefore been explored as alternatives. Sun and Qian [13] applied a convolutional neural network (CNN) and achieved 71% accuracy across 95 herbal

medicine categories. For *Ziziphus* species, Jeon et al. [14] achieved more than 98% accuracy using DenseNet.

However, RGB imaging captures only three visible bands and therefore relies primarily on surface-level color and morphological features [15]. This limitation reduces its ability to discriminate between similar species such as *P. ternata* and *P. pedatisecta*. This challenge increases when classification must generalize across physical forms. Herbal medicines are marketed not only as tubers but also as slices and powders [6], and processing progressively diminishes the morphological cues on which RGB-based models rely. RGB imaging alone is therefore insufficient for safety-critical authentication.

Hyperspectral imaging (HSI) addresses this limitation. By measuring reflectance across hundreds of narrow spectral bands, HSI captures a sample's intrinsic spectral fingerprint [16]. In the visible and near-infrared range (VNIR; 400–1000 nm), the signal reflects electronic transitions of chromophores and higher-order overtones. The short-wave infrared range (SWIR; 900–1700 nm) provides complementary information, where second-order overtones of O–H, N–H, and C–H bonds reflect compositional differences associated with moisture, lipids, and proteins [17, 18]. Combined with deep learning, HSI offers a promising nondestructive approach to automated authentication. HSI captures chemically discriminative information beyond the capabilities of RGB imaging and has demonstrated effectiveness in medicinal plant classification and origin discrimination [19, 20].

Current HSI–deep learning workflows have three recurring limitations that weaken their reliability. The first is dependence on preprocessing choices. Most models rely on a specific preprocessing method, and performance can vary considerably depending on that choice [21]. Because the transformation is fixed before training, it cannot be jointly optimized with the model. Identifying the optimal preprocessing combination also requires substantial time and domain expertise [22]. Some studies have sought to reduce this burden by feeding raw spectra directly into the model [23] or by automatically searching for Savitzky–Golay (SG) filter parameters [24]. Neither approach fully resolves this issue. The former does

not exploit complementary information from multiple spectral representations, whereas the latter still treats preprocessing as a separate stage from model optimization.

The second limitation is single-form validation. Most HSI–deep learning studies evaluate classification performance on only one physical form [19, 25], even though medicinal plant materials are distributed in multiple processed forms. To the best of our knowledge, no study has systematically evaluated physical form as an experimental factor while holding the target species pair constant and varying the form across in-form, mixed-form, and all-form conditions. Existing studies instead classify species within a single form, leaving robustness to form variation unmeasured. A unified model and multi-form evaluation protocol are therefore needed to assess practical robustness.

The third and most consequential limitation concerns the interpretability of model decisions. For quality control, accurate classification alone is insufficient; a model should also provide a transparent and scientifically defensible basis for its decisions. Explainable artificial intelligence (XAI) can address this need by identifying the spectral regions that drive classification [26]. However, in agricultural spectroscopy, attribution analyses typically end with band identification, and the resulting wavebands are seldom evaluated against independent analytical evidence [27]. Without chemical cross-validation, model explanations provide limited assurance that predictions rely on genuine compositional differences rather than spurious correlations. Addressing this gap requires integrating XAI-based attribution with independent chemical profiling.

To address these three limitations, this study develops HyGD-Net, a framework that integrates spectral preprocessing, classification, and chemically grounded interpretation for quality control. The study makes three contributions. (1) We internalize derivative preprocessing through learnable filters within an end-to-end network, enabling joint optimization of spectral transformation and classification without manual preprocessing selection. (2) We construct a multi-form evaluation protocol that holds the target species pair constant while varying physical form across seven in-form, mixed-form, and all-form scenarios, allowing direct measurement of robustness to form variation. (3) We cross-validate the model's

explanations against independent chemical evidence. By comparing Integrated Gradients consensus wavebands with gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) profiles, we connect prediction, explanation, and chemical evidence within a unified workflow.

2. Materials and methods

This study followed a four-stage workflow (Fig. 1). First, samples were prepared in multiple physical forms, and hyperspectral data were acquired. Next, the HyGD-Net architecture and benchmark classifiers were implemented. Discriminative wavebands were then identified using Integrated Gradients (IG) attribution analysis. Finally, these wavebands were chemically cross-validated using gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS).

2.1. Sample preparation

This study used dried tubers of *P. ternata* and *P. pedatisecta* obtained from the National Institute for Korean Medicine Development (NIKOM, Gyeongsan, Republic of Korea). A total of 320 *P. ternata* and 348 *P. pedatisecta* tubers were collected. Sample authenticity was confirmed using two approaches. First, nine qualified experts appointed by the Ministry of Food and Drug Safety conducted sensory evaluations. Second, representative samples from each species were analyzed using GC-TOF-MS.

Each tuber was subsequently processed into three physical forms: tuber (T), slice (S), and powder (P). Intact tubers were imaged before processing (320 *P. ternata* and 348 *P. pedatisecta* tubers). The tubers were then sliced transversely to a thickness of approximately 3.0 ± 0.5 mm (272 *P. ternata* and 288 *P. pedatisecta* samples), with small tubers excluded. The slices were subsequently ground into powder, yielding 300 g of powder from *P. ternata* and 360 g from *P. pedatisecta*. Because commercial powder products typically contain material pooled from multiple individual tubers, ground material from multiple samples was pooled to replicate this condition. All samples were stored in sealed containers under dry conditions at room temperature until analysis.

2.2. Hyperspectral image acquisition and preprocessing

2.2.1. Hyperspectral image acquisition

Hyperspectral images were acquired using two push-broom cameras (NIREOS S.P.L., Milan, Italy), each covering a different wavelength range. Imaging was performed inside an imaging station shielded

from external light by blackout curtains (Fig. 2A). The HERA VNIR camera covered the visible and near-infrared range from 400 to 1000 nm, with 120 spectral bands and a spatial resolution of 1280 × 1024 pixels. The HERA SWIR camera covered the short-wave infrared range from 900 to 1700 nm, with 140 spectral bands and a spatial resolution of 640 × 512 pixels. The VNIR camera had spectral resolutions of <1.5 nm at 400 nm and <10 nm at 1000 nm, whereas the SWIR camera had spectral resolutions of <5 nm at 900 nm and <16 nm at 1700 nm.

Illumination was provided by two halogen lamps and two light-emitting diode (LED) sources. Images were acquired at a camera-to-sample distance of 85 cm against a black background. Tuber and slice samples were arranged in a regular grid, with six samples per frame. For powder samples, 30-g portions were dispensed into Petri dishes, and two dishes were imaged per frame (Fig. 2B). All images were acquired and stored using HERA Acquisition software.

2.2.2. Reflectance calibration and spectral extraction

Because hyperspectral images are sensitive to sensor noise and illumination variability, the raw data were calibrated using Eq. (1):

$$R = (I_{raw} - I_{dark}) / (I_{\text{white}} - I_{dark}) \quad (1)$$

where R is the calibrated reflectance, I_{raw} is the raw signal, I_{dark} is the dark reference acquired with the camera lens covered by a cap, and I_{white} is the white reference acquired from a Spectralon panel (Labsphere Inc., North Sutton, NH, USA) with approximately 99% reflectance. After calibration, pseudo-RGB images were generated using selected wavelengths (VNIR, 450/550/650 nm; SWIR, 1200/1600/1700 nm) for region-of-interest (ROI) extraction (Fig. 2B).

For tuber and slice samples, automatic segmentation was performed using the Segment Anything Model (SAM), and background pixels were masked to zero [28]. Each ROI was divided into four quadrants. Pixel spectra within each quadrant were averaged to produce one spectral vector, yielding four spectra per sample. When an ROI was too small for quadrant-based extraction, a single spatially averaged

spectrum was used instead. For powder samples, a 10 × 10-pixel region centered within each Petri dish was manually defined, yielding 100 pixel-level spectra. Finally, mean spectral profiles were computed for each species and physical form across the VNIR (400–1000 nm) and SWIR (900–1700 nm) ranges.

2.3. Dataset construction and partitioning

Hyperspectral datasets were constructed for the three physical forms (tuber, slice, and powder) of *P. ternata* and *P. pedatisecta*. Table 1 summarizes the resulting datasets.

Table 1. Number of spectra per species, physical form, and imaging modality in the final dataset.

Modality	Species	Tuber	Slice	Powder
VNIR	*P. ternata*	1280	1080	1000
	P. pedatisecta	1308	1142	1200
SWIR	*P. ternata*	1280	1088	1000
	P. pedatisecta	1308	1148	1200

Note. VNIR = visible and near-infrared; SWIR = short-wave infrared.

To reduce the risk of data leakage, all ROIs derived from a single image were assigned to the same partition, ensuring that spectra from the same acquisition were not split across the training, validation, and test sets. Each dataset was then partitioned into training, validation, and test sets in a ratio of 64:16:20. No data augmentation was applied.

2.4. Experimental design

Seven experimental scenarios were designed and grouped into three categories (Table 2). The first category comprised in-form experiments (Exp. 1–3), in which classification was performed within a single physical form: tuber, slice, or powder. The second category comprised mixed-form experiments (Exp. 4–6), in which two physical forms were combined (tuber + slice, tuber + powder, or slice + powder). The third category comprised the all-form experiment (Exp. 7), in which six-class classification

was performed across all three physical forms simultaneously, simulating conditions in which diverse physical forms coexist.

Table 2. Experimental scenarios designed to evaluate classification performance in in-form, mixed-form, and all-form settings.

Category	Experiment	Tuber	Slice	Powder	Number of classes
In-form experiment	Exp. 1	√			2
	Exp. 2		√		2
	Exp. 3			√	2
Mixed-form experiment	Exp. 4	√	√		4
	Exp. 5	√		√	4
	Exp. 6		√	√	4
All-form experiment	Exp. 7	√	√	√	6

Note. Exp. = experiment; √ = physical form included in the scenario.

2.5. Classification models

The proposed HyGD-Net was evaluated against six benchmark classifiers comprising two conventional machine learning models, a support vector machine (SVM) and random forest (RF), and four deep learning models, a CNN, EfficientNetV2-S, ResNet, and Swin Transformer. For a fair comparison, all models received the same one-dimensional spectral inputs and used identical data partitions. Each model was trained to achieve its best validation performance. The deep learning models

used validation-based early stopping with restoration of the best-performing weights, and tuning was conducted without preferential treatment of any architecture. The source code for the proposed model is publicly available at [anonymized repository].

2.5.1. Conventional machine learning models

A support vector machine (SVM) [29] and random forest (RF) [30] were used as benchmark models. For the SVM, the regularization parameter was set to C = 10 and γ = "scale", and the input spectra were standardized using StandardScaler. The RF model was configured with n_estimators = 200, max_depth = 20, and max_features = "sqrt", without input normalization. Both models received one-dimensional spectral vectors as input. The conventional machine learning models were implemented using scikit-learn in Python.

2.5.2. Deep learning models

Four deep learning architectures were evaluated: a CNN, EfficientNetV2-S, ResNet, and Swin Transformer. A custom one-dimensional CNN comprising three convolutional layers was used to extract spectral features [31]. EfficientNetV2-S was adapted for hyperspectral data by replacing two-dimensional convolutional operations with one-dimensional counterparts [32]. ResNet was implemented as a lightweight one-dimensional architecture inspired by ResNet-18 and comprised residual blocks with skip connections [33]. Swin Transformer was similarly adapted to one-dimensional inputs to model long-range spectral dependencies [34]. Because suitable pretrained weights were unavailable for the hyperspectral inputs used in this study, all models were trained from scratch with random initialization.

2.6. Proposed HyGD-Net architecture

HyGD-Net (Hyperspectral Gated Derivative Network) was designed to enable joint optimization of spectral preprocessing and classification within a single end-to-end framework. In conventional workflows, preprocessing operations such as derivative transformation and standard normal variate (SNV) transformation are fixed before model training and therefore cannot be optimized jointly with model

parameters. To address this limitation, HyGD-Net internalizes derivative transformations as learnable neural network components rather than external preprocessing operations, allowing them to be adaptively optimized for the classification objective. The overall architecture comprises three core modules: (1) a learnable filter module, (2) a triple-branch parallel CNN, and (3) a position-wise gated fusion module (Fig. 3). Each module is described below.

2.6.1. Learnable filter module

The first module implements a Savitzky–Golay (SG) derivative filter [35] as a learnable convolutional layer, enabling adaptive optimization of spectral preprocessing. The SG filter fits a polynomial of order p to 2m+1 consecutive spectral points and computes the dth-order derivative at the center point. For an input spectrum x = [x1, x2, …, xN], the dth-order SG derivative at position i can be expressed as shown in Eq. (2):

$$\hat{x}_i^{(d)} = \sum_{j=-m}^{m} c_j^{(d)} x_{i+j} \quad (2)$$

where $c_j^{(d)}$ denotes the SG coefficient corresponding to derivative order d.

Because this operation is mathematically equivalent to one-dimensional convolution with the SG coefficients as the convolution kernel, the filter can be reformulated as a Conv1D layer. In this study, a minimal three-point window was used (m = 1, such that 2m+1 = 3) with polynomial order p = 2. Under these settings, the SG coefficients reduce to the classical central-difference and second-difference operators, providing a theoretically grounded initialization. Kernel weights were initialized with these coefficients and subsequently optimized through backpropagation during training.

