

Contextualizing CNN attribution maps using predefined image-derived descriptors in medicinal plant seed classification

Su-Min Chin

Kyung Hee University

Yea-Jin Park

Kyung Hee University

Dae-Hyun Jung

daehyun@khu.ac.kr

Kyung Hee University

Article

Keywords: Medicinal plant seeds, Seed image classification, CNN attribution, Integrated Gradients, Descriptor contextualization, Foreground restriction

Posted Date: August 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10586547/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 16th, 2026.
See the published version at <https://doi.org/10.1038/s41598-026-75686-8>.

Contextualizing CNN attribution maps using predefined image-derived descriptors in medicinal plant seed classification

Su-Min Chin^{a,b}, Yea-Jin Park^{a,b}, Dae-Hyun Jung^{a,b,*}

^a Department of Smart Farm Science, Kyung Hee University, Yongin, Gyeonggi-do, Republic of Korea; e-mail: sgn08178@khu.ac.kr, lou0412@khu.ac.kr

^b BK21 Interdisciplinary Program in IT-Bio Convergence System, Kyung Hee University, Yongin, Republic of Korea

* Corresponding author, e-mail address: daehyun@khu.ac.kr

Abstract

Deep learning classifiers can achieve high predictive performance in medicinal plant seed image classification, but prediction accuracy alone does not clarify which visual properties are linked to model responses. This issue is particularly relevant for visually similar medicinal plant seeds, related taxa, and potential substitutes that may share overlapping morphology. Here, we analyzed RGB images from five medicinal plant seed classes by comparing CNN attribution maps with predefined image-derived descriptor maps within the seed foreground. ResNet50, EfficientNet-B0, and ConvNeXt-Small showed near-ceiling classification performance under controlled imaging and preprocessing conditions, and zero-baseline absolute Integrated Gradients (IG) maps from ConvNeXt-Small were used as the primary attribution reference. Among the tested descriptor maps, LAB L, Brightness, and fast Fourier transform (FFT) low-pass showed the highest positive foreground-restricted spatial associations with IG attribution maps. Descriptor summary feature classification further showed that foreground-level descriptor statistics contained class-discriminative visual information. The analysis organizes CNN attribution patterns within reproducible image-derived visual contexts in medicinal plant seed image classification.

Keywords: Medicinal plant seeds; Seed image classification; CNN attribution; Integrated Gradients; Descriptor contextualization; Foreground restriction

29 **1. Introduction**

30 Image-based analysis provides a non-destructive and reproducible way to record visible
31 variation in plant materials [1–3]. Plant organs and materials such as whole plants, leaves, fruits,
32 and seeds can be imaged repeatedly under standardized conditions, providing a scale and
33 consistency that are difficult to achieve by manual observation alone. In particular, RGB
34 images are widely used in plant image analysis and image-based classification because they
35 are relatively easy to acquire and contain information on color, brightness, size, shape, surface
36 texture, and spatial arrangement. These image-based measurements are linked to various
37 research purposes, including species or cultivar discrimination, quality assessment, sample
38 screening, and comparison of morphological variation [1,2]. As image-based models become
39 more widely used, plant image classification increasingly requires performance evaluation
40 alongside analysis of the visual context in which model responses occur.

41 Seed image analysis follows the same general direction [4–10]. Seeds are plant materials
42 directly involved in propagation, genetic resource conservation, quality assessment, and
43 species identification. Image-based analysis enables repeated extraction of seed characteristics
44 such as size, shape, color, texture, and contour. Previous seed image analysis studies have used
45 size- and shape-related traits, including area, length, aspect ratio, and circularity, as well as
46 visual traits such as surface color, brightness, texture, and contour for species discrimination
47 or quality assessment [4–8]. Under imaging conditions in which a single seed is separated from
48 the background, color distribution, brightness variation, local texture, edge structure, and
49 contour within the seed foreground become the main visual units for downstream analysis.

50 Medicinal plant seeds are a relevant domain for this type of analysis because source
51 identification and quality control are important in medicinal plant materials [11–13]. Closely
52 related species and potential substitutes may be considered together in practical contexts, and

visually similar seed samples may differ only in subtle combinations of visible traits. In such cases, class differences are difficult to summarize using a single visual property. Instead, several image properties, such as seed coat color, brightness, local texture, contour, size, and shape, may appear as overlapping spatial patterns. Medicinal seed image classification therefore requires analysis of class label prediction together with the foreground spatial patterns associated with model responses.

Deep learning-based image classification has been increasingly applied to seed image analysis [10,14]. Convolutional neural networks (CNNs) can learn local image patterns together with spatial structure, making them suitable for seed image datasets in which multiple visual properties occur simultaneously. CNN-based models have shown high classification performance across plant image classification tasks and can serve as practical classifiers when the foreground object is clearly defined [14–17]. Accuracy-based evaluation mainly summarizes whether a model separates classes, while the visual patterns within the seed foreground that are associated with prediction-related responses require additional analysis. This makes attribution-based analysis useful for examining the spatial distribution of CNN responses in seed images.

Attribution methods have been used to examine prediction-related spatial responses in CNN classifiers [18–21]. These methods estimate how locations or regions in an input image are related to a specific model output and present the result as an attribution map. Integrated Gradients (IG), for example, calculates pixel- or region-level attribution by accumulating gradients of the prediction score along a path from a baseline input to the actual input [18]. Attribution maps are useful for examining prediction-related spatial responses [18–21]. However, heatmaps alone do not specify the image properties represented by high-response regions. Even within the same foreground region, attribution responses may overlap with color

differences, brightness variation, local texture, edge responses, contour-related structures, or local image patterns in which these properties appear together.

To address this interpretive gap, this study compares attribution maps with predefined image-derived descriptor maps to place attribution responses within an image-derived visual context. The descriptor maps in this study are image-derived reference maps calculated from the original RGB image and the seed foreground mask. They represent low-level visual properties within the seed foreground using consistent criteria. The descriptors are organized into color and intensity, spatial frequency, texture, and edge- and shape-related categories, allowing foreground image responses to be compared from multiple visual perspectives [22–27]. This configuration provides a basis for examining which predefined descriptor responses are spatially co-distributed with IG attribution responses within the seed object. The main analytical unit of this study is therefore the foreground-restricted spatial association between IG attribution patterns and predefined descriptor responses.

Foreground-restricted spatial association quantifies the degree to which attribution maps and descriptor responses are co-distributed within the seed object. Separately, descriptor-derived foreground information can be summarized and evaluated for class-level separability. For this purpose, this study extracts foreground-level summary statistics from descriptor maps and uses them as inputs to a tabular classifier in descriptor summary feature classification. Spatial association analysis addresses attribution-descriptor co-distribution within the seed foreground, whereas descriptor summary feature classification evaluates whether descriptor-derived summary information contains class-discriminative visual information. These two analyses provide complementary axes for contextualizing attribution patterns: spatial co-distribution and descriptor-level class separability.

This study presents a foreground-restricted analytical workflow for contextualizing CNN

101 attribution maps in RGB image classification of five medicinal plant seed species. Seed objects
102 and foreground masks were aligned through foreground-centered preprocessing, and zero-
103 baseline absolute IG maps generated from ConvNeXt-Small were compared with predefined
104 image-derived descriptor maps. Descriptor summary feature classification was also performed
105 to evaluate whether descriptor-derived foreground-level information contained class-
106 discriminative visual information. The proposed workflow provides a methodological basis for
107 organizing CNN attribution patterns in medicinal seed image classification using two analytical
108 axes: foreground-restricted spatial association and descriptor summary feature classification.

109

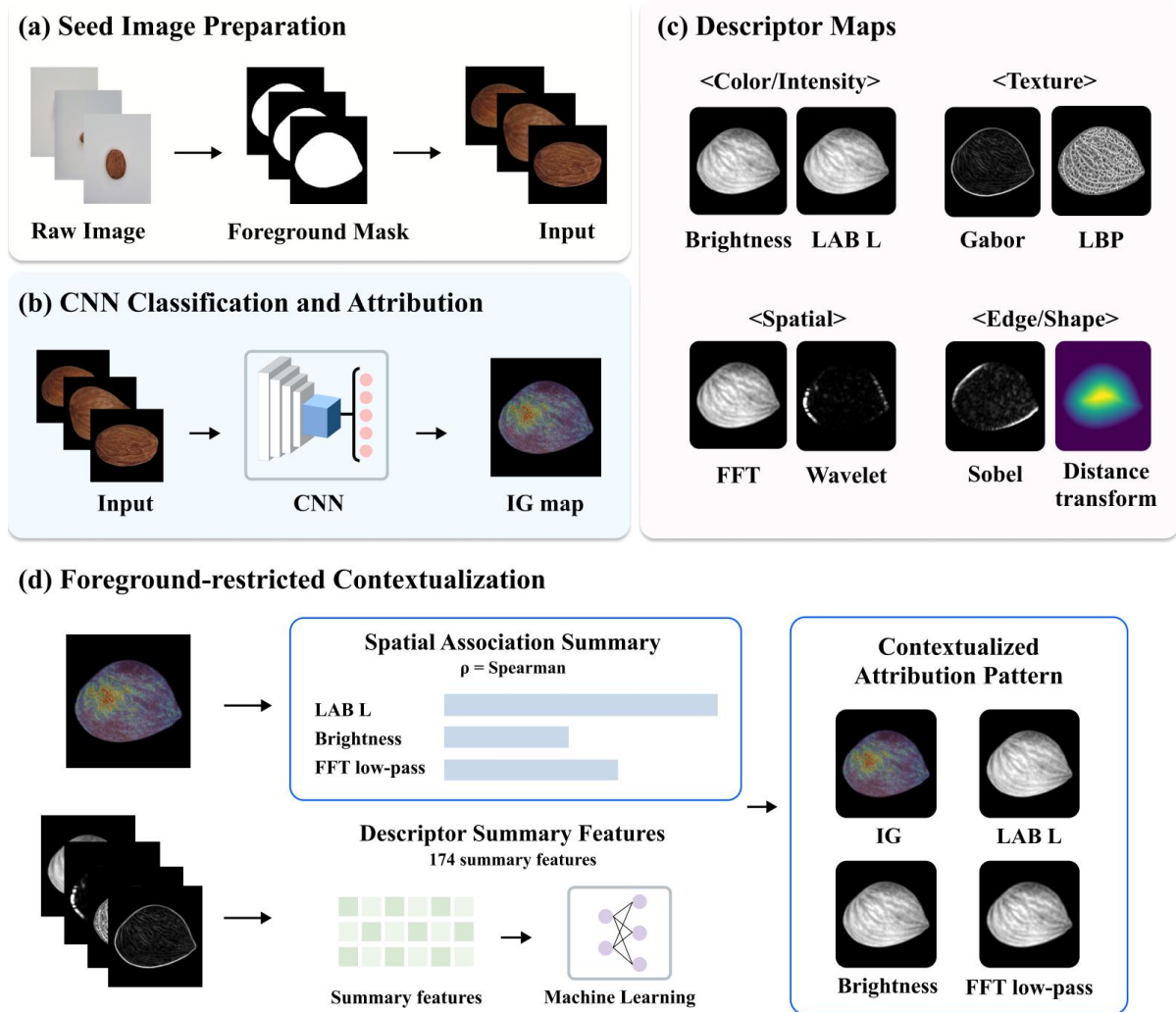


Fig. 1 Foreground-restricted attribution contextualization workflow for medicinal plant seed image classification. (a) Seed images were preprocessed to generate foreground masks and foreground-centered model inputs. (b) The preprocessed seed images were classified using a CNN classifier, and zero-baseline absolute Integrated Gradients (IG) maps were generated from the primary classifier. (c) Predefined image-derived descriptor maps were generated from the same seed images to represent broad visual properties, including color/intensity, spatial frequency, texture, and edge/shape-related responses. (d) The IG attribution maps and descriptor maps were compared within the seed foreground using Spearman spatial association, and foreground-restricted descriptor summary features were used for descriptor summary classification. The workflow contextualizes attribution patterns within predefined seed visual descriptor responses.

2. Materials and methods

2.1 Plant materials, image dataset, and preprocessing

Medicinal plant seed samples were obtained from the National Institute for Korean Medicine

Development (NIKOM, Republic of Korea). The sample identities were verified through organoleptic inspection by nine qualified experts commissioned by the Ministry of Food and Drug Safety (MFDS, Republic of Korea). RGB images were acquired in a PULUZ photo light chamber equipped with manufacturer-specified 25-lumen LED illumination and a white background. Images were captured using an iPhone 8 (Apple Inc., Cupertino, CA, USA) and a Samsung Galaxy Note 10 (Samsung Electronics, Suwon, Republic of Korea) at a resolution of 4032×3024 pixels. In this study, an RGB image dataset was constructed for five medicinal plant seed classes. The target classes consisted of *Armeniaca vulgaris* (ARSE), *Armeniaca sibirica* (ARSS), *Prunus japonica* (PJNA), *Prunus davidiana* (PRDA), and *Prunus persica* (PRPE). The full dataset included 1,124 RGB images, with 205 images for ARSE, 223 for ARSS, 228 for PJNA, 228 for PRDA, and 240 for PRPE. The class composition and train-test split are presented in Table 1.

Table 1. Dataset composition and train and test split

Class code	Species	Total images	Train images	Test images
ARSE	<i>Armeniaca vulgaris</i>	205	164	41
ARSS	<i>Armeniaca sibirica</i>	223	179	44
PJNA	<i>Prunus japonica</i>	228	182	46
PRDA	<i>Prunus davidiana</i>	228	182	46
PRPE	<i>Prunus persica</i>	240	192	48
Total	-	1,124	899	225

All RGB images were processed using the same preprocessing pipeline. The preprocessing procedure was designed to standardize CNN input images and to generate foreground-centered seed images and binary foreground masks for subsequent pixel-wise attribution and descriptor analyses. The overall preprocessing workflow is summarized in Fig. 2.