Given an input spectrum x, this module simultaneously generates three complementary representations, as defined in Eq. (3):

$$x^{(0)} = x,\ x^{(1)} = h_1 * x,\ x^{(2)} = h_2 * x \quad (3)$$

where $*$ denotes one-dimensional convolution. All outputs retain the same spectral length through "same" padding. The raw spectral branch passes the input spectrum directly, whereas the first- and

second-order derivative branches apply learnable filters optimized through backpropagation to generate task-adaptive representations.

The derivative filters comprise two bias-free Conv1D layers corresponding to the first- and second-order derivatives, each with a kernel size of 3 (2m+1) and "same" padding. The weights are initialized using the SG coefficients in Eqs. (4) and (5) and are subsequently updated during training to optimize task-adaptive spectral representations:

$$x^{(1)} = h_1 * x,\ h_1 = [-0.5,\ 0,\ 0.5] \tag{4}$$

$$x^{(2)} = h_2 * x,\ h_2 = [1,\ -2,\ 1] \tag{5}$$

2.6.2. Triple-branch convolutional architecture

To extract complementary features from the raw, first-order derivative, and second-order derivative spectra, three parallel convolutional branches were employed. The branches shared the same architecture but had independent parameters, enabling representation-specific feature learning. Each branch comprised three sequential convolutional blocks with increasing numbers of filters (32, 64, and 128). Each block consisted of a Conv1D layer with kernel size k = 5, followed by rectified linear unit (ReLU) activation, batch normalization, and max pooling with a pool size of 2, as shown in Eq. (6):

$$F_i^{(l)} = \mathrm{MaxPool}\big(\mathrm{BN}\big(\mathrm{ReLU}\big(\mathrm{Conv1D}\big(F_i^{(l-1)}\big)\big)\big)\big) \tag{6}$$

Given an input spectrum with N spectral bands, the spectral dimension was progressively reduced according to Eq. (7):

$$N \rightarrow \lfloor N/2 \rfloor \rightarrow \lfloor N/4 \rfloor \rightarrow \lfloor N/8 \rfloor \tag{7}$$

where N = 120 for VNIR and N = 140 for SWIR. The final outputs of the three branches are denoted F_{raw}, F_{d1}, and F_{d2}, respectively.

2.6.3. Position-wise gated fusion module

HyGD-Net employs a position-wise gated fusion mechanism that adaptively modulates the contribution of each branch at each spectral position rather than simply concatenating the branch outputs

[36]. This mechanism allows the model to selectively emphasize informative representations among the raw, first-order derivative, and second-order derivative features.

The outputs of the three branches are projected to a common feature dimension D = 64 using a pointwise convolution (k = 1), as defined in Eq. (8):

$$f_i = \mathrm{Conv1D}_{k=1}(F_i),\ i \in \{\mathrm{raw},\ \mathrm{d1},\ \mathrm{d2}\} \tag{8}$$

The projected features are concatenated along the channel dimension to form a global context representation, as shown in Eq. (9):

$$H = [f_{\mathrm{raw}} \,\|\, f\mathrm{d1} \,\|\, f\mathrm{d2}] \tag{9}$$

where ‖ denotes channel-wise concatenation. Branch-specific gates are then generated from H according to Eq. (10):

$$g_i = \sigma(\mathrm{Conv1D}_{k=1}^{(i)}(H)),\ i \in \{\mathrm{raw},\ \mathrm{d1},\ \mathrm{d2}\} \tag{10}$$

where σ denotes the sigmoid function. The fused representation is computed as the element-wise weighted sum of the branch features, as given in Eq. (11):

$$F_{fused} = \Sigma_{i \in \{\mathrm{raw,d1,d2}\}}\, g_i \odot f_i \tag{11}$$

where ⊙ denotes element-wise multiplication. Finally, the fused features are passed through a pointwise convolution (k = 1) with Gaussian error linear unit (GELU) activation, followed by layer normalization, as shown in Eq. (12):

$$F_{out} = \mathrm{LayerNorm}[\mathrm{GELU}(\mathrm{Conv1D}_{k=1}(Ffused))] \tag{12}$$

2.6.4. Classification head

Global average pooling (GAP) was applied to the output of the gated fusion module to obtain a fixed-length feature vector. This operation was followed by a fully connected layer with 256 hidden units and GELU activation, a dropout layer (p = 0.5), and a softmax output layer that produced a probability distribution over C classes. The value of C was set to 2, 4, or 6 depending on the experimental scenario.

2.7. Ablation study design

To quantify the contribution of each HyGD-Net component, four architectural configurations were evaluated, comprising the full model and three reduced variants (Table 3).

Table 3. Ablation study configurations used to evaluate the contribution of each HyGD-Net module, including the Savitzky–Golay (SG) filter.

Variant	Learnable SG filter	triple branch	Gated fusion	Description
Full	√	√	√	Complete HyGD-Net
Fixed-SG	Frozen	√	√	SG weights fixed at Eqs. (4) and (5)
No-Gate	√	√	Concatenation	Gated fusion replaced by channel concatenation
Single-Branch	√	Single branch	None	Single-branch architecture only

Note. SG = Savitzky–Golay; √ = component included.

In the Fixed-SG variant, the SG filter weights were frozen during training, enabling assessment of the contribution of learnable preprocessing. In the No-Gate variant, features from all branches were concatenated without adaptive weighting, enabling evaluation of the gated fusion mechanism. In the Single-Branch variant, all spectral representations were processed by a single network, removing both branch separation and gated fusion.

2.8. Training configuration

All experiments were conducted on a workstation equipped with an NVIDIA RTX A5000 GPU with 24 GB of video random-access memory (VRAM). Deep learning models were implemented in TensorFlow 2.20.0 with Python 3.10, whereas the conventional machine learning models, SVM and RF, were implemented using scikit-learn. Image processing and ROI extraction used OpenCV, rembg (U^2-Net), and the Segment Anything Model (SAM), whereas reflectance calibration and spectral extraction used NumPy and SciPy. For each deep learning model, training was terminated when validation loss ceased to improve according to the specified early-stopping criterion, and the weights from the best validation epoch were restored. All models were trained with a single fixed random seed, and the hyperparameters used to train HyGD-Net are listed in Table 4.

Table 4. Hyperparameter settings used for HyGD-Net training.

Parameter	Value
Optimizer	Adam ($\beta1 = 0.9$, $\beta2 = 0.999$)
Initial learning rate	3×10^{-4}
Learning-rate scheduler	ReduceLROnPlateau (patience = 10, factor = 0.5, min_lr = 1×10^{-6}
Batch size	32
Maximum epochs	200
Early stopping	Patience = 20, monitor = val_loss
Loss function	Cross-entropy
Dropout rate	0.5
Random seed	42

Note. min_lr = minimum learning rate; val_loss = validation loss.

2.9. Performance evaluation metrics

Classification performance was evaluated using accuracy (ACC), sensitivity (SEN), and specificity (SPE). Because ACC was computed on the native multiclass task in each experimental scenario, it was defined as the proportion of correctly classified samples, as defined in Eq. (13):

$$\mathrm{ACC} = N_{\mathrm{correct}} / N_{\mathrm{total}} \tag{13}$$

where $N_{correct}$ and N_{total} denote the numbers of correctly classified and total samples, respectively. SEN and SPE were computed after collapsing predictions into a binary species-level decision and were defined in Eqs. (14) and (15) as follows:

$$\mathrm{SEN} = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN}) \tag{14}$$

$$\mathrm{SPE} = \mathrm{TN} / (\mathrm{TN} + \mathrm{FP}) \tag{15}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative predictions, respectively, with *P. ternata* treated as the positive class.

The metrics were therefore computed at two different levels. ACC was measured on the native 2-, 4-, or 6-class task for each scenario and thus reflected multiclass classification performance. By contrast, SEN and SPE were computed after collapsing predictions into a binary species-level decision (*P. ternata* versus *P. pedatisecta*) and therefore reflected species discrimination alone.

This distinction is important when interpreting the results. For example, an intraspecies cross-form error, such as classifying a *P. ternata* tuber as a *P. ternata* slice, reduces ACC because the predicted class differs from the true class. However, because the predicted and true labels correspond to the same species, the error does not affect SEN or SPE. Consequently, a model that discriminates species accurately but confuses physical forms may exhibit high SEN and SPE despite comparatively lower ACC.

2.10. Model interpretation using explainable artificial intelligence

Integrated Gradients (IG) was used to quantify waveband-level importance in the deep learning models (Sundararajan et al., 2017). IG attributes a prediction to individual input features by integrating

model-output gradients along a straight-line path from a baseline input x′ to the observed input x, as defined in Eq. (16):

$$IG_i(x) = (x_i - x'_i) \int_0^1 [\partial F(x' + \alpha(x - x')) / \partial x_i] \, d\alpha \quad (16)$$

where x is the input spectrum, x′ is the baseline spectrum, and F(x) is the model output for the target class. A zero-valued spectrum was used as the baseline. The integral was approximated using 200 interpolation steps and trapezoidal integration. IG attributions were computed for five deep learning models across seven experimental scenarios and two spectral regions, yielding 70 conditions in total. For each condition, absolute attribution values were averaged across test samples to obtain a waveband-level importance profile. The corresponding 95% confidence intervals were estimated using 1,000 bootstrap iterations.

A two-stage consensus procedure was then applied to identify robust discriminative wavelengths. Both stages used the same greedy neighbor-merging rule with a ±10-nm tolerance but differed in their inputs. Stage 1 operated within each experiment. The top 10 wavelengths from all models were merged, and a cluster was retained only when at least two models contributed to it, yielding model-level consensus. Stage 2 operated across experiments. Cluster centers retained after Stage 1 were merged again, and a cluster was retained only when it occurred in at least three experiments, yielding global consensus wavelengths. Separately, the selection frequency of a wavelength was defined as the number of the seven experiments in which that wavelength ranked among the top 10 IG bands.

2.11. GC-TOF-MS chemical analysis

Gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) was performed using a Pegasus GC HRT system (LECO Corp., USA) to provide independent chemical evidence for the XAI-derived consensus wavebands. Dried samples (200 mg) were mixed with 2 mL of 30% NaCl solution, and 2 μL of 0.2 mg/mL 3-pentanol was added as an internal standard. Volatile organic compounds (VOCs) were extracted by headspace solid-phase microextraction (HS-SPME) using a 50/30-μm DVB/CAR/PDMS

fiber (Supelco, USA) at 70 °C for 20 min with agitation at 500 rpm, followed by thermal desorption in the GC injector for 3 min.

Separation was performed on an Rtx-5MS column (30 m × 0.25 mm × 0.25 μm; Restek, USA) in split injection mode (40:1), with helium as the carrier gas at a flow rate of 1.0 mL/min. The oven temperature was increased stepwise from 45 °C to 220 °C and then held at 220 °C for 3 min. The ion source was maintained at 250 °C with electron ionization at 70 eV and mass spectra were acquired in scan mode over m/z 35–450. VOCs were identified by comparison with the NIST mass spectral library, with 3-pentanol serving as the internal standard. Detection was evaluated qualitatively based on the presence or absence of each compound in the two species rather than quantitative peak-area comparisons.

3. Results

3.1. Spectral characteristics across physical forms

The two species showed greater spectral separation in the SWIR range than in the VNIR range. Their mean reflectance spectra across the three physical forms are compared in Fig. 4.

In the VNIR region (Fig. 4A–C), the two species exhibited similar spectral patterns. Reflectance was low below 500 nm, increased gradually between 500 and 700 nm, and stabilized above 750 nm. The spectral profiles largely overlapped, with limited interspecies differences observed primarily within the 500–700 nm range. Intersample variability was greater for tuber and slice samples than for powder samples.

By contrast, the SWIR region (Fig. 4D–F) exhibited more complex spectral profiles with multiple absorption features. Spectral differences between the two species were more pronounced than those in the VNIR region, particularly within the 900–1300 nm and 1400–1600 nm intervals.

3.2. Classification performance of HyGD-Net across VNIR and SWIR spectral regions

The proposed model and six benchmark models were evaluated using raw spectra across seven experimental scenarios and two spectral regions, VNIR (400–1000 nm) and SWIR (900–1700 nm), covering in-form (Exp. 1–3), mixed-form (Exp. 4–6), and all-form (Exp. 7) settings. HyGD-Net achieved the highest or joint-highest accuracy in every scenario, with accuracy (ACC) ranging from 91.55% to 100.00% (Fig. 5). Its advantages over the benchmark models were particularly pronounced in several mixed-form settings, where variation in physical form increased classification complexity. Complete scenario-specific results for both spectral regions are presented in Tables 5 and 6.

Table 5. Scenario specific classification performance (accuracy, sensitivity, and specificity, %) of HyGD-Net and six benchmark models using raw visible and near-infrared (VNIR; 400–1000 nm) spectra across Experiments 1–7.

Model	Metric	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
HyGD-Net	Acc	91.64	93.71	100.00	91.55	95.09	96.50	92.76
	Sens	90.78	95.20	100.00	90.78	97.70	95.10	94.40
	Spec	92.58	92.13	100.00	92.37	92.11	98.08	90.92
CNN	Acc	87.92	88.99	98.41	86.56	91.10	93.22	92.12
	Sens	88.65	90.39	97.50	85.10	89.66	92.75	92.55
	Spec	87.11	87.50	99.50	88.14	92.76	93.75	98.41
ResNet	Acc	86.25	92.36	99.09	85.13	89.57	95.03	90.51
	Sens	86.23	92.35	99.17	84.90	95.79	94.24	92.00
	Spec	86.23	92.35	99.17	85.38	82.46	95.91	88.84
EfficientNetV2-S	Acc	58.36	78.65	98.41	85.34	83.84	89.38	89.87
	Sens	20.92	84.28	97.50	84.51	72.99	88.06	89.96
	Spec	99.61	72.69	99.50	86.23	96.27	90.87	97.96
Swin Transformer	Acc	78.44	77.53	96.59	80.75	75.05	84.97	86.78
	Sens	78.60	77.43	96.79	81.13	77.96	84.86	87.20
	Spec	78.60	77.43	96.79	93.56	91.56	94.97	97.34
SVM	Acc	82.53	80.00	99.32	77.70	82.92	85.54	78.20
	Sens	82.63	79.88	99.29	78.82	89.27	82.52	76.13
	Spec	82.63	79.88	99.29	76.48	75.66	88.94	80.51

Model	Metric	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
RF	Acc	86.06	75.73	98.18	81.67	90.59	84.63	84.74
	Sens	86.05	75.67	98.17	83.73	92.53	81.88	85.60
	Spec	86.05	75.67	98.17	79.45	88.38	87.74	83.78

Note. Acc = accuracy; Sens = sensitivity; Spec = specificity; Exp. = experiment.

Table 6. Scenario specific classification performance (accuracy, sensitivity, and specificity, %) of HyGD-Net and six benchmark models using raw short-wave infrared (SWIR; 900–1700 nm) spectra across Experiments 1–7.

Model	Metric	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
HyGD-Net	Acc	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Sens	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Spec	100.00	100.00	100.00	100.00	100.00	100.00	100.00
CNN	Acc	99.81	99.78	100.00	100.00	99.90	99.55	100.00
	Sens	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Spec	99.61	99.54	100.00	100.00	99.78	99.04	100.00
ResNet	Acc	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Sens	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Spec	100.00	100.00	100.00	100.00	100.00	100.00	100.00
EfficientNetV2-S	Acc	99.81	100.00	99.32	100.00	100.00	100.00	100.00
	Sens	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Model	Metric	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
Swin Transformer	Spec	99.61	100.00	98.50	100.00	100.00	100.00	100.00
	Acc	98.88	99.55	100.00	100.00	100.00	99.77	100.00
	Sens	98.86	99.57	100.00	100.00	100.00	99.77	100.00
	Spec	98.86	99.57	100.00	100.00	100.00	99.92	100.00
SVM	Acc	99.44	99.78	100.00	99.90	99.69	99.89	99.58
	Sens	99.41	99.77	100.00	100.00	100.00	100.00	99.73
	Spec	99.41	99.77	100.00	99.79	99.34	99.76	99.41
RF	Acc	93.87	97.10	100.00	93.91	97.44	98.87	95.30
	Sens	93.70	97.07	100.00	95.30	97.51	99.36	97.60
	Spec	93.70	97.07	100.00	92.41	97.37	98.33	92.73

Note. Acc = accuracy; Sens = sensitivity; Spec = specificity; Exp. = experiment.

3.2.1. Performance comparison in the VNIR region

Across all seven experimental scenarios using VNIR (400–1000 nm) spectra, HyGD-Net consistently achieved the highest accuracy (Fig. 5).

In the in-form scenarios (Exp. 1–3), HyGD-Net achieved accuracies of 91.64%, 93.71%, and 100.00%, respectively. Its advantage over the second-best-performing model was largest in Exp. 1, where it exceeded CNN by 3.72 percentage points (pp). By contrast, EfficientNetV2-S achieved only 58.36% accuracy in the same experiment. Its confusion matrix showed that 223 of 282 *P. ternata* samples were misclassified as *P. pedatisecta*, indicating a pronounced bias toward the *P. pedatisecta* class.

Across the mixed-form scenarios (Exp. 4–6), HyGD-Net maintained accuracies within a relatively narrow range of 91.55–96.50%. ResNet showed greater variation, achieving 92.36% accuracy in Exp. 2

but dropping to 85.13% in Exp. 4. In the all-form scenario (Exp. 7), HyGD-Net achieved 92.76% accuracy, 94.40% sensitivity, and 90.92% specificity. The conventional machine learning models performed less favorably, with SVM and RF achieving accuracies of 78.20% and 84.74%, respectively. The other deep learning benchmark models achieved accuracies ranging from 86.78% to 92.12%.

3.2.2. Performance comparison in the SWIR region

Overall classification performance was higher in the SWIR region (900–1700 nm) than in the VNIR region (Table 6). HyGD-Net and ResNet achieved 100.00% accuracy, sensitivity, and specificity across all seven scenarios. The remaining deep learning models also achieved near-saturated performance. EfficientNetV2-S and CNN maintained accuracies above 99% in every scenario, whereas Swin Transformer achieved 100.00% accuracy in four scenarios and a minimum accuracy of 98.88% in Exp. 1. SVM likewise maintained accuracies at or above 99.44% across all scenarios.

RF exhibited the greatest variability among the models. Under mixed-form conditions, its accuracy ranged from 93.91% to 98.87%, whereas its overall minimum was 93.87% in Exp. 1. This value was 6.13 pp below the 100.00% accuracy achieved by HyGD-Net and ResNet in the same experiment.

3.2.3. Accuracy variability across scenarios

Accuracy variability across the seven scenarios is summarized in the box plots in Fig. 5B and 5D, illustrating the sensitivity of each model to changes in physical-form composition.

In the VNIR range (Fig. 5B), HyGD-Net exhibited the lowest variability, with a mean accuracy of 94.46% and a standard deviation of 2.81 pp. CNN ranked second in stability, with a mean accuracy of 91.19% and a standard deviation of 3.67 pp, followed by ResNet with a standard deviation of 4.52 pp. The remaining models exhibited standard deviations exceeding 6 pp. EfficientNetV2-S showed the greatest variability, with accuracy ranging from 58.36% to 98.41%, a mean of 83.41%, and a standard deviation of 11.69 pp.

In the SWIR range (Fig. 5D), variability decreased markedly across models. HyGD-Net and ResNet showed no variability, maintaining 100.00% accuracy in every scenario. EfficientNetV2-S, CNN, and Swin Transformer each achieved mean accuracies above 99.7%, with standard deviations below 0.4 pp. SVM showed similarly low variability, with a mean accuracy of 99.75% and a standard deviation of 0.18 pp. RF was the exception, with accuracy ranging from 93.87% to 100.00%, a mean of 96.64%, and the largest standard deviation among the SWIR models, 2.21 pp.

3.3. Component-wise ablation study of HyGD-Net

The HyGD-Net ablation results are presented in Fig. 6. Following the design described in Section 2.7, three reduced variants, Fixed-SG, No-Gate, and Single-Branch, were compared with the full model. In the upper panel, circle size and accompanying percentage labels indicate the reduction in accuracy associated with removing or modifying the corresponding component. The lower bar plot shows the cumulative increase in accuracy from the baseline architecture to the full HyGD-Net as components are added sequentially.

3.3.1. Contribution of the learnable SG filter

The contribution of the learnable filter was assessed by comparing HyGD-Net with a variant using fixed SG coefficients across the seven scenarios. In the VNIR region, the Fixed-SG variant showed lower accuracy in five of the seven experiments (Fig. 6), with reductions ranging from 0.41 to 3.16 pp. The largest decrease occurred in Exp. 1 (3.16 pp), followed by Exp. 7 (1.76 pp), whereas smaller decreases were observed in Exp. 4–6. The two models achieved identical accuracies in Exp. 2 and Exp. 3. In the SWIR region, the two configurations performed identically, each achieving 100.00% accuracy across all scenarios.

3.3.2. Contribution of the triple-branch architecture

Classification performance was compared between the triple-branch CNN used in HyGD-Net and a single-branch variant across the seven scenarios.

In the VNIR region, the single-branch model achieved lower accuracy in every scenario, with a mean decrease of 6.30 pp. Differences ranged from 1.36 to 10.34 pp and were largest in Exp. 2 (Δ = 10.34 pp), Exp. 4 (Δ = 9.68 pp), Exp. 6 (Δ = 8.25 pp), and Exp. 7 (Δ = 7.60 pp). Smaller differences were observed in Exp. 1 (2.98 pp), Exp. 5 (3.99 pp), and Exp. 3 (1.36 pp); in Exp. 3, both models achieved accuracies above 98%. In the SWIR region, the two architectures produced nearly identical results. A difference of only 0.10 pp was observed in Exp. 5, whereas identical accuracies were obtained in the other six scenarios.

3.3.3. Contribution of gated fusion

Classification performance was compared between the full HyGD-Net model with gated fusion and a No-Gate variant using simple concatenation across the seven scenarios.

Across the VNIR experiments, the No-Gate variant consistently achieved lower accuracy, with reductions ranging from 0.23 to 4.09 pp. The largest decrease occurred in Exp. 1 (4.09 pp), followed by Exp. 5 (2.66 pp), Exp. 4 (1.43 pp), and Exp. 2 (1.35 pp). Smaller differences of 0.23–0.62 pp were observed in Exp. 3, Exp. 6, and Exp. 7, where overall accuracy was already high. In the SWIR region, both configurations achieved 100.00% accuracy in every experiment.

3.3.4. Learnable SG filter weight analysis

To determine whether the learnable SG filter adapted its coefficients during training, deviations of the learned weights from their initial Savitzky–Golay coefficients were analyzed (Table 7). D1 and D2 denote the first- and second-derivative branches, respectively.

Within the VNIR range, deviations remained consistently small across all experiments. The maximum absolute deviation (|Δ|) ranged from 0.0011 to 0.0019 for D1 and from 0.0006 to 0.0019 for D2, corresponding to less than 0.40% and 0.10%, respectively, relative to the maximum absolute magnitudes of the initial coefficients.

Slightly larger deviations were observed in the SWIR range. The maximum |Δ| ranged from 0.0005 to 0.0030 for D1 and from 0.0012 to 0.0049 for D2, corresponding to maximum relative changes of 0.60% and 0.25%, respectively. The largest D2 deviation occurred in Exp. 6, followed by Exp. 7 and Exp. 4.

Table 7. Maximum absolute deviations of learned filter coefficients from their initial Savitzky–Golay (SG) values for the D1 (first-derivative) and D2 (second-derivative) branches.

Region	Filter	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
VNIR	D1	0.0013	0.0012	0.0017	0.0011	0.0012	0.0017	0.0019
	D2	0.0015	0.0010	0.0006	0.0019	0.0010	0.0011	0.0017
SWIR	D1	0.0012	0.0030	0.0012	0.0005	0.0010	0.0008	0.0022
	D2	0.0029	0.0020	0.0014	0.0036	0.0012	0.0049	0.0042

Note. D1 = first-derivative branch; D2 = second-derivative branch; SG = Savitzky–Golay; Exp. = experiment.

3.3.5. Gated fusion weight analysis

The learned gate values of the gated fusion module were analyzed across wavelength positions for all seven experimental scenarios (Fig. 7). Values for the three branches, raw, first-order derivative, and second-order derivative, clustered near the sigmoid midpoint of 0.5, and none approached saturation near 0 or 1.

In the VNIR range, the mean gate values (mean ± standard deviation across wavelength position) were 0.4955 ± 0.0019 for the raw branch, 0.5010 ± 0.0014 for the first-order derivative branch, and 0.4984 ± 0.0018 for the second-order derivative branch. The first-order derivative branch had the highest gate value at 63.8% of wavelength positions. Interbranch differences were most pronounced within the 400–500 nm region and diminished above 500 nm as the gate values converged toward 0.5.

In the SWIR range, the mean gate values were 0.4925 ± 0.0050 for the raw branch, 0.5011 ± 0.0036 for the first-order derivative branch, and 0.4926 ± 0.0055 for the second-order derivative branch. The first-order derivative branch had the highest gate value at 79.8% of wavelength positions, whereas the raw branch exhibited greater variability. Interbranch differences were most pronounced within the 900–1200 nm range and diminished at longer wavelengths.

3.4. XAI-based wavelength importance analysis

Integrated Gradients (IG) attribution analysis was performed for five deep learning models across seven scenarios and two spectral regions, yielding 70 conditions in total (Fig. 8). Consensus wavelengths were subsequently identified from attribution patterns that recurred across models and experiments according to the procedure described in Section 2.10.

In the VNIR region, the 400–500 nm and 900–1000 nm intervals showed strong intermodel agreement across experiments. The 400–500 nm interval was more consistently represented, being selected by most models in nearly all experiments. Application of the ±10-nm tolerance criterion yielded eight global consensus wavelengths, with the most representative described below. A consensus wavelength at approximately 444 nm, associated with a merged cluster spanning 440–453 nm, recurred in five experiments, while a neighboring wavelength at approximately 461 nm showed similarly consistent recurrence. Within the 900–1000 nm interval, a consensus wavelength at approximately 997 nm also recurred in five experiments, with additional consensus wavelengths at approximately 981, 962, and 937 nm.

In the SWIR region, recurrent agreement was observed within the 900–1000 nm and 1100–1200 nm intervals across experiments. Within the 1100–1200 nm interval, approximately 1157 nm, associated with a cluster spanning 1150–1159 nm, was the most reproducible consensus wavelength and recurred in five experiments. Additional consensus wavelengths at approximately 1120 and 1171 nm formed a closely spaced group within the same broader region. Within the 900–1000 nm interval, approximately 991 nm

was the most representative consensus wavelength, while additional wavelengths at approximately 931, 1017, and 1088 nm extended across the broader 900–1100 nm range. Because the 900–1000 nm interval lies within the overlap between the VNIR and SWIR sensor ranges, consensus wavelengths were identified in this region in both analyses.