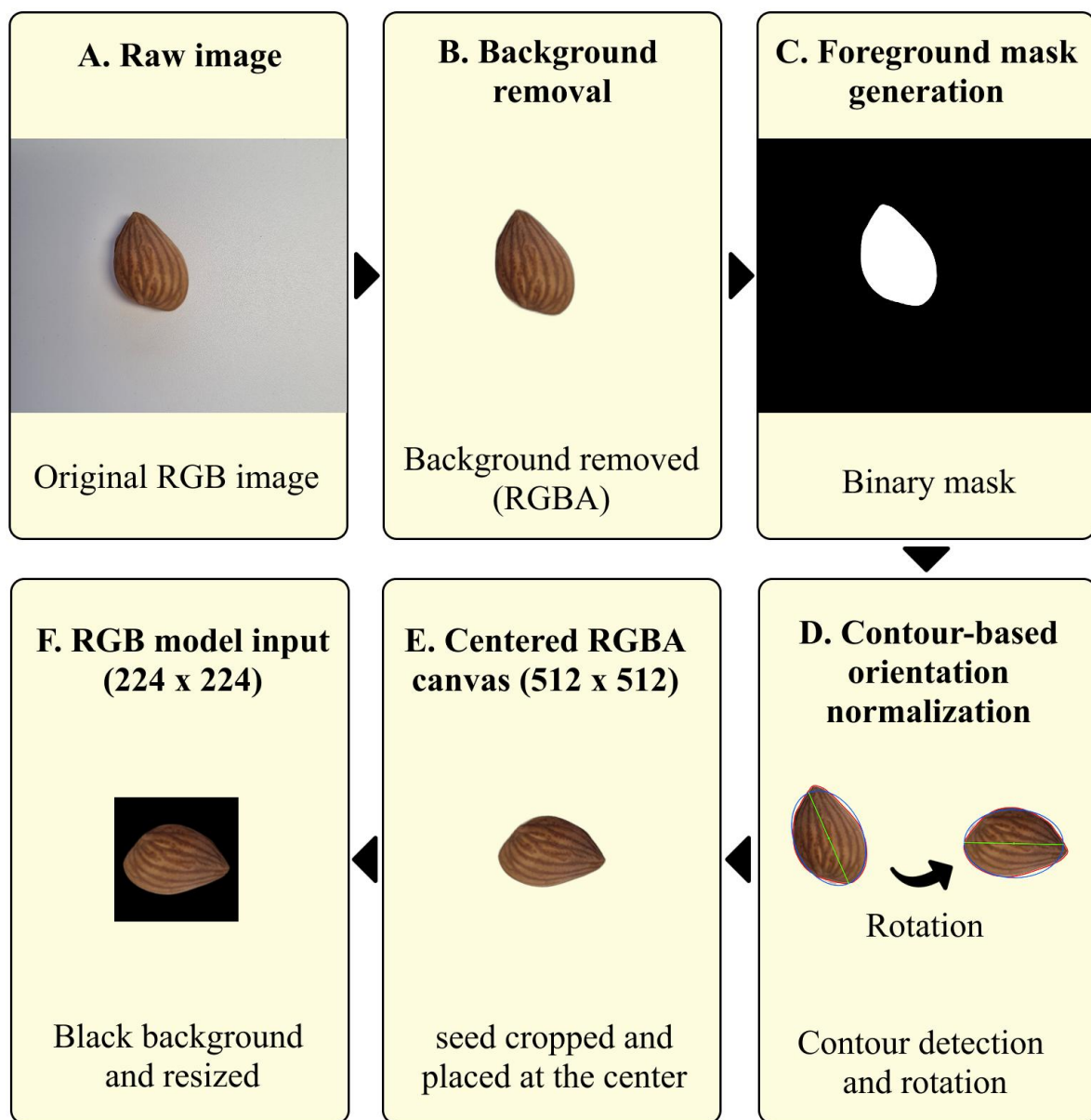


Fig. 2 Image preprocessing workflow for medicinal plant seed classification and foreground mask generation. (a) Raw RGB image. (b) Background removal and RGBA extraction. (c) Binary foreground mask generation. (d) Contour-based orientation normalization. (e) Placement on a centered 512×512 RGBA canvas. (f) Resizing to the 224×224 RGB model input.

First, background removal was performed using the U²-Net-based background removal function implemented in the rembg library to separate the seed object from the background [28]. This step was conducted to reduce background signals and define the seed foreground as the spatial domain for downstream analysis. Alpha matting was not applied to avoid excessive smoothing of the seed boundary and surface texture. The background removal output was saved

as an RGBA image, and a binary foreground mask was generated from the alpha channel.

Contours were detected from the generated foreground mask, and the largest contour was defined as the single seed object. For orientation normalization, ellipse fitting was applied to the foreground contour to estimate object orientation, and both the RGB image and the foreground mask were rotated according to the estimated angle. After rotation, the seed object was cropped based on the bounding box of the largest contour, and the cropped object was aligned so that its major axis followed the horizontal direction. The same geometric transformations were applied to the RGB image and the foreground mask to preserve spatial correspondence between the image and mask.

The aligned seed object was placed at the center of a 512×512 pixel RGBA canvas. The padding inside the canvas was set to 30 pixels, and the maximum size of the seed object was limited to 452 pixels. Object resizing was performed using Lanczos interpolation while preserving the original aspect ratio. The final RGBA image was converted into an RGB image with a black background through alpha compositing and saved in PNG format to avoid texture artifacts caused by JPEG compression. The final binary foreground mask was saved from the alpha channel of the same RGBA image.

The preprocessing results were inspected through manual quality control. Background removal errors, cropping errors, seed boundary loss, residual background pixels, and orientation errors were visually checked. When necessary, horizontal flipping, vertical flipping, or rotation by a specified angle was applied. Manual correction was performed for 24 images in total, including 15 horizontal flips, 4 vertical flips, and 5 rotation corrections. No editing was applied to modify seed surface color, texture, or outer contour, and manual correction was limited to image alignment and mask quality adjustment.

At the model input stage, the preprocessed 512×512 pixel RGB images were resized to 224

× 224 pixels. Normalization was performed using the ImageNet mean and standard deviation to match the ImageNet-pretrained CNN backbones [29]. The final foreground mask was transformed to match the coordinate space of the model input and descriptor maps and was used as the spatial boundary defining the seed foreground in subsequent pixel-wise attribution analysis and descriptor map comparison. Accordingly, downstream analyses compared attribution responses and descriptor responses within the seed object while excluding background pixels.

2.2 Dataset partitioning and CNN classifier training

The preprocessed dataset was divided into a training set and an independent test set at an 8:2 ratio using stratified sampling to preserve the class distribution. Among the 1,124 images, 899 images were assigned to the training set and 225 images to the test set. The random seed was fixed at 42. The test set was not used for model training, early stopping, checkpoint selection, or hyperparameter tuning. It was used only for final model evaluation and as the image set for subsequent attribution analysis.

To evaluate the model training settings and performance stability, stratified 5-fold cross-validation was performed within the training set. In each fold, validation loss was calculated using the corresponding validation split, and the model checkpoint from the epoch with the lowest validation loss was saved. Early stopping was applied when validation loss did not improve for 15 consecutive epochs. After cross-validation, 10% of the training set was separated as a stratified validation subset during final model training, and this subset was used only for early stopping and checkpoint selection.

ConvNeXt-Small, ResNet50, and EfficientNet-B0 were compared as CNN backbones for

medicinal seed image classification [30–32]. All three models used ImageNet-pretrained weights, and the final classification head was replaced with a 5-class output layer corresponding to the five seed classes [29]. Training images were processed using resizing, random horizontal flipping, random rotation, color jitter, and ImageNet normalization. Validation and test images were processed using only resizing and ImageNet normalization. Random erasing and random vertical flipping were not used.

Model optimization was performed using the AdamW optimizer with learning rate warm-up and a cosine annealing schedule [33]. The loss function was cross-entropy loss with label smoothing. After backbone comparison, subsequent attribution and descriptor analyses were performed using the final ConvNeXt-Small checkpoint.

Training was conducted with a batch size of 32 for up to 100 epochs. The initial learning rate was set to 4×10^{-4} , the minimum learning rate to 1×10^{-6} , and the weight decay to 0.01. Warm-up was applied for 5 epochs, and the label smoothing value was set to 0.1.

2.3 Attribution map generation

Attribution maps were generated to examine the prediction-related spatial response of the trained CNN classifier within the seed image coordinate space. In this study, Integrated Gradients (IG) was calculated using the ConvNeXt-Small primary classifier, and the zero-baseline absolute IG map was used as the main attribution map [18]. Positive IG, Gaussian-blurred baseline IG, and Grad-CAM were used as supplementary comparisons to examine differences in spatial responses according to attribution settings and methods [18–21].

Integrated Gradients was applied to generate input-level attribution maps for the prediction output of the ConvNeXt-Small primary classifier. For an input image x , a baseline image x' ,

and the predicted class logit F_c , Integrated Gradients was defined as follows:

$$IG_{i(x)} = (x_i - x'_i) \int_0^1 \frac{\partial F_c(x' + \alpha(x - x'))}{\partial x_i} d\alpha \quad (1)$$

where i denotes a pixel-channel location. In this study, x and x' were defined in the raw RGB input space. The input image was represented as an RGB tensor in the range $[0, 1]$, and the zero baseline was set as a black RGB image in which all pixel values were 0. The Gaussian-blurred baseline was generated by applying Gaussian blur to the corresponding input image, with the Gaussian blur sigma set to 20. Each interpolated image was normalized using the ImageNet mean and standard deviation during model forward propagation and then passed to the ConvNeXt-Small classifier. The integral in Eq. 1 was approximated using 100 interpolation steps.