3.5. GC-TOF-MS analysis for chemical validation

GC-TOF-MS analysis was performed on *P. ternata* and *P. pedatisecta* to provide independent chemical evidence for the key attribution wavebands identified by IG analysis. Eleven volatile compounds exhibited distinct detection patterns between the two species. Nine compounds were identified by NIST library matching, whereas two remained unidentified and were reported by retention index (RI) (Table 8). The detection patterns comprised three groups.

Hexanal, heptanal, and benzaldehyde were detected in both species. Three compounds were detected only in *P. ternata*: 2-heptanone, phenylacetaldehyde, and one unidentified compound. Five compounds were detected only in *P. pedatisecta*: hexanoic acid, (E,E)-2,4-heptadienal, (E)-2-octenal, nonanal, and one unidentified compound. Thus, more species-specific compounds were detected in *P. pedatisecta* than in *P. ternata*. Aldehydes predominated among the identified compounds, accounting for seven compounds, whereas one ketone (2-heptanone) and one carboxylic acid (hexanoic acid) were also identified. The analysis was qualitative, and peak areas were not compared quantitatively. These species-specific volatile profiles were subsequently compared with the IG consensus wavelengths and near-infrared overtone absorption regions associated with the corresponding functional groups (Fig. 9).

Table 8. Volatile compounds identified in *P. ternata* and *P. pedatisecta* by gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) analysis.

No.	Compound	Rt (s)	RI	Similarity	*P. ternata*	*P. pedatisecta*
1	Hexanal	253.9	801.99	930	+	+

No.	Compound	Rt (s)	RI	Similarity	*P. ternata*	*P. pedatisecta*
2	2-Heptanone	378.0	893.65	901	+	−
3	Heptanal	393.7	903.65	916	+	+
4	Benzaldehyde	505.2	960.95	932	+	+
5	Hexanoic acid	541.0	979.34	737	−	+
6	(E,E)-2,4-Heptadienal	610.9	1012.40	856	−	+
7	Unknown	617.8	1015.30	—	+	−
8	Phenylacetaldehyde	688.6	1045.00	932	+	−
9	(E)-2-Octenal	721.3	1058.70	914	−	+
10	Unknown	809.1	1095.50	—	−	+
11	Nonanal	835.3	1106.00	909	−	+

Note. Rt = retention time; RI = retention index; + = detected; − = not detected; — = no similarity value provided.

4. Discussion

This study addressed three limitations that have constrained the application of hyperspectral imaging to herbal authentication: dependence on manually selected preprocessing, validation restricted to a single physical form, and a lack of independent chemical evidence supporting the wavebands that influence model decisions. Addressing these limitations within a unified workflow provides a more robust and verifiable framework for authentication.

4.1. Preprocessing-free and form-robust classification with HyGD-Net

Without external derivative preprocessing, HyGD-Net classified raw spectral inputs with accuracies of 91.55–100.00% in the VNIR range and 100.00% across all SWIR scenarios, achieving the highest or joint-highest performance under identical evaluation conditions. Conventional HSI pipelines often depend on predefined preprocessing methods, such as Savitzky–Golay derivatives, standard normal variate (SNV) transformation, and multiplicative scatter correction (MSC), each of which requires selection and tuning before model training. HyGD-Net instead embeds learnable derivative filters within the model, enabling joint optimization of spectral transformation and classification while exploiting complementary spectral representations.

Beyond reducing dependence on externally selected derivative preprocessing, HyGD-Net maintained high and stable performance across experimental scenarios. In the VNIR range, it achieved a mean accuracy of 94.46% with a standard deviation of 2.81 percentage points (pp), the lowest variability among all evaluated models. The benchmark models exhibited standard deviations ranging from 3.67 to 11.69 pp. ResNet, for example, varied by 13.96 pp between its highest and lowest accuracies, from 99.09% in Exp. 3 to 85.13% in Exp. 4.

This stability was observed not only in in-form settings but also under mixed-form and all-form conditions, which more closely represent distribution environments in which multiple processed forms coexist. Surface morphology, particle size, and light-scattering behavior vary substantially among tuber,

slice, and powder forms, whereas species-related compositional signatures are expected to be more conserved than external morphology. By internalizing derivative transformations, HyGD-Net integrates the raw spectrum with task-adaptive first- and second-order derivative representations. This multi-representation strategy may reduce sensitivity to form-dependent spectral variation while preserving features relevant to interspecies discrimination.

The perfect SWIR accuracy warrants careful consideration because near-ceiling performance can raise concerns about data leakage or overly favorable partitioning. Two observations make a simple model-specific explanation less compelling, although neither excludes leakage. First, near-saturated SWIR performance was observed across multiple architectures rather than in HyGD-Net alone. ResNet and HyGD-Net achieved 100.00% accuracy in all seven scenarios, while CNN, EfficientNetV2-S, Swin Transformer, and SVM remained at or above 98.88%. Second, the same partitioning strategy yielded substantially more dispersed VNIR performance, with accuracies ranging from 58.36% to 100.00%. These contrasting patterns are consistent with stronger intrinsic class separability in the SWIR region. In addition, the discriminative SWIR wavelengths overlap with C–H overtone regions relevant to the species-specific compounds discussed in Section 4.4. Nevertheless, the SWIR results should be interpreted cautiously because high performance across multiple models does not by itself rule out residual sample dependence or split-related optimism.

4.2. Contribution of individual module: Evidence from ablation and weight analysis

4.2.1. Adaptive optimization of the learnable filter

During training, the learnable filter weights deviated only slightly from their initial SG coefficients, with relative changes ranging from 0.03% to 0.60%. This limited deviation is consistent with the module's design objective of adapting to task-specific data while remaining close to the mathematically grounded SG initialization widely used in spectroscopy. The learned filters therefore retained the local derivative structure imposed at initialization while adapting to discrimination between *P. ternata* and *P. pedatisecta*.

The ablation study supported the value of this adaptation: learnable filters outperformed fixed SG filters in five of the seven VNIR experiments.

This design differs from previous approaches in two respects. Models that feed raw spectra directly into a network without explicit derivative modules do not incorporate established spectroscopic priors such as SG derivatives [23], whereas methods that tune SG parameters retain preprocessing as a separate stage and therefore limit joint optimization [24]. The learnable SG filter in HyGD-Net preserves the spectroscopic prior through informed initialization while jointly optimizing spectral transformation and classification within a single end-to-end framework.

4.2.2. Triple-branch CNN as the core module

Replacing the triple-branch architecture with a single-branch variant produced the largest mean performance reduction among the evaluated architectural modifications in the VNIR region, 6.30 pp, indicating the central contribution of this module to HyGD-Net. The learnable SG filters provide three complementary representations, namely the raw spectrum and first- and second-order derivative spectra, but performance also depends on effective extraction of discriminative features from each representation.

In the Single-Branch variant, the spectral representations are processed through a shared network, limiting representation-specific feature learning. Because the raw and derivative representations have distinct statistical characteristics, shared processing may reduce the model's ability to extract features tailored to each representation. By contrast, the triple-branch architecture assigns an independent CNN branch to each representation, enabling representation-specific feature extraction while preserving complementary information across branches.

The effect was pronounced in both in-form and mixed-form settings, with the largest accuracy reductions observed in Exp. 2 (Δ = 10.34 pp) and Exp. 4 (Δ = 9.68 pp). Mixed-form scenarios introduce greater spectral heterogeneity across physical forms, which may increase the value of independently extracting features from complementary spectral representations.

4.2.3. Position-wise gated fusion as an adaptive weighting mechanism

Across all seven VNIR experiments, gated fusion improved accuracy relative to simple concatenation by 0.23–4.09 pp, demonstrating a consistent contribution across scenarios. The learned gate values remained near the sigmoid midpoint of 0.5, while the most pronounced interbranch differences occurred within the 400–500 nm region in VNIR and the 900–1200 nm region in SWIR. These regions overlap with wavelength intervals highlighted by the IG analysis, suggesting that representation-specific modulation may be strongest in spectrally informative regions.

This position-dependent modulation is consistent with the complementary roles of the three representations. The raw spectrum preserves overall spectral structure, whereas derivative representations emphasize local spectral changes and can attenuate slowly varying baseline components. In discriminative regions, relative differences among gate values increase the contribution of particular representations; elsewhere, the gate values remain closer together. Because the gates are sigmoid outputs rather than normalized mixture weights, however, values near 0.5 should not be interpreted as equal branch contributions. Even within this bounded range, the small position-dependent shifts were sufficient to yield consistent accuracy gains over concatenation, indicating that the mechanism refines rather than replaces the contribution of each branch.

4.3. Classification across physical forms

4.3.1. Performance degradation across physical forms

The results support the concern raised in the Introduction that evaluation on a single physical form may not adequately characterize performance under more heterogeneous form compositions. Several benchmark models showed substantial cross-scenario variation, and ResNet differed by 13.96 pp between its best and worst VNIR scenarios.

Differences in classification difficulty were also evident across physical forms. The relatively high performance observed for powder samples in Exp. 3 may reflect reduced morphological heterogeneity

after pulverization, including less variation in surface structure, sample size, and localized light scattering. Under this interpretation, lower spectral heterogeneity may make species-related compositional differences more readily detectable. By contrast, the tuber condition in Exp. 1 was comparatively challenging for several models, potentially because intact samples exhibit greater surface inhomogeneity, size variation, and measurement-position-dependent light scattering. These form-dependent differences indicate that evaluation using a single physical form alone may not fully characterize model performance under conditions in which multiple processed forms are encountered.

4.3.2. Cross-form robustness of HyGD-Net

A direct comparison between the slice-only in-form scenario (Exp. 2) and the tuber-plus-slice mixed-form scenario (Exp. 4) illustrates differences in cross-scenario stability. ResNet accuracy decreased by 7.23 pp, from 92.36% to 85.13%, whereas HyGD-Net accuracy decreased by 2.16 pp, from 93.71% to 91.55%. In Exp. 4, HyGD-Net outperformed ResNet by 6.42 pp. Across all scenarios, HyGD-Net maintained relatively stable accuracy together with high species-level sensitivity and specificity, indicating that its learned representations remained discriminative as physical-form composition varied.

This robustness may be associated with the multi-order spectral representations used by HyGD-Net. As multiple physical forms are combined, within-class spectral distributions may broaden and form-dependent reflectance variation may increase. Derivative representations attenuate slowly varying baseline components and emphasize local changes in spectral shape, potentially helping to distinguish species-related information from form-associated variation. The triple-branch architecture independently extracts features from each representation, while the gated fusion module modulates their contributions across spectral positions. Together, these mechanisms help preserve discriminative information as form-related variability increases.

4.4. XAI-based identification and chemical validation of discriminative wavelengths

Although XAI techniques have been increasingly applied in agricultural spectroscopy, identified wavelengths are rarely evaluated against independent chemical analyses [27]. In addition to assessing classification performance, this study examined whether model attribution patterns were consistent with chemically meaningful evidence by comparing IG-derived consensus wavelengths with GC-TOF-MS profiles.

4.4.1. Cross-modality and cross-model convergence of discriminative wavelengths

Across modalities, consistent IG attribution patterns were observed in the VNIR and SWIR domains. Despite differences in sensor type, spatial resolution, and spectral resolution, both modalities identified consensus wavelengths within the overlapping 900–1000 nm region. Specifically, global consensus wavelengths at approximately 997 nm in VNIR and 991 nm in SWIR converged on the same broader spectral region across separately acquired modalities. This agreement suggests that the region may contain discriminative information that is not specific to a single sensor modality.

At the model and experimental levels, the robustness of the IG results was further supported by intermodel and cross-scenario agreement. Despite substantial architectural differences among convolutional, residual, attention-based, and gated-fusion models, consensus wavelengths recurred across multiple experimental scenarios encompassing different physical-form compositions. This recurrence indicates that the identified regions are comparatively robust to model architecture and form-associated spectral variability.

Taken together, convergence across spectral modalities, model architectures, and experimental scenarios supports the interpretation that the models repeatedly use related spectral regions for interspecies discrimination. This consistency reduces, but does not eliminate, the likelihood that the identified regions arise solely from model-specific or scenario-specific artifacts. Chemical interpretation therefore requires additional evidence, as examined through the GC-TOF-MS comparison below.

4.4.2. Cross-validation of XAI consensus wavelengths with GC-TOF-MS chemical analysis

The chemical plausibility of the identified consensus wavelengths was further examined through comparison with the GC-TOF-MS results (Fig. 9). Among the nine identified compounds, four were detected exclusively in *P. pedatisecta*, namely hexanoic acid, (E,E)-2,4-heptadienal, (E)-2-octenal, and nonanal, whereas two, 2-heptanone and phenylacetaldehyde, were detected exclusively in *P. ternata*. These qualitative detection patterns indicate interspecies differences in volatile composition. Near-infrared absorption regions associated with the functional groups present in these compounds were subsequently compared with the IG-derived consensus wavelengths.

In the VNIR region, the most stable consensus wavelength, approximately 444 nm within the broader 440–453 nm cluster, is consistent with the $\pi \rightarrow \pi^*$ electronic transition region associated with conjugated double-bond systems in carotenoids [38]. Variation in this spectral region has been associated with differences in pigment composition, including carotenoid and flavonoid contributions [39]. In addition, consensus wavelengths within the 930–1000 nm range, including approximately 997 nm in VNIR and 991 nm in SWIR, are consistent with an O–H overtone absorption region near 950–1000 nm [40]. Hexanoic acid, detected exclusively in *P. pedatisecta*, contains a carboxyl group and therefore provides one plausible chemical context for interspecies differences involving O–H-associated spectral features. However, this qualitative association does not establish that hexanoic acid specifically caused the observed attribution pattern.