Signed attribution values for each RGB channel were saved for each sample. The absolute attribution map was calculated by summing the attribution magnitude across RGB channels, and the positive attribution map was generated by summing the RGB channel attributions and then applying ReLU. In the main analysis, the absolute attribution map obtained from the zero baseline was used as the main IG map. This map served as the reference map for comparing the overall spatial magnitude of prediction-related attribution responses with predefined descriptor maps. The positive attribution map and Gaussian-blurred baseline map were used in supplementary analyses. No smoothing, percentile clipping, or colormap transformation was applied to IG maps used for quantitative analysis. Visualization overlay maps were generated after removing attribution values outside the foreground mask and were displayed on the original RGB image.

Grad-CAM was calculated for the final checkpoints of ConvNeXt-Small, ResNet50, and EfficientNet-B0 as a complementary layer-level attribution visualization [19]. The predicted class logit of each model was used as the target score, and the target layer was set to a layer with spatial activations before the classifier head. The target layers were stages.2.blocks.26 for ConvNeXt-Small, layer4.2 for ResNet50, and blocks.4 for EfficientNet-B0. Forward activations and backward gradients were stored at each target layer, and the spatial mean of the channel-wise gradients was used as the weight to calculate the Grad-CAM map. The calculated CAM was processed with ReLU and then upsampled to 224×224 pixels using bilinear interpolation.

The selected-layer Grad-CAM of ConvNeXt-Small was used together with IG for representative attribution visualization and was also included in supplementary spatial association analysis with predefined descriptor maps. Grad-CAM maps from ResNet50 and EfficientNet-B0 were generated as complementary backbone-level visualizations and were not included in the main descriptor association analysis. Therefore, the main descriptor contextualization in this study was performed using the ConvNeXt-Small zero-baseline absolute IG map.

2.4 Predefined image-derived visual descriptor maps

To compare the spatial pattern of IG attribution maps within an image-derived visual context, predefined visual descriptor maps were generated for each seed image. In this study, descriptor maps were used as image-derived reference maps that represent low-level visual properties within the seed foreground using consistent computational criteria. This allowed the attribution response to be compared with predefined descriptor responses that were spatially distributed within the seed foreground.

The descriptor maps were organized into four categories: color and intensity, spatial frequency, texture, and edge and shape. These categories were used to cover broad visual properties that can be observed within medicinal plant seed foregrounds. Each descriptor map was calculated from the preprocessed RGB image, grayscale image, or foreground mask. All descriptor maps were generated or transformed in the same 224×224 pixel coordinate system as the attribution maps, allowing direct comparison with the attribution maps in subsequent pixel-wise spatial association analysis.

Color and intensity descriptors included Brightness, LAB L, LAB Chroma, and HSV Saturation. Brightness was calculated from RGB intensity values, and LAB L was extracted from the lightness channel of the LAB color space. LAB Chroma was calculated as the chromatic magnitude using the LAB a and b channels, and HSV Saturation was generated from the saturation channel of the HSV color space. These maps were used as reference maps representing intensity-, lightness-, chroma-, and saturation-related image variation within the seed foreground.

Spatial frequency descriptors were calculated from grayscale intensity images and included FFT low-pass, FFT high-pass, and wavelet detail maps [22,23]. The FFT low-pass map was generated to represent coarse-scale intensity variation, whereas the FFT high-pass map was generated to represent fine-scale spatial variation. Wavelet detail maps were used to represent scale- and direction-specific image responses. These descriptors were used as reference maps for comparing image variation occurring at different spatial scales.

Texture descriptors included Gabor response maps and local binary pattern maps [24,25]. Gabor response maps were calculated using predefined frequency and orientation combinations, and each combination was saved as a separate descriptor map. Local binary patterns were calculated to summarize local intensity relationships among neighboring pixels. These maps

were used to represent local and oriented texture-like image patterns within the seed foreground.

Edge and shape descriptors included Sobel edge response, Laplacian-based local variation, distance transform, and Fourier descriptor maps [26,27]. Sobel edge response was calculated from the gradient magnitude of the grayscale image, and Laplacian-based local variation was generated from the Laplacian response. The distance transform was calculated to represent the distance of each pixel inside the foreground mask from the seed boundary. Fourier descriptor maps were generated from the foreground contour to provide a contour-derived shape representation.

All descriptor maps were converted to the range [0, 1] using map-wise min-max normalization. This step was applied to adjust differences in response scale among descriptor types and to compare spatial patterns within the foreground. In the subsequent spatial association analysis, only pixels inside the foreground mask were used, and descriptor-attribution comparisons were performed within the seed object while excluding background pixels. The category, represented visual property, and description of each descriptor map are summarized in Table 2. The full descriptor inventory is provided in Supplementary Table S1.

Table 2. Predefined image-derived visual descriptor maps used in this study.

Category	Descriptor map	Visual property represented	Description
Color and intensity	Brightness	Overall image intensity	Distribution of RGB intensity values within the seed foreground
Color and intensity	LAB L	Perceptual lightness	Lightness-channel variation separated from chromatic components
Color and intensity	LAB Chroma	Chromatic magnitude	Spatial distribution of chromatic magnitude calculated from LAB a and b channels
Color and intensity	HSV Saturation	Color saturation	Distribution of saturation

			values in HSV color space
Spatial frequency	FFT low-pass	Low spatial-frequency response	Coarse-scale intensity variation in the grayscale image
Spatial frequency	FFT high-pass	High spatial-frequency response	Fine-scale intensity variation and local spatial changes
Spatial frequency	Wavelet detail maps	Scale- and direction-specific spatial response	Multi-scale directional detail responses in the grayscale image
Texture	Gabor response	Orientation- and frequency-selective response	Local oriented image responses computed at predefined frequencies and orientations
Texture	Local binary pattern	Local intensity relationship	Local binary relationships among neighboring pixel intensities
Edge and shape	Sobel edge response	First-order edge response	Local gradient magnitude in the grayscale image
Edge and shape	Laplacian-based local variation	Second-order intensity response	Local second-order intensity variation in the grayscale image
Edge and shape	Distance transform	Mask-derived interior distance	Distance of each foreground pixel from the seed boundary
Edge and shape	Fourier descriptor	Contour-derived shape representation	Frequency-domain representation derived from the foreground contour

310

311 2.5 Attribution–descriptor spatial association analysis

312 Spearman rank correlation was calculated to quantify the spatial association between
313 attribution maps and predefined descriptor maps [34]. This analysis was performed to evaluate
314 the degree to which CNN attribution responses and descriptor responses were co-distributed
315 within the seed foreground. The main analysis was conducted using the zero-baseline absolute
316 IG map generated from ConvNeXt-Small, which was used as the main attribution map
317 representing the overall spatial magnitude of prediction-related responses. For supplementary

comparisons, the zero-baseline positive IG map and the selected-layer Grad-CAM map from ConvNeXt-Small were used. All attribution maps and descriptor maps were compared after alignment to the same 224×224 pixel coordinate system.

Spatial association analysis was performed using only pixels inside the foreground mask. For each test image, the attribution values within the foreground mask were denoted as (a_F) , and the descriptor responses extracted from the same foreground pixel locations were denoted as (d_F) , where (F) represents the set of pixels inside the foreground mask. The foreground-restricted rank-based spatial association between an attribution map and a descriptor map was calculated as the Spearman correlation coefficient, as shown in Eq. 2.

$$\rho = \text{corr}(\text{rank}(a_F), \text{rank}(d_F)) \quad (2)$$

This procedure was performed independently for each image and each descriptor map. Background pixels were excluded from the correlation calculation, and the rank-based spatial correspondence between the two maps within the foreground was summarized as an image-level correlation coefficient.