In the SWIR region, the consensus wavelength near 1157 nm lies within a C–H second-overtone region [18]. Previous studies have reported second overtones of sp^2 C–H bonds in aromatic and olefinic structures at shorter wavelengths, approximately 1130–1140 nm, and second overtones of sp^3 C–H bonds in alkyl structures at longer wavelengths, approximately 1190–1210 nm [41]. Compounds such as (E,E)-2,4-heptadienal and (E)-2-octenal contain both unsaturated and alkyl structural moieties, whereas nonanal contains a saturated carbon chain. The consensus wavelength near 1157 nm may therefore lie within a

broader region influenced by overlapping C–H overtone contributions associated with interspecies compositional differences.

Overall, the VNIR consensus wavelengths were associated primarily with pigment-related electronic transitions and O–H overtone regions, whereas the SWIR consensus wavelengths were concentrated within C–H overtone regions. This complementary pattern suggests that the two spectral domains capture distinct but potentially synergistic chemical information. Importantly, the qualitative agreement between IG-derived wavelength importance and GC-TOF-MS-informed functional-group interpretation is consistent with the possibility that the learned discriminative patterns reflect underlying compositional differences rather than arbitrary numerical features. However, the GC-TOF-MS analysis provides chemical corroboration rather than direct compound-specific validation of individual wavelengths.

4.5. Limitations and future work

Several limitations should be acknowledged. First, all samples were obtained from a single supplier, NIKOM, and variability across geographical origins and harvest years was therefore not adequately represented. Second, GC-TOF-MS analysis was limited to volatile constituents, precluding direct evaluation of the contributions of nonvolatile macromolecules, such as starch, other polysaccharides, and proteins, to the SWIR consensus wavelengths. In addition, SWIR classification reached ceiling-level accuracy across all models, making it difficult to distinguish performance differences among architectures in this range. Such differences can be adequately assessed only with more challenging and diverse sample sets in future studies.

Future work should therefore (i) include samples spanning diverse cultivars, geographical origins, and harvest years to assess model generalizability and (ii) incorporate complementary liquid chromatography–mass spectrometry (LC-MS) and Fourier-transform infrared (FT-IR) spectroscopy to extend chemical evaluation to nonvolatile constituents. In addition, field applicability should be assessed by developing

and validating a portable hyperspectral system based on the key wavelength regions identified in this study.

5. Conclusion

This study proposed HyGD-Net, an end-to-end hyperspectral imaging-based deep learning architecture for the nondestructive authentication of morphologically similar *P. ternata* and *P. pedatisecta* across multiple processed forms. HyGD-Net integrates a learnable derivative filter, a triple-branch CNN, and position-wise gated fusion to jointly optimize spectral transformation and classification using VNIR and SWIR spectra without external derivative preprocessing. The model therefore reduces dependence on manually selected preprocessing while providing a unified architecture for multi-form authentication of medicinal plant materials.

HyGD-Net achieved 100.00% accuracy across all SWIR scenarios. In the VNIR range, it achieved a mean accuracy of 94.46% across seven scenarios spanning tuber, slice, and powder forms. These results demonstrate strong performance under in-form, mixed-form, and all-form conditions, which more closely represent post-harvest distribution environments in which medicinal plant materials occur in multiple processed forms. Ablation analyses further supported the contributions of the architectural components, particularly in the VNIR region: the learnable derivative filter enabled task-adaptive spectral transformation, the triple-branch CNN extracted representation-specific complementary features, and the gated fusion module modulated feature contributions across spectral positions.

To address the black-box nature of deep learning predictions, an Integrated Gradients-based, multi-model, multi-scenario consensus analysis was used to identify discriminative wavelengths that recurred across physical-form conditions. GC-TOF-MS profiling provided complementary chemical context for these XAI-derived wavelengths by identifying species-specific volatile profiles and comparing them with relevant functional-group absorption regions. Interpretability was therefore provided by the surrounding analytical workflow rather than by intrinsic model transparency. HyGD-Net itself is not intrinsically interpretable; instead, the broader workflow links prediction, attribution-based explanation, and independent chemical evidence. Taken together, the findings support HyGD-Net as a preprocessing-free,

form-robust framework with chemically corroborated explanations for the post-harvest authentication of medicinal plant materials.

Acknowledgements

Not applicable.

Conflict of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval

Not applicable. This study did not involve human participants, human-derived data or materials, or animals.

Consent for publication

Not applicable.

Availability of data and material

The source code of HyGD-Net, including the training pipeline, the learnable Savitzky–Golay derivative module, and the Integrated Gradients analysis scripts, is publicly available at https://github.com/YEAJIN412/HyGD-Net. The hyperspectral image datasets and GC-TOF-MS raw data generated and analyzed in this study are not publicly deposited because they form part of an ongoing project funded by the Korean Ministry of Food and Drug Safety, but are available from the corresponding author upon reasonable request.

Contributions

Yea-Jin Park: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Yu-Jin Jeon: Investigation, Validation. So Jin Park: Investigation, Validation. Da Hye Ryu: Investigation, Formal analysis, Data curation. Hye Min Kim: Investigation, Formal analysis, Data curation. Ho-Youn Kim: Supervision, Methodology, Formal analysis, Resources. Soo Hyun Park: Investigation, Formal analysis, Data curation. Dae-Hyun Jung: Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Funding

This research was supported by a grant (23192MFDS106) from the Ministry of Food and Drug Safety in 2023-2027.

References

1. Fortune Business Insights. Herbal medicine market size, share, trends, report, 2034 [Internet]. 2026 [cited 2026 Jul 16]. Available from: https://www.fortunebusinessinsights.com/herbal-medicine-market-106320

2. Burki T. WHA adopts new strategy on traditional medicine. Lancet. 2025;405:1897-1898. doi:10.1016/S0140-6736(25)01142-0

3. Convention on Biological Diversity. Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization [Internet]. 2011 [cited 2026 Jul 16]. Available from: https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf

4. Ichim MC. The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration. Front Pharmacol. 2019;10:1227. doi:10.3389/fphar.2019.01227

5. Ichim MC, Häser A, Nick P. Microscopic authentication of commercial herbal products in the globalized market: potential and limitations. Front Pharmacol. 2020;11:876. doi:10.3389/fphar.2020.00876

6. Bai J, Qi J, Yang L, Wang Z, Wang R, Shi Y. A comprehensive review on ethnopharmacological, phytochemical, pharmacological and toxicological evaluation, and quality control of Pinellia ternata (Thunb.) Breit. J Ethnopharmacol. 2022;298:115650. doi:10.1016/j.jep.2022.115650

7. Yu HL, Wang W, Wu H, Shen M, Zhang YB, Li SH. Effect of processing on toxic components lectin from four kinds of Araceae toxic medicines. Zhongguo Zhong Yao Za Zhi. 2019;44:5398-5404. doi:10.19540/j.cnki.cjcmm.20190916.302

8. Zhang T, Xu F, Ruhsam M, Feng L, Zhang M, Wang Z, et al. A nucleotide signature for the identification of Pinelliae Rhizoma (Banxia) and its products. Mol Biol Rep. 2022;49:7753-7763. doi:10.1007/s11033-022-07600-0

9. Yu HL, Zhao TF, Wu H, Pan YZ, Zhang Q, Wang KL, et al. Pinellia ternata lectin exerts a pro-inflammatory effect on macrophages by inducing the release of pro-inflammatory cytokines, the activation of the nuclear factor-kappaB signaling pathway and the overproduction of reactive oxygen species. Int J Mol Med. 2015;36:1127-1135. doi:10.3892/ijmm.2015.2315

10. Zhou Y, Zhang D, Li H, Zhang H, Fang J, Ma Y, et al. The scientific basis and advantage of human experiential assessment in the quality control of Chinese herbal medicines exampling as Schisandrae Chinensis Fructus. Sci Rep. 2018;8:5695. doi:10.1038/s41598-018-23619-5

11. Chen S, Yin X, Han J, Sun W, Yao H, Song J, et al. DNA barcoding in herbal medicine: retrospective and prospective. J Pharm Anal. 2023;13:431-441. doi:10.1016/j.jpha.2023.03.008

12. Shah M, Patel N, Tripathi N, Vyas VK. Capillary electrophoresis methods for impurity profiling of drugs: a review of the past decade. J Pharm Anal. 2022;12:15-28. doi:10.1016/j.jpha.2021.06.009

13. Sun X, Qian H. Chinese herbal medicine image recognition and retrieval by convolutional neural network. PLoS One. 2016;11:e0156327. doi:10.1371/journal.pone.0156327

14. Jeon YJ, Park SJ, Lee H, Kim HY, Jung DH. Deep learning-based model for effective classification of Ziziphus jujuba using RGB images. AgriEngineering. 2024;6:4604-4619. doi:10.3390/agriengineering6040263

15. Wäldchen J, Mäder P. Plant species identification using computer vision techniques: a systematic literature review. Arch Comput Methods Eng. 2018;25:507-543. doi:10.1007/s11831-016-9206-z

16. Bhargava A, Sachdeva A, Sharma K, Alsharif MH, Uthansakul P, Uthansakul M. Hyperspectral imaging and its applications: a review. Heliyon. 2024;10:e33208. doi:10.1016/j.heliyon.2024.e33208

17. Ru C, Li Z, Tang R. A hyperspectral imaging approach for classifying geographical origins of Rhizoma Atractylodis Macrocephalae using the fusion of spectrum-image in VNIR and SWIR ranges (VNIR-SWIR-FuSI). Sensors. 2019;19:2045. doi:10.3390/s19092045

18. Weyer LG, Lo SC. Spectra-structure correlations in the near-infrared. In: Chalmers JM, Griffiths PR, editors. Handbook of vibrational spectroscopy. 1st ed. Chichester: Wiley; 2001. doi:10.1002/0470027320.s4102

19. Cai Z, Huang Z, He M, Li C, Qi H, Peng J, et al. Identification of geographical origins of Radix Paeoniae Alba using hyperspectral imaging with deep learning-based fusion approaches. Food Chem. 2023;422:136169. doi:10.1016/j.foodchem.2023.136169

20. Xing W, Wang X, Ma Z, Xing Y, Dun X, Cheng X. Rapid discrimination of Platycodonis radix geographical origins using hyperspectral imaging and deep learning. Optics. 2025;6:52. doi:10.3390/opt6040052

21. Jiao Y, Li Z, Chen X, Fei S. Preprocessing methods for near-infrared spectrum calibration. J Chemom. 2020;34:e3306. doi:10.1002/cem.3306

22. Rinnan Å, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem. 2009;28:1201-1222. doi:10.1016/j.trac.2009.07.007

23. Chen YY, Wang ZB. End-to-end quantitative analysis modeling of near-infrared spectroscopy based on convolutional neural network. J Chemom. 2019;33:e3122. doi:10.1002/cem.3122

24. Zhang G, Hao H, Wang Y, Jiang Y, Shi J, Yu J, et al. Optimized adaptive Savitzky-Golay filtering algorithm based on deep learning network for absorption spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc. 2021;263:120187. doi:10.1016/j.saa.2021.120187

25. Jiang Z, Zhong L, Xue J, Lv J, Zhou F, Zhou Y, et al. Data fusion based on near-infrared spectroscopy and hyperspectral imaging technology for rapid adulteration detection of Ganoderma lucidum spore powder. Microchem J. 2023;193:109190. doi:10.1016/j.microc.2023.109190

26. Ryo M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. Artif Intell Agric. 2022;6:257-265. doi:10.1016/j.aiia.2022.11.003

27. Ahmed MT, Ahmed MW, Kamruzzaman M. A systematic review of explainable artificial intelligence for spectroscopic agricultural quality assessment. Comput Electron Agric. 2025;235:110354. doi:10.1016/j.compag.2025.110354

28. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2023. p. 4015-4026.

29. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273-297. doi:10.1007/BF00994018

30. Breiman L. Random forests. Mach Learn. 2001;45:5-32. doi:10.1023/A:1010933404324

31. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-444. doi:10.1038/nature14539

32. Tan M, Le Q. EfficientNetV2: smaller models and faster training. In: Proceedings of the 38th International Conference on Machine Learning (ICML); 2021. p. 10096-10106.

33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770-778.

34. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021. p. 10012-10022.

35. Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 1964;36:1627-1639. doi:10.1021/ac60214a047

36. Arevalo J, Solorio T, Montes-y-Gómez M, González FA. Gated multimodal units for information fusion. arXiv preprint. 2017. Available from: https://arxiv.org/abs/1702.01992

37. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML); 2017. p. 3319-3328.

38. Lichtenthaler HK, Buschmann C. Chlorophylls and carotenoids: measurement and characterization by UV-VIS spectroscopy. Curr Protoc Food Anal Chem. 2001;1:F4.3.1-F4.3.8. doi:10.1002/0471142913.faf0403s01

39. Taniguchi M, LaRocca CA, Bernat JD, Lindsey JS. Digital database of absorption spectra of diverse flavonoids enables structural comparisons and quantitative evaluations. J Nat Prod. 2023;86:1087-1119. doi:10.1021/acs.jnatprod.2c00720

40. Beć KB, Grabska J, Huck CW. Near-infrared spectroscopy in bio-applications. Molecules. 2020;25:2948. doi:10.3390/molecules25122948

41. Ma L, Peng Y, Pei Y, Zeng J, Shen H, Cao J, et al. Systematic discovery about NIR spectral assignment from chemical structural property to natural chemical compounds. Sci Rep. 2019;9:9503. doi:10.1038/s41598-019-45945-y

Captions

Fig. 1. Schematic overview of the experimental workflow. (a) Preparation of *P. ternata* and *P. pedatisecta* in three physical forms (tuber, slice, and powder) and hyperspectral image acquisition in the visible and near-infrared (VNIR; 400–1000 nm) and short-wave infrared (SWIR; 900–1700 nm) spectral regions. (b) Architecture of the proposed HyGD-Net (left) and six benchmark classifiers (right) used for comparative performance evaluation. (c) Identification of consensus discriminative wavebands through Integrated Gradients-based explainable artificial intelligence (XAI) analysis across multiple scenarios, followed by chemical cross-validation using gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) to assess the plausibility of the model's decision-making basis.

Fig. 2. Hyperspectral imaging and spectral extraction pipeline. (a) Imaging chamber equipped with halogen and light-emitting diode (LED) light sources. (b) Region-of-interest (ROI) extraction for tuber (top), slice (middle), and powder (bottom) samples, with representative spectra.

Fig. 3. Architecture of the Hyperspectral Gated Derivative Network (HyGD-Net). (a) Overall network pipeline from the input spectrum to the classification output through learnable filtering, a triple-branch convolutional neural network (CNN), gated fusion, and a classification head. (b) Learnable filter comprising Savitzky–Golay-initialized one-dimensional convolution (Conv1D) kernels and trainable residuals. (c) Position-wise gated fusion with sigmoid-based branch weighting.

Fig. 4. Mean reflectance spectra of *P. ternata* and *P. pedatisecta* across physical forms in the visible and near-infrared (VNIR) and short-wave infrared (SWIR) regions. (a–c) VNIR (400–1000 nm): (a) tuber, (b) slice, and (c) powder. (d–f) SWIR (900–1700 nm): (d) tuber, (e) slice, and (f) powder.

Fig. 5. Classification accuracy and variability of HyGD-Net and six benchmark models across seven experimental scenarios. (a) Scenario-specific classification accuracy in the visible and near-infrared (VNIR; 400–1000 nm) range. (b) Boxplot of the accuracy distribution across seven experiments in the VNIR range. (c) Scenario-specific classification accuracy in the short-wave infrared (SWIR; 900–1700 nm) range. (d) Boxplot of the accuracy distribution across seven experiments in the SWIR range.

Fig. 6. Ablation results of the Hyperspectral Gated Derivative Network (HyGD-Net) across seven experimental scenarios. Circle size and percentage labels indicate decreases in performance associated with removing or modifying each component, whereas the bar plots show cumulative performance gains from the baseline architecture to the full model.

Fig. 7. Distribution of learned gate values in the gated fusion module across wavelength positions for the seven experimental scenarios. (a) Visible and near-infrared (VNIR; 400–1000 nm) range. (b) Short-wave infrared (SWIR; 900–1700 nm) range. Within each panel, subplots correspond to Exp. 1–7. Lines denote the mean gate value at each wavelength position for the raw, first-order derivative, and second-order derivative branches; shaded bands denote ±1 standard deviation across test samples at each wavelength position.

Fig. 8. Integrated Gradients (IG) attribution analysis across five deep learning models and seven experimental scenarios. (a) Visible and near-infrared (VNIR; 400–1000 nm) range. (b) Short-wave infrared (SWIR; 900–1700 nm) range. Consensus wavelength bands are highlighted.

Fig. 9. Cross-validation of consensus wavelengths using gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) analysis. (a) Integrated Gradients attribution scores across the full spectral range (400–1700 nm), with identified consensus wavelengths. (b) Species-specific volatile compounds detected by GC-TOF-MS. (c) Correspondence between consensus wavelengths and near-infrared overtone absorption regions associated with the identified functional groups.

Figures

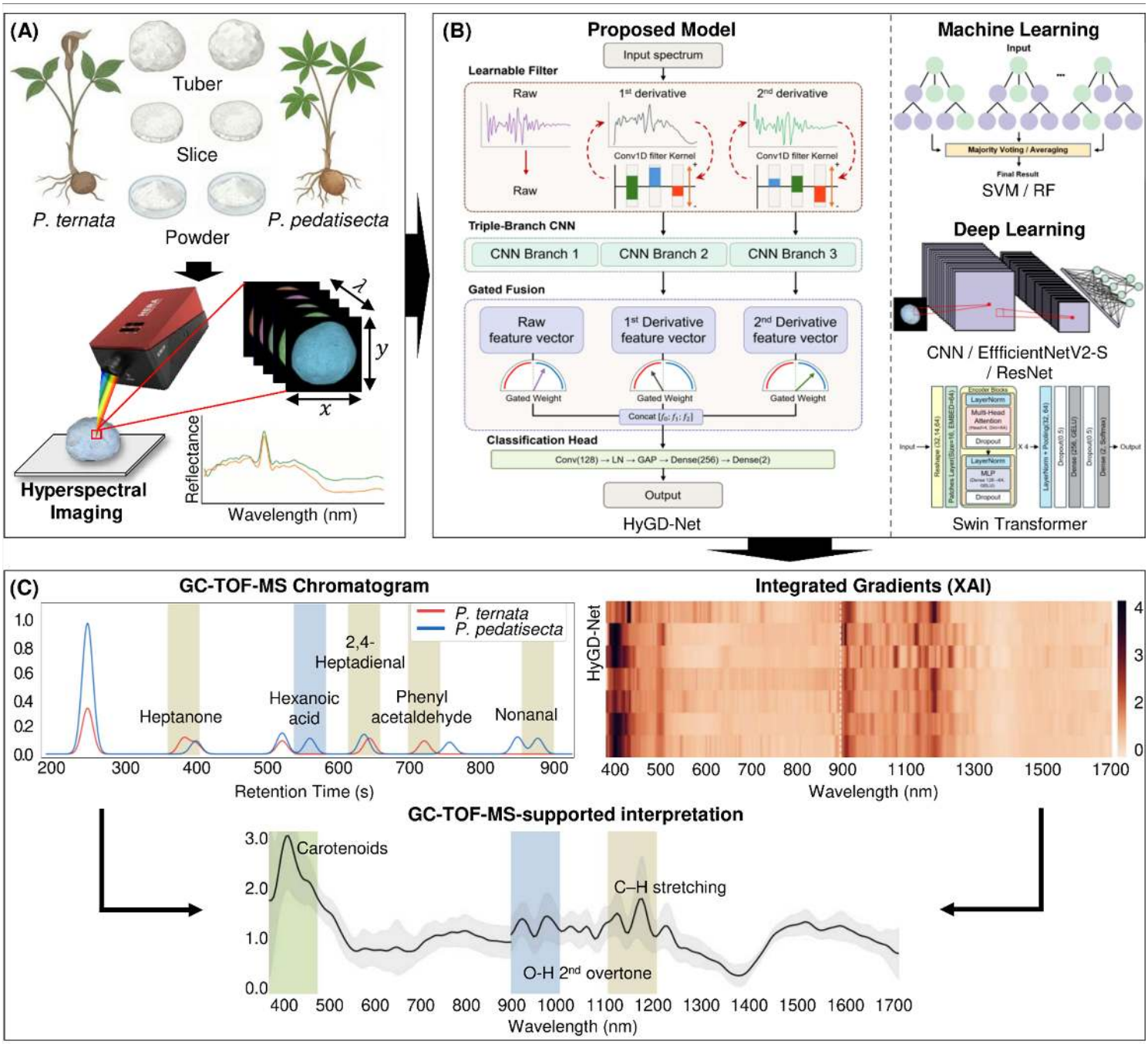

Figure 1

Schematic overview of the experimental workflow. (a) Preparation of *P. ternata* and *P. pedatisecta*in three physical forms (tuber, slice, and powder) and hyperspectral image acquisition in the visible and near-infrared (VNIR; 400–1000 nm) and short-wave infrared (SWIR; 900–1700 nm) spectral regions. (b) Architecture of the proposed HyGD-Net (left) and six benchmark classifiers (right) used for comparative performance evaluation. (c) Identification of consensus discriminative wavebands through Integrated Gradients-based explainable artificial intelligence (XAI) analysis across multiple scenarios, followed by

chemical cross-validation using gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) to assess the plausibility of the model's decision-making basis.

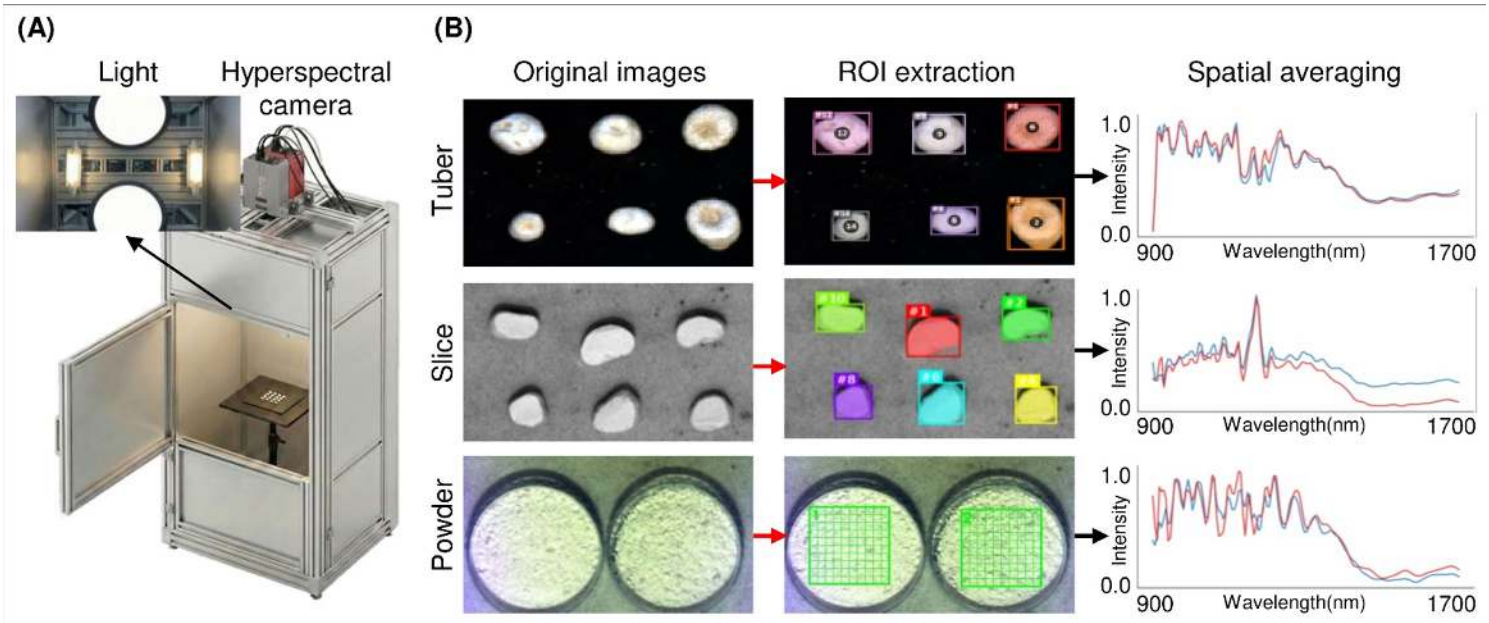

Figure 2

Hyperspectral imaging and spectral extraction pipeline. (a) Imaging chamber equipped with halogen and light-emitting diode (LED) light sources. (b) Region-of-interest (ROI) extraction for tuber (top), slice (middle), and powder (bottom) samples, with representative spectra.