Image-level correlation coefficients were used as the basic analytical units and were then summarized as classwise and descriptor-level results. For the classwise summary, image-level correlation coefficients from test images belonging to the same class were collected, and the class-level mean and standard deviation were calculated. For the descriptor-level summary, image-level correlation coefficients obtained from all test images were summarized for each descriptor to compare foreground-restricted spatial associations across the tested descriptor maps.

Class-average map analysis was also performed to examine class-level spatial patterns. For each seed class, attribution maps from test images belonging to the same class were averaged to generate a class-average attribution map. Descriptor maps were averaged in the same manner to generate class-average descriptor maps. The Spearman correlation coefficient between the class-average attribution map and each class-average descriptor map was then calculated within the foreground region. This analysis was used as a supplementary analysis to summarize class-level spatial patterns.

Statistical significance was assessed using image-level correlation coefficients for each descriptor. A one-sample t-test was applied to evaluate whether the distribution of image-level correlation coefficients for each descriptor differed from zero. Benjamini–Hochberg false discovery rate correction was applied to control false positives arising from multiple comparisons, and the significance level was set at 0.05 [35]. The main text focuses on major descriptors, while image-level statistics for the full descriptor set are provided in Supplementary Table S2. Supplementary attribution comparisons based on positive IG and ConvNeXt-Small Grad-CAM are provided in Supplementary Tables S3 and S4, respectively.

2.6 Descriptor summary feature classification

Descriptor summary feature classification was performed as a supplementary analysis to evaluate foreground-restricted summary features extracted from predefined descriptor maps as inputs to tabular classifiers. This analysis used the same train/test split as the CNN classification experiment, and the independent test set was used only for final performance evaluation.

A total of 29 valid individual descriptor maps were used in this analysis. A valid descriptor

map was defined as a map for which both the descriptor NPY file and the foreground mask were available for all 1,124 samples, with no NaN or Inf values in foreground pixel values and no empty foreground cases. No descriptor maps were excluded, and all valid descriptor maps were available for all 1,124 samples.

For each image and descriptor map, the foreground mask was resized to match the spatial size of the descriptor map using nearest-neighbor interpolation. Pixels with mask values greater than 127 were defined as the seed foreground, and only these foreground pixel values were used for summary feature calculation. Background pixels were excluded from feature extraction. Descriptor maps were analyzed after normalization to the range [0, 1].

Six foreground-restricted summary statistics were calculated from each descriptor map: mean, standard deviation, median, 90th percentile, top 10% mean, and interquartile range. Standard deviation was calculated as the population standard deviation. The 90th percentile was calculated from foreground pixel values, and the interquartile range was calculated as the difference between the 75th and 25th percentiles. The top 10% mean was calculated by selecting the highest 10% of foreground pixel values and averaging them. Each image was therefore converted into a 174-dimensional numeric feature vector consisting of 29 descriptor maps $\times$ 6 summary statistics.

Descriptor summary features were converted into a wide-format tabular matrix and used as classifier inputs. Feature columns were named by combining the descriptor key and statistic name. Class labels were encoded in the order ARSE, ARSS, PJNA, PRDA, and PRPE.

Logistic regression was used as the primary fixed classifier for tabular classification, and radial basis function support vector machine (RBF SVM) was used as a secondary comparison [36]. Both classifiers were implemented in scikit-learn pipelines with StandardScaler. StandardScaler was fitted only on the training set, and the test set was transformed using the

scaling parameters estimated from the training set [37]. Logistic regression was configured with the lbfgs solver, $C = 1.0$, $\text{max_iter} = 5000$, and $\text{random_state} = 42$. RBF SVM was configured with an RBF kernel, $C = 1.0$, $\text{gamma} = \text{scale}$, $\text{probability} = \text{false}$, and $\text{random_state} = 42$. Both classifiers were evaluated under fixed settings, with no additional hyperparameter tuning.

Performance was also compared across feature sets. In addition to the all-valid descriptor set containing all 29 valid descriptor maps, the following subsets were evaluated: selected 10-map set, Top 3, Top 5, and Top 7 IG-associated descriptor subsets, low or negative IG-associated subset, color and intensity subset, spatial frequency subset, texture-only subset, texture and wavelet-detail subset, and edge and shape subset. For each subset, the same six summary statistics were calculated for each included descriptor map to construct the feature vector. The descriptor maps included in each subset and the corresponding number of features are summarized in Supplementary Table S5.

Classification performance was evaluated on the independent test set using accuracy, macro precision, macro recall, and macro F1-score. Logistic regression results are presented as the primary analysis, and RBF SVM results are presented as a secondary comparison to examine differences associated with classifier choice.

2.7 Software and reproducibility

All final experiments and analyses were conducted in a Python 3.11.15 environment. PyTorch 2.12.0.dev20260316+cu128 and torchvision 0.26.0.dev20260316+cu128 were used for deep learning model training and gradient-based attribution calculation. The ConvNeXt-Small, ResNet50, and EfficientNet-B0 model architectures were implemented using timm 0.6.7, and

411 IG attribution was calculated using Captum 0.8.0.

412 Image preprocessing was performed using rembg 2.0.75, Pillow 12.2.0, and OpenCV 4.10.0.

413 Descriptor maps were calculated using OpenCV 4.10.0, scikit-image 0.26.0, NumPy 2.4.4,

414 SciPy 1.17.1, and PyWavelets 1.8.0 [38–40]. Descriptor summary feature table construction

415 and tabular data processing were performed using pandas 2.3.3. Statistical analysis and

416 descriptor summary feature classification were performed using SciPy 1.17.1, statsmodels

417 0.14.6, and scikit-learn 1.8.0 [37,40]. Figures 1 and 2 were designed and assembled by the

418 authors, whereas Figures 3 and 4 were generated using Python, matplotlib 3.10.8, and

419 seaborn 0.13.2.

420 Experiments were performed on a workstation equipped with one NVIDIA RTX PRO 6000

421 Blackwell Max-Q Workstation Edition GPU. The execution environment used NVIDIA

422 driver 580.159.03, PyTorch CUDA 12.8, and cuDNN 9.19.0. The CUDA version displayed

423 by nvidia-smi was 13.0.

424 To support reproducibility, the random seed was fixed at 42. During CNN training, the

425 Python random seed, NumPy random seed, PyTorch random seed, and CUDA random seed

426 were all set. The cuDNN deterministic option was enabled, and the benchmark option was

427 disabled. Train/test split files, model configuration files, descriptor map inventories,

428 attribution outputs, descriptor summary feature tables, statistical result tables, and figure-

429 generation scripts were stored and managed within the same repository structure.

430 During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist with

431 manuscript drafting and code development. All AI-assisted content and code were reviewed,

432 edited, and verified by the authors, who take full responsibility for the integrity and accuracy

433 of the work. No generative AI image-generation tools were used.

3. Results

3.1 CNN backbone performance comparison

To examine whether seed class discrimination was stable under the controlled imaging and preprocessing conditions, classification performance was compared among ConvNeXt-Small, ResNet50, and EfficientNet-B0. All three backbones showed near-ceiling performance in both stratified 5-fold cross-validation and the independent test set. ResNet50 achieved a cross-validation accuracy of 0.9989 ± 0.0025 and a macro F1-score of 0.9989 ± 0.0024 , with both accuracy and macro F1-score reaching 1.0000 on the independent test set. EfficientNet-B0 also achieved a cross-validation accuracy of 0.9989 ± 0.0025 and a macro F1-score of 0.9989 ± 0.0024 , with both accuracy and macro F1-score reaching 1.0000 on the independent test set. ConvNeXt-Small achieved a cross-validation accuracy of 0.9989 ± 0.0025 and a macro F1-score of 0.9988 ± 0.0026 , and showed an accuracy of 0.9956 and a macro F1-score of 0.9957 on the independent test set (Table 3).

These results indicate that all three CNN backbones provided stable classification performance under the dataset and preprocessing conditions used in this study. Subsequent attribution–descriptor contextualization analyses were performed using the final ConvNeXt-Small checkpoint. ConvNeXt-Small showed near-ceiling performance on the independent test set, and its prediction outputs were used to generate zero-baseline absolute IG maps and selected-layer Grad-CAM maps. Grad-CAM maps from ResNet50 and EfficientNet-B0 were used as backbone-level complementary visualizations, while the main foreground-restricted descriptor association analysis was conducted using the zero-baseline absolute IG maps generated from ConvNeXt-Small (Table 3).

Table 3. CNN backbone performance comparison.

Model	CV accuracy	CV macro F1	Test accuracy	Test macro F1
ResNet50	0.9989 ± 0.0025	0.9989 ± 0.0024	1.0000	1.0000
EfficientNet-B0	0.9989 ± 0.0025	0.9989 ± 0.0024	1.0000	1.0000
ConvNeXt-Small	0.9989 ± 0.0025	0.9988 ± 0.0026	0.9956	0.9957

The mean cross-validation accuracy was 0.9989 for all three models, and the macro F1-score ranged from 0.9988 to 0.9989. Taken together with the independent test set results, the cross-validation results showed that all three CNN backbones achieved near-ceiling performance under the dataset and preprocessing conditions used in this study. The confusion matrix and the single misclassified ConvNeXt-Small test sample are shown in Supplementary Fig. S3.

3.2 Attribution maps from the primary classifier

Attribution maps were generated for 225 seed images in the independent test set using the final ConvNeXt-Small classifier. In the main analysis, the zero-baseline absolute IG map was used as the main attribution representation. This map was used to represent the overall spatial magnitude of prediction-related responses and was applied in subsequent foreground-restricted spatial association analysis with predefined descriptor maps.

To examine differences in spatial patterns according to the attribution setting, absolute IG maps generated using a Gaussian-blurred baseline were compared with zero-baseline absolute IG maps. The foreground-restricted Spearman correlation between the two IG maps was 0.4395 ± 0.0981 . Based on this result, the zero-baseline absolute IG map was fixed as the

reference for the main analysis, while the Gaussian-blurred baseline IG map was presented as a supplementary comparison. Class-average zero-baseline and Gaussian-blurred baseline IG maps and per-sample spatial agreement are shown in Supplementary Fig. S2.

Selected-layer Grad-CAM was generated from the final ConvNeXt-Small classifier as a complementary layer-level attribution visualization. The foreground coverage of the top 20% Grad-CAM response region was 0.9739 ± 0.0566 . This result indicates that most high-response regions in Grad-CAM were located within the seed foreground. Positive IG and Grad-CAM results are presented together in the supplementary attribution comparison.

Figure 3 shows zero-baseline absolute IG overlays and selected-layer Grad-CAM overlays for representative test images. In the subsequent analyses, the ConvNeXt-Small zero-baseline absolute IG map was used as the main attribution map to calculate foreground-restricted spatial association with predefined descriptor maps. Additional representative attribution and descriptor examples are provided in Supplementary Fig. S1.

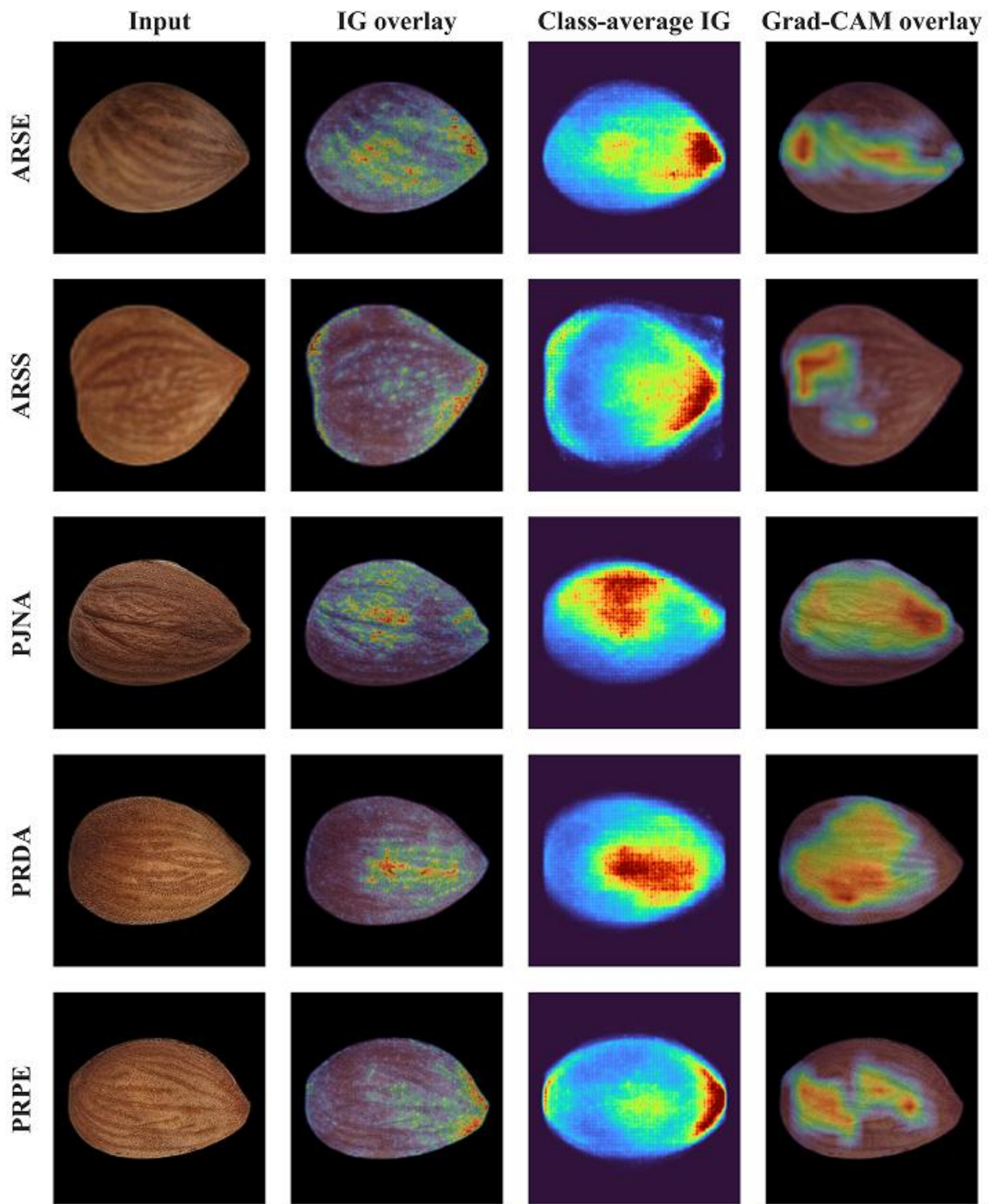


Fig. 3 Attribution maps from the primary classifier. Representative input images, zero-baseline absolute IG overlays, class-average IG maps, and selected-layer Grad-CAM overlays are shown for each seed class. IG was used as the main input-level attribution map, whereas Grad-CAM was shown as a complementary selected-layer visualization.

3.3 Spatial association between IG attribution and predefined descriptor maps

Foreground-restricted spatial association between the zero-baseline absolute IG attribution maps from ConvNeXt-Small and the predefined descriptor maps was evaluated using Spearman correlation. Among the tested descriptor maps, several descriptors in the color and intensity category and the spatial frequency category showed relatively high positive associations. LAB L showed the highest mean Spearman correlation coefficient, with a value of 0.4857 ± 0.1586 , and Brightness showed a similar value of 0.4849 ± 0.1569 . In the spatial frequency category, FFT low-pass also showed a high positive association, with a mean Spearman correlation coefficient of 0.4804 ± 0.1695 (Fig. 4).