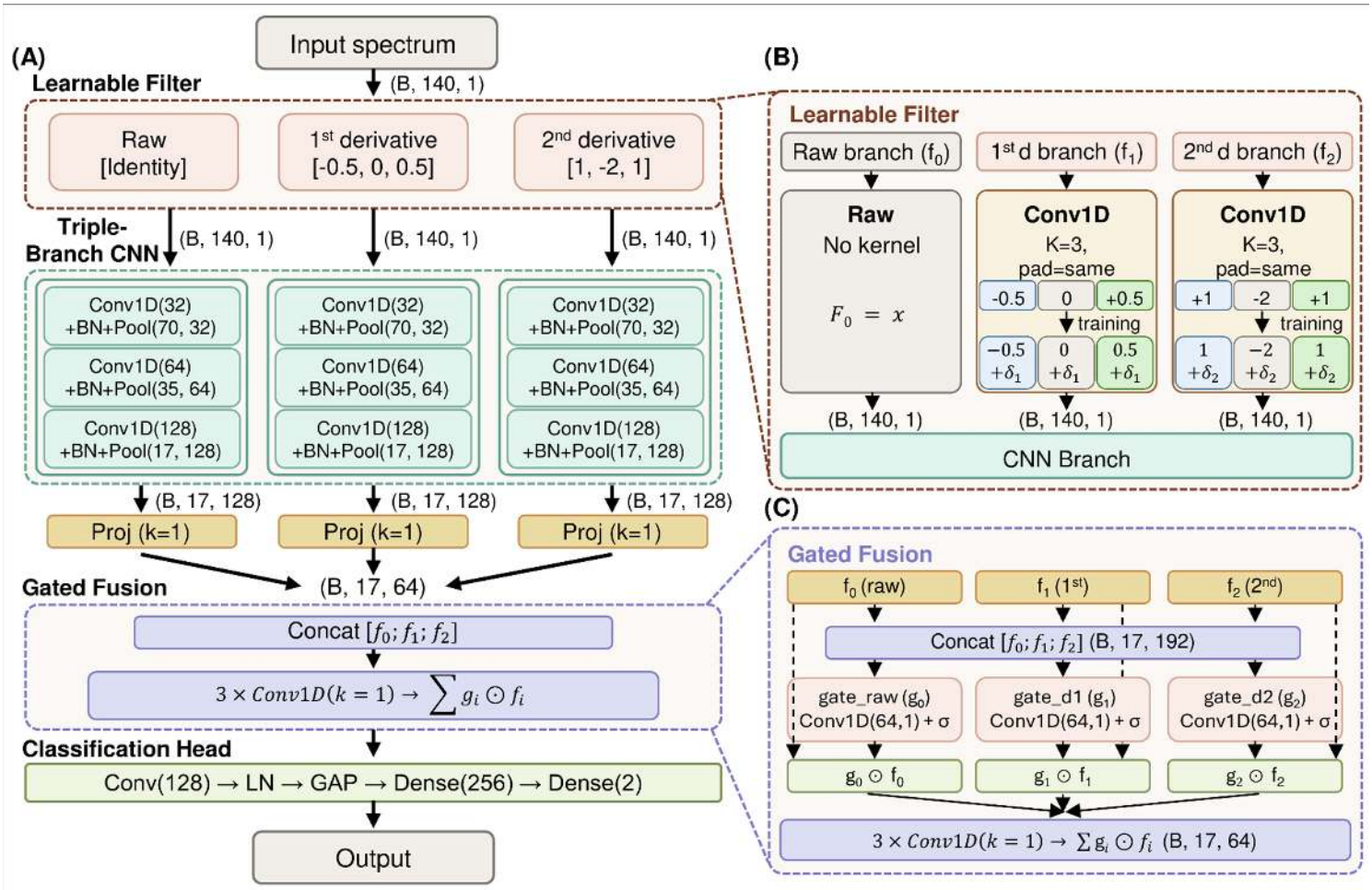

Figure 3

Architecture of the Hyperspectral Gated Derivative Network (HyGD-Net). (a) Overall network pipeline from the input spectrum to the classification output through learnable filtering, a triple-branch convolutional neural network (CNN), gated fusion, and a classification head. (b) Learnable filter comprising Savitzky–Golay-initialized one-dimensional convolution (Conv1D) kernels and trainable residuals. (c) Position-wise gated fusion with sigmoid-based branch weighting.

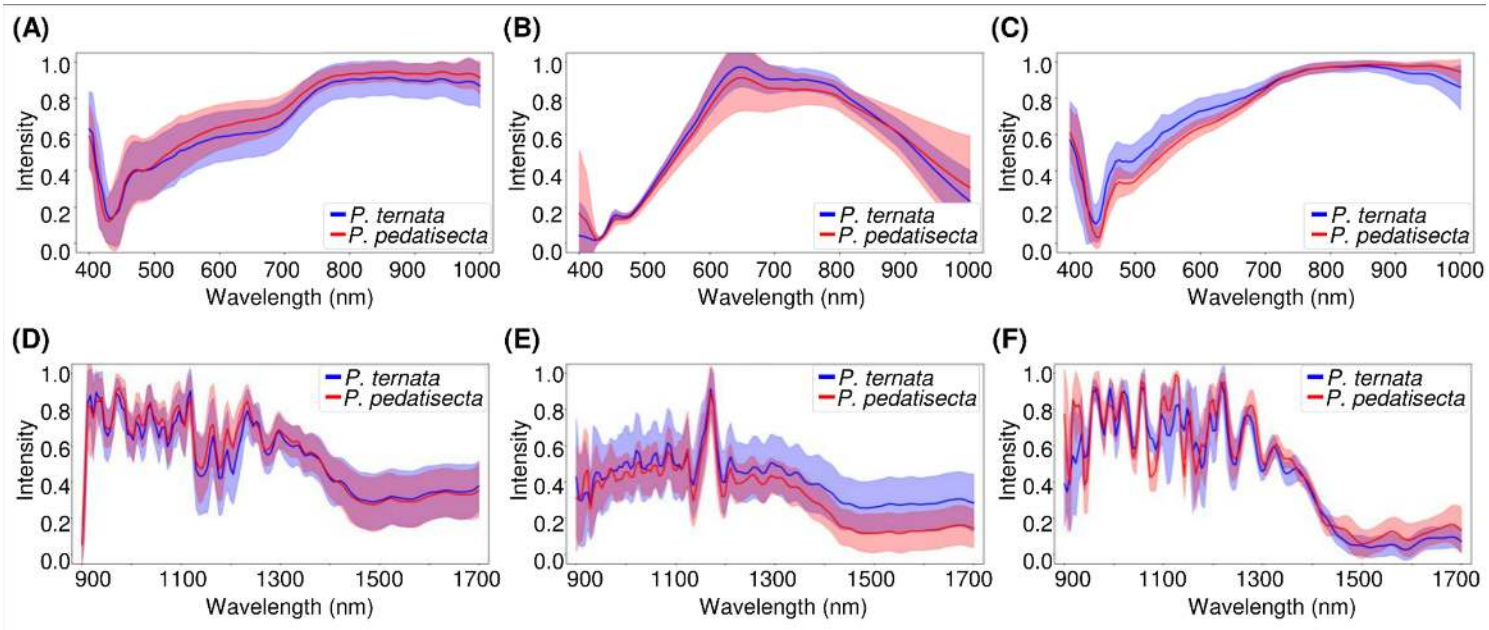

Figure 4

Mean reflectance spectra of *P. ternata* and *P. pedatisecta*across physical forms in the visible and near-infrared (VNIR) and short-wave infrared (SWIR) regions. (a–c) VNIR (400–1000 nm): (a) tuber, (b) slice, and (c) powder. (d–f) SWIR (900–1700 nm): (d) tuber, (e) slice, and (f) powder.

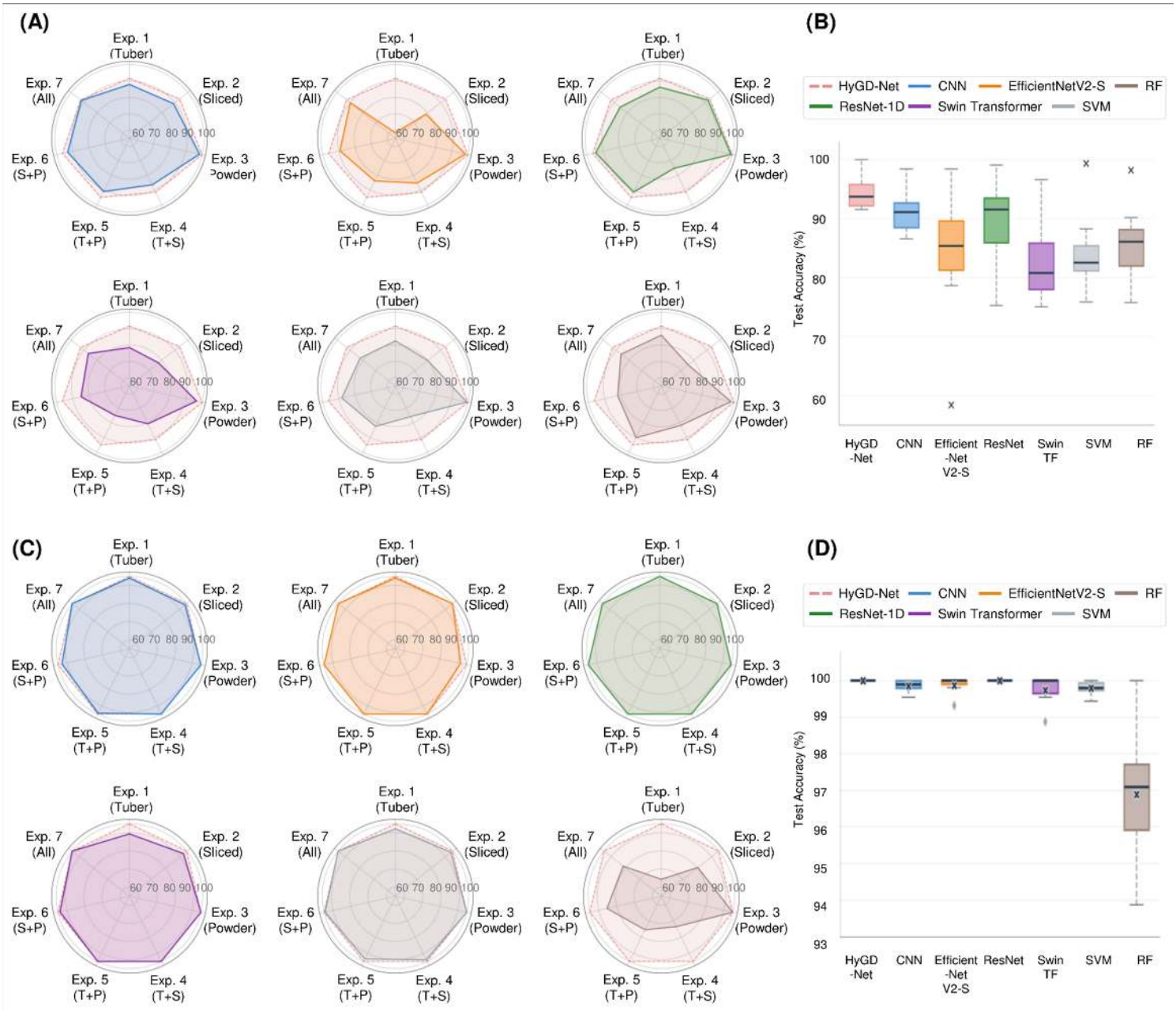

Figure 5

Classification accuracy and variability of HyGD-Net and six benchmark models across seven experimental scenarios. (a) Scenario-specific classification accuracy in the visible and near-infrared (VNIR; 400–1000 nm) range. (b) Boxplot of the accuracy distribution across seven experiments in the VNIR range. (c) Scenario-specific classification accuracy in the short-wave infrared (SWIR; 900–1700 nm) range. (d) Boxplot of the accuracy distribution across seven experiments in the SWIR range.

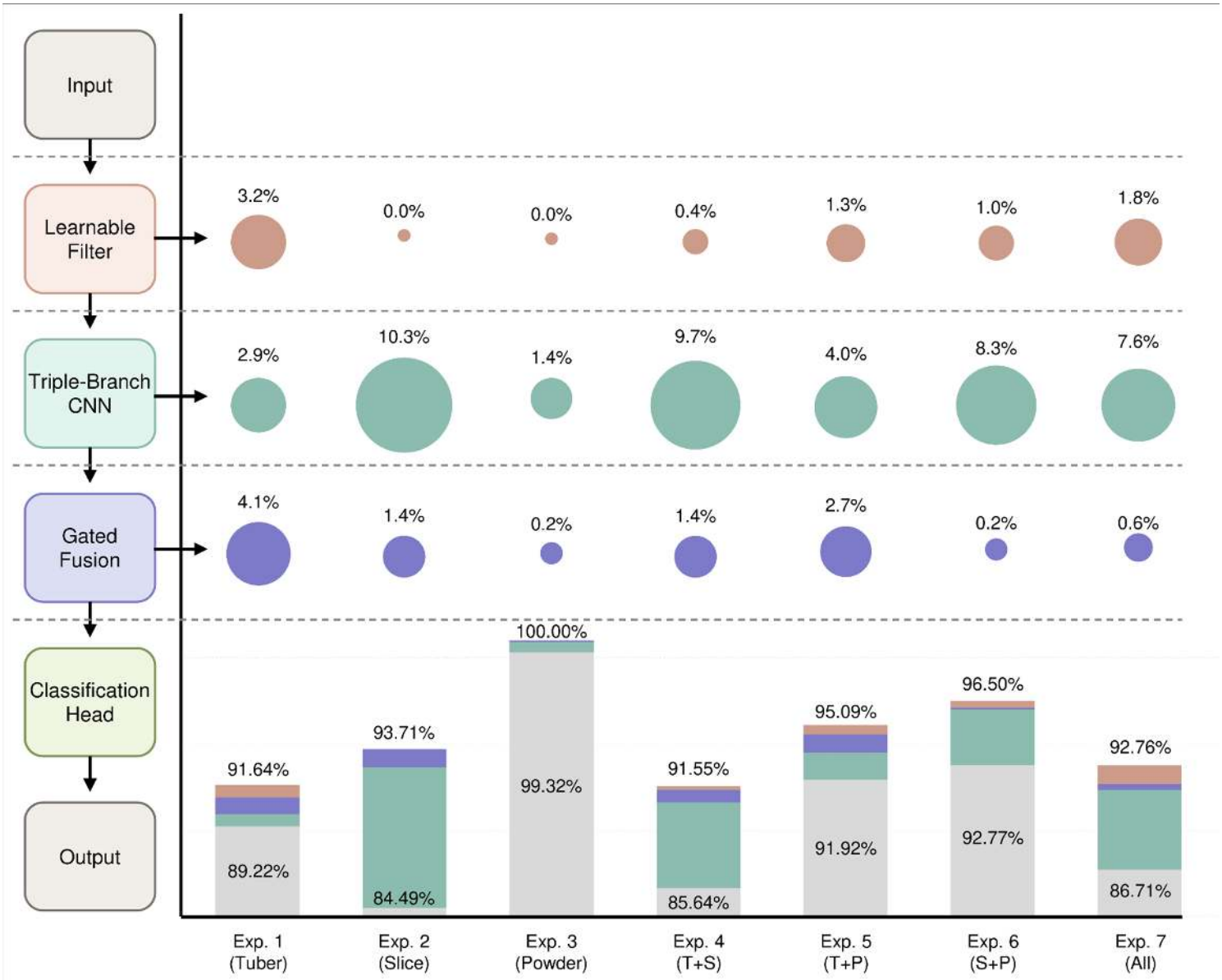

Figure 6

Ablation results of the Hyperspectral Gated Derivative Network (HyGD-Net) across seven experimental scenarios. Circle size and percentage labels indicate decreases in performance associated with removing or modifying each component, whereas the bar plots show cumulative performance gains from the baseline architecture to the full model.