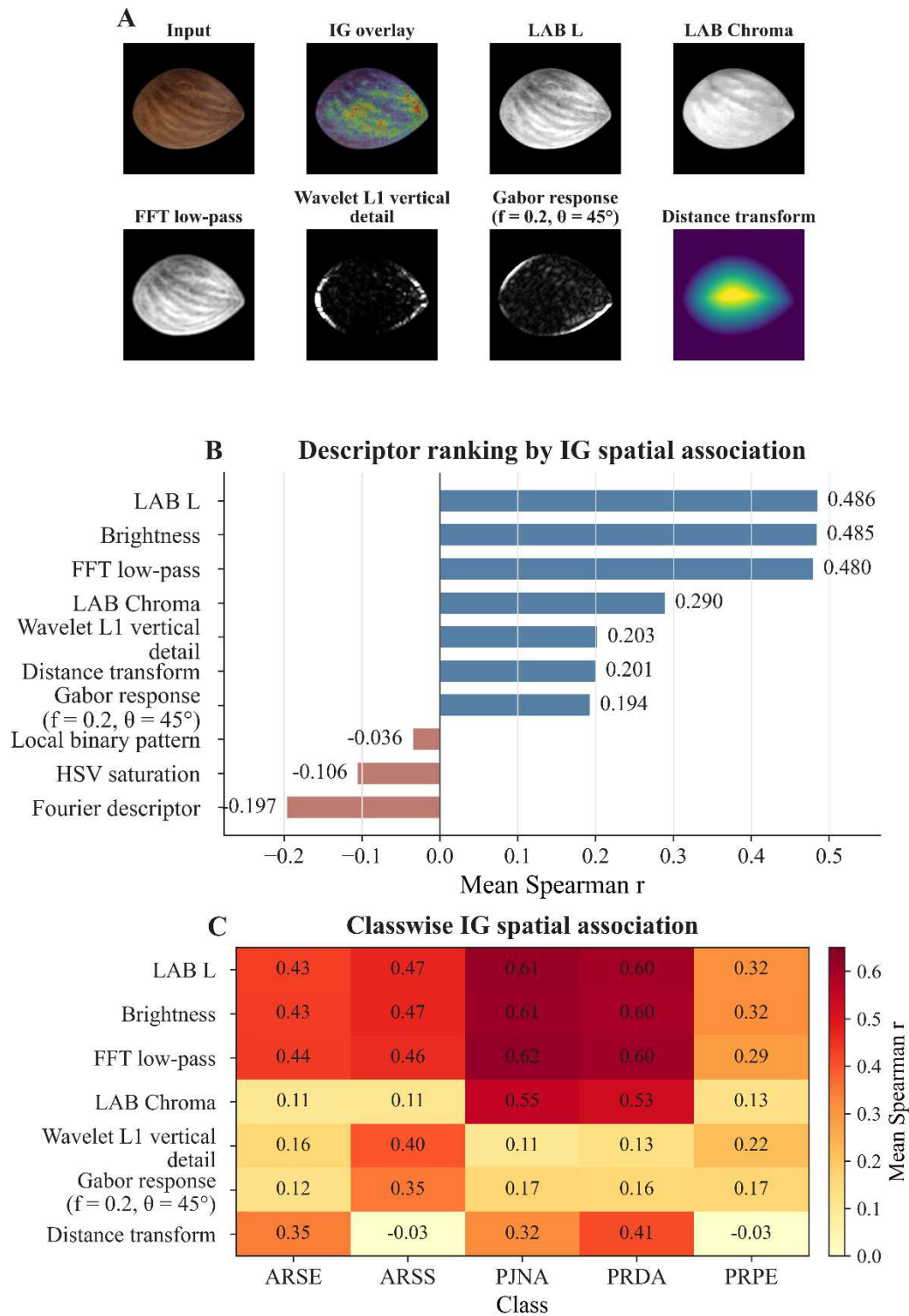


Fig. 4 Spatial association between IG attribution and predefined image-derived descriptor maps. (A) Representative input image, IG overlay, and selected descriptor maps generated from the same seed image. (B) Mean foreground-restricted Spearman correlation between zero-baseline absolute IG maps and selected descriptor maps across the test set. (C) Classwise mean Spearman correlation for the selected descriptors.

After the highest-ranked descriptors, positive associations were observed for several color-,

spatial detail-, texture-, and distance-related descriptors. LAB Chroma showed a mean Spearman correlation coefficient of 0.2898 ± 0.2568 , while Wavelet L1 vertical detail and Distance transform showed values of 0.2032 ± 0.1324 and 0.2007 ± 0.2312 , respectively. In the texture category, the Gabor response at frequency 0.2 and orientation 45 degrees showed a mean Spearman correlation coefficient of 0.1935 ± 0.1233 .

Some descriptors showed low or negative associations. The Fourier descriptor showed the lowest mean Spearman correlation coefficient, with a value of -0.1973 ± 0.2322 , and HSV Saturation also showed a negative association, with a value of -0.1064 ± 0.1675 . LBP showed a low association close to zero, with a value of -0.0356 ± 0.0404 . Image-level statistics and classwise results for the full descriptor set are provided in Supplementary Table S2. The full classwise descriptor-association pattern is shown in Supplementary Fig. S4.

Overall, the zero-baseline absolute IG attribution maps from ConvNeXt-Small showed the highest positive foreground-restricted spatial associations with LAB L, Brightness, and FFT low-pass among the tested descriptor maps. These results quantitatively show that the main IG attribution response tended to be co-distributed with lightness-related and low-frequency descriptor responses within the seed foreground.

3.4 Descriptor summary features retain class-discriminative visual information

Descriptor summary feature classification was performed to evaluate whether foreground-restricted summary statistics extracted from predefined descriptor maps contained information for seed class discrimination. A total of 29 valid individual descriptor maps were used in this analysis, and six summary statistics were calculated from the pixel values inside the foreground mask for each descriptor map. As a result, each image was converted into a

174-dimensional numeric feature vector. This analysis was conducted as a supplementary analysis, separate from the preceding IG–descriptor spatial association analysis, to examine the extent to which descriptor maps captured class-discriminative visual variation within the seed foreground.

When logistic regression was used as the primary fixed classifier, the feature set containing all 29 valid descriptor maps achieved an accuracy of 0.9778 and a macro F1-score of 0.9785 on the independent test set (Table 4).

Table 4. Classification performance using foreground-restricted descriptor summary features.

Feature set	Model	Test accuracy	Macro F1
All valid descriptor maps	Logistic regression	0.9778	0.9785
Selected 10-map set	Logistic regression	0.9511	0.9522
Top 5 IG-associated subset	Logistic regression	0.9111	0.9127
Top 7 IG-associated subset	Logistic regression	0.9111	0.9124
Texture and wavelet-detail subset	Logistic regression	0.9644	0.9657

The selected 10-map set also maintained performance close to that of the full descriptor set, with an accuracy of 0.9511 and a macro F1-score of 0.9522. Among subsets composed of descriptors with high IG association, the Top 3 subset showed an accuracy of 0.8044 and a macro F1-score of 0.8045, while the Top 5 subset increased to 0.9111 and 0.9127, respectively. The Top 7 subset showed an accuracy of 0.9111 and a macro F1-score of 0.9124,

which was similar to the Top 5 subset.

The texture and wavelet-detail subset also achieved an accuracy of 0.9644 and a macro F1-score of 0.9657 with logistic regression. This result indicates that descriptor-level visual variation contributing to seed class discrimination was also present outside the descriptor subset with the highest IG spatial associations. The low or negative IG-associated subset showed an accuracy of 0.8622 and a macro F1-score of 0.8668, further showing that spatial association ranking and descriptor summary classification performance provide information from different analytical perspectives.