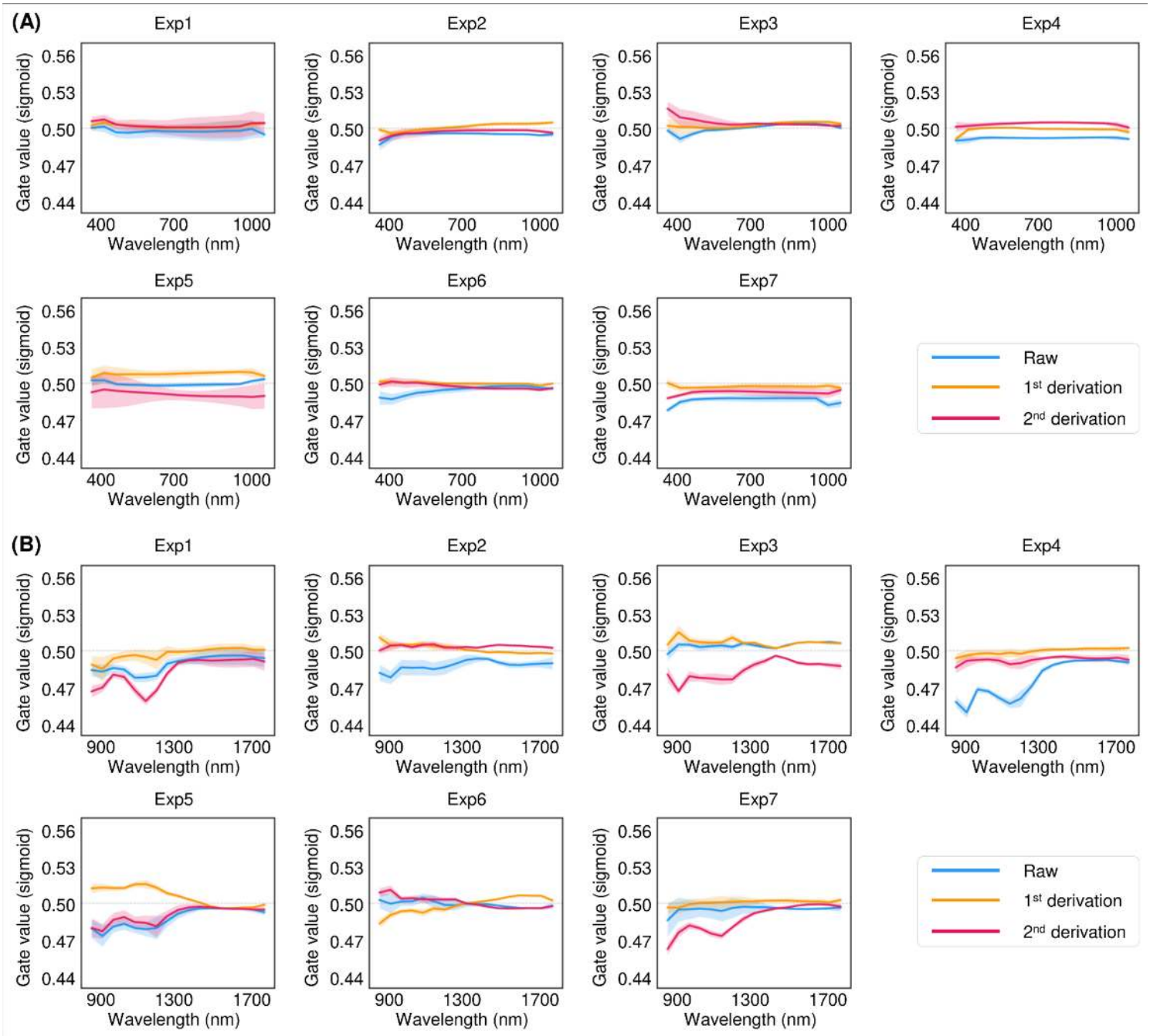

Figure 7

Distribution of learned gate values in the gated fusion module across wavelength positions for the seven experimental scenarios. (a) Visible and near-infrared (VNIR; 400–1000 nm) range. (b) Short-wave infrared (SWIR; 900–1700 nm) range. Within each panel, subplots correspond to Exp. 1–7. Lines denote the mean gate value at each wavelength position for the raw, first-order derivative, and second-order derivative branches; shaded bands denote ±1 standard deviation across test samples at each wavelength position.

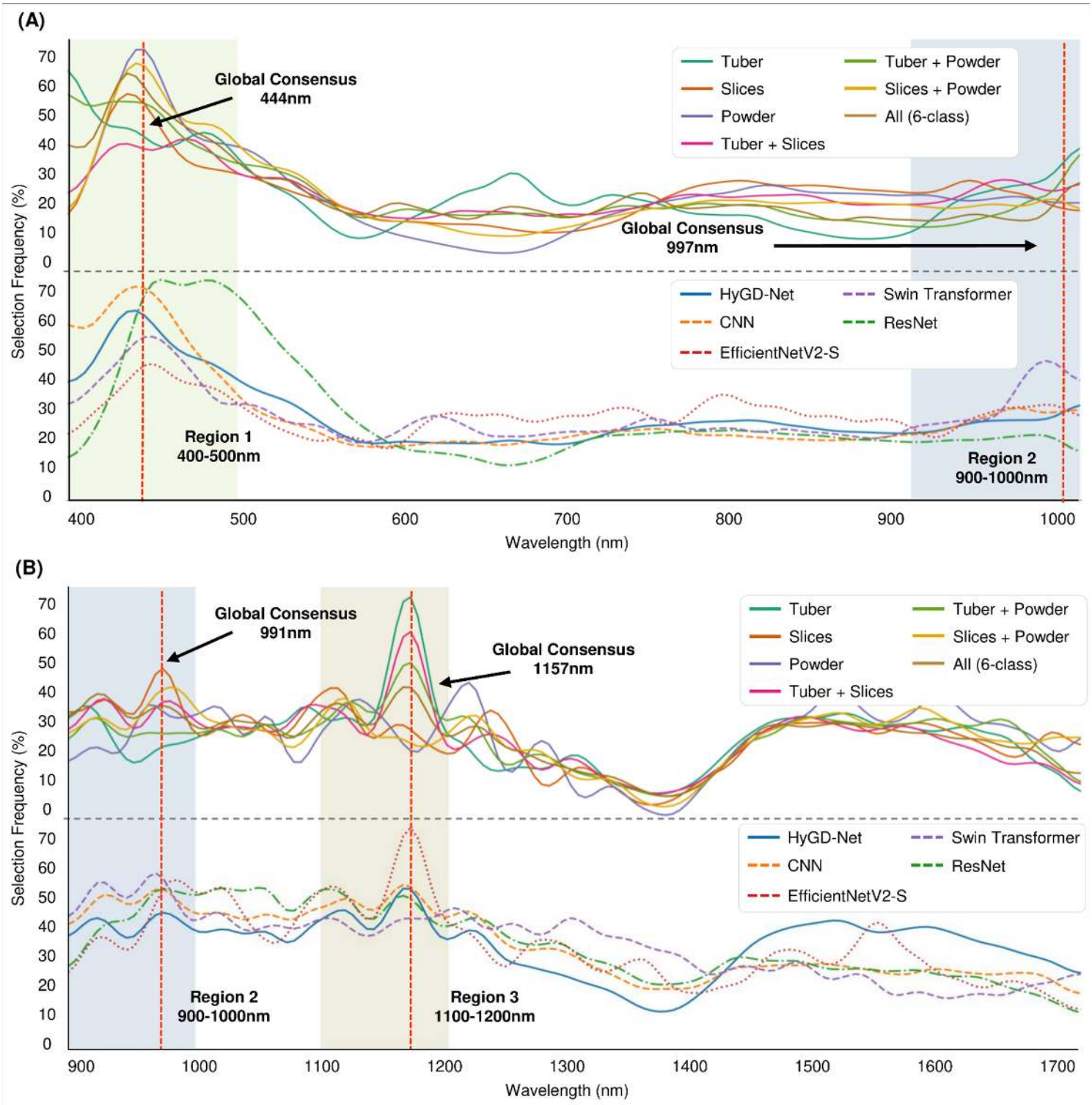

Figure 8

Integrated Gradients (IG) attribution analysis across five deep learning models and seven experimental scenarios. (a) Visible and near-infrared (VNIR; 400–1000 nm) range. (b) Short-wave infrared (SWIR; 900–1700 nm) range. Consensus wavelength bands are highlighted.

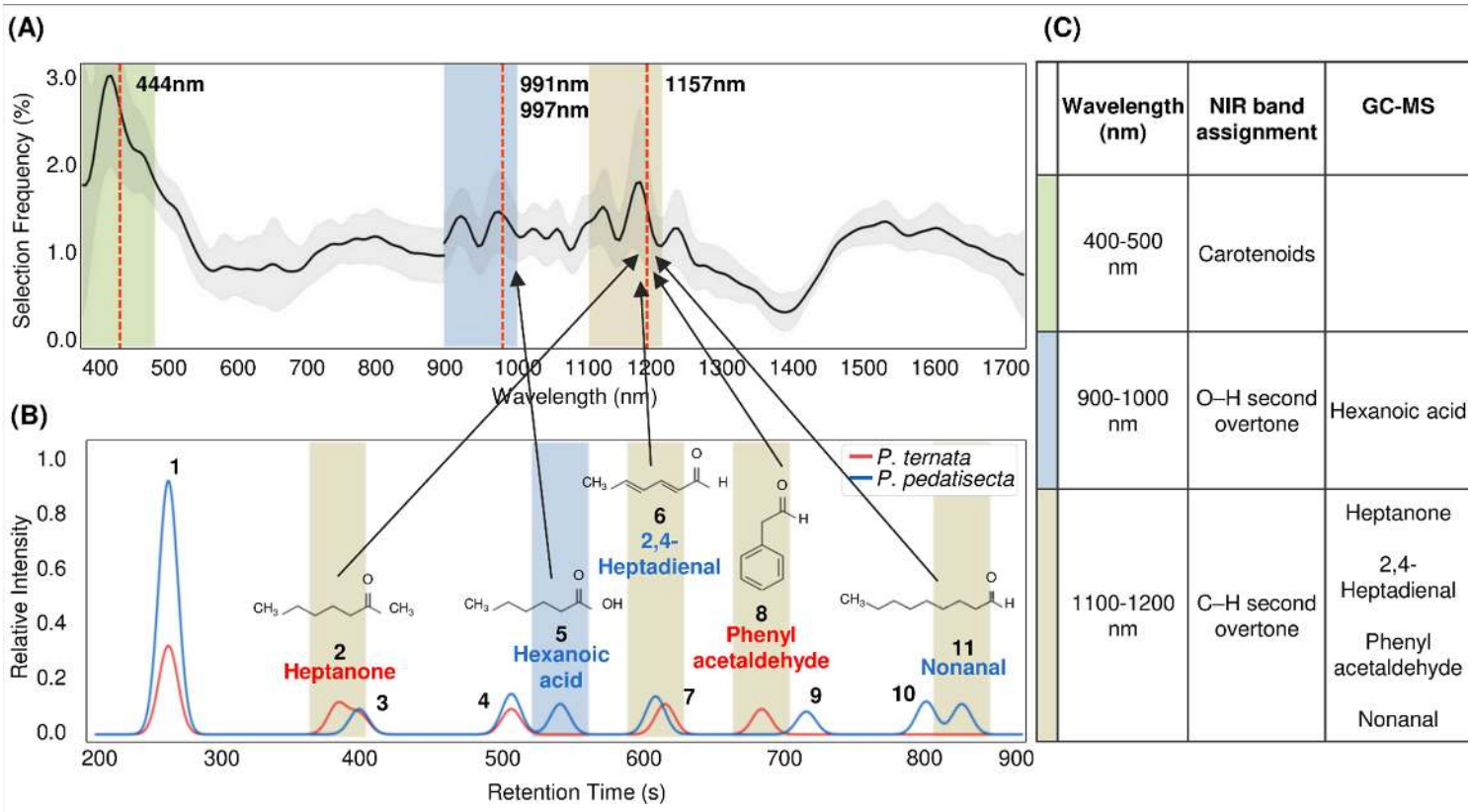

Figure 9

Cross-validation of consensus wavelengths using gas chromatography–time-of-flight mass spectrometry (GC-TOF-MS) analysis. (a) Integrated Gradients attribution scores across the full spectral range (400–1700 nm), with identified consensus wavelengths. (b) Species-specific volatile compounds detected by GC-TOF-MS. (c) Correspondence between consensus wavelengths and near-infrared overtone absorption regions associated with the identified functional groups.