In the secondary comparison using RBF SVM, the full set of 29 valid descriptor maps achieved an accuracy of 0.9733 and a macro F1-score of 0.9742. The selected 10-map set showed an accuracy of 0.9244 and a macro F1-score of 0.9257, and the Top 7 IG-associated subset showed an accuracy of 0.9244 and a macro F1-score of 0.9246. The RBF SVM results also showed that the full descriptor set maintained high classification performance. The full descriptor summary classification results for all feature subsets and both classifiers are provided in Supplementary Table S5.

These results show that foreground-restricted summary features extracted from predefined descriptor maps contain class-discriminative visual information for medicinal seed class discrimination.

3.5 Supplementary comparison using positive IG and Grad-CAM

To compare with the main descriptor association results based on the zero-baseline absolute IG map, foreground-restricted descriptor associations were also calculated using positive IG maps and selected-layer Grad-CAM maps from ConvNeXt-Small. This analysis was

performed as a supplementary comparison to examine the extent to which descriptor association patterns were maintained when the attribution representation was changed.

In the positive IG maps, several top-ranked descriptors overlapped with those from the zero-baseline absolute IG maps. Based on the Top 10 descriptors, positive IG and absolute IG shared seven descriptors. Brightness, LAB L, and FFT low-pass were also included among the top-ranked descriptors for positive IG. Descriptor association coefficients for positive IG tended to be lower overall than those for absolute IG, and the rankings of some wavelet- and texture-related descriptors differed. The full results of descriptor association based on positive IG are provided in Supplementary Table S3.

Descriptor association analysis using selected-layer Grad-CAM from ConvNeXt-Small also showed partial overlap with the absolute IG results. Based on the Top 10 descriptors, Grad-CAM and absolute IG shared six descriptors. In Grad-CAM, Distance transform, LAB Chroma, and HSV Saturation showed relatively higher rankings than in absolute IG, while some wavelet detail and Gabor response descriptors showed different ranking patterns depending on the attribution representation. The full results of descriptor association based on Grad-CAM are provided in Supplementary Table S4.

Overall, Brightness, LAB L, and FFT low-pass remained among the top-ranked descriptors in positive IG, whereas some descriptor rankings in Grad-CAM differed from those in absolute IG. In this study, the zero-baseline absolute IG results are presented as the main analysis, and the positive IG and Grad-CAM results are summarized as supplementary comparisons of descriptor association patterns according to the attribution setting.

4. Discussion

4.1 Foreground-restricted descriptor contextualization as a methodological contribution

This study presents a foreground-restricted workflow for contextualizing CNN attribution maps using predefined image-derived descriptor maps in medicinal plant seed classification. In the proposed workflow, seed foreground masks, attribution maps, and descriptor maps were aligned in a shared coordinate system, and attribution–descriptor spatial association was calculated within the seed object after excluding background pixels. This foreground restriction reflects the analytical focus of the study, as visual information relevant to seed image classification is mainly contained in seed coat color, intensity variation, local texture, edge structure, contour, and shape-related variation.

Attribution heatmaps indicate spatial regions associated with CNN predictions, but they do not specify which image properties are represented in those regions. Within the seed foreground, high attribution responses may coincide with lightness variation, chromatic differences, coarse-scale intensity structure, local texture, edge responses, or contour-related patterns. Comparing attribution maps with predefined descriptor maps therefore provides a structured way to place heatmap locations within reproducible image-derived visual contexts.

The descriptor maps used in this study represented color and intensity, spatial frequency, texture, edge, and shape-related responses under consistent computational criteria. The interpretation of these maps was limited to their use as visual references for quantifying foreground co-distribution between attribution responses and descriptor responses. This distinction allowed attribution–descriptor spatial association to be interpreted separately from descriptor-level class separability.

The near-ceiling classification performance of the three CNN backbones under controlled

imaging and preprocessing conditions provided a practical basis for the subsequent attribution analysis. Stable classifier performance supported the generation of prediction-related attribution maps from reliable model outputs. In this study, classification accuracy mainly served to support the downstream analysis, while the methodological focus lay in how attribution patterns could be contextualized within the seed foreground.

This workflow is particularly relevant for image-based analysis of visually similar medicinal plant seeds, related taxa, and potential substitutes. In such cases, class label prediction alone gives limited information about the visual basis of model responses. Foreground-restricted descriptor contextualization summarizes how attribution responses and predefined descriptor responses are spatially co-distributed inside the seed object. The resulting framework provides a methodological basis for structured interpretation of CNN attribution patterns in medicinal seed image classification and related plant image analysis tasks.

4.2 Attribution maps as method-dependent spatial patterns

IG and Grad-CAM generated from ConvNeXt-Small provided attribution patterns at different spatial levels [18,19]. IG accumulates gradient-based attribution between input pixel-channel values and the predicted class logit, producing an input-level attribution pattern across the seed foreground. Grad-CAM is derived from activations and gradients in a selected layer and produces a coarser layer-level response pattern. In this study, the top 20% response region of Grad-CAM was located mostly within the seed foreground, indicating that the selected-layer response was largely aligned with the preprocessed seed object region.

The spatial pattern of the IG attribution map varied according to the baseline setting [18,20,21]. The mean foreground-restricted Spearman correlation between the zero-baseline

absolute IG map and the Gaussian-blurred baseline absolute IG map was 0.4395 ± 0.0981 , showing that the baseline setting affected the spatial distribution of IG responses. Based on this result, the zero-baseline absolute IG map was fixed as the main attribution map for descriptor comparison. Positive IG and Grad-CAM were used as supplementary comparisons to examine how attribution patterns changed across attribution settings and methods.

This method-dependent variability supports the use of a clearly defined attribution reference when comparing CNN attribution maps with predefined descriptor maps. In the present workflow, the zero-baseline absolute IG map served as the primary reference for quantifying foreground-restricted spatial association with descriptor responses. The supplementary attribution maps were used to assess whether similar descriptor association tendencies appeared under alternative attribution settings, without treating any single attribution map as a complete explanation of the model decision.

4.3 Spatial association with predefined image-derived descriptor maps

The zero-baseline absolute IG map showed the highest positive foreground-restricted spatial associations with LAB L, Brightness, and FFT low-pass among the tested descriptor maps. LAB L and Brightness represent lightness and intensity variation within the seed foreground, while FFT low-pass represents coarse-scale intensity structure in the grayscale image [22]. This pattern suggests that the input-level attribution map of ConvNeXt-Small was spatially co-distributed with lightness-related, intensity-related, and low-frequency image responses within the seed foreground. In this sense, predefined descriptor maps served as references for organizing attribution patterns within image-derived visual contexts.

Positive associations were also observed for LAB Chroma, Wavelet L1 vertical detail,

Distance transform, and some Gabor responses [23,24]. These descriptors represent chromatic magnitude, scale-specific spatial detail, interior distance from the boundary, and orientation-selective texture-like responses, respectively. In seed images, intensity, color, local spatial detail, and contour-related responses may vary together within the foreground. The co-distribution of multiple descriptor responses with IG attribution supports interpreting attribution patterns within combinations of image-derived responses.

Fourier descriptor, HSV Saturation, and LBP showed low or negative associations [25,27]. Fourier descriptor is closer to a contour-derived shape representation, whereas LBP summarizes local binary relationships among neighboring pixels. These descriptors differ from the input-level IG map in spatial scale or response type, which may explain their lower association with foreground IG responses. Overall, the results in Section 3.3 provide a quantitative basis for comparing the visual descriptor contexts in which IG attribution maps are spatially distributed within the seed foreground.

4.4 Descriptor summary features as class-discriminative visual context

Descriptor summary feature classification was performed to examine whether predefined descriptor maps contained foreground-level visual information relevant to seed class discrimination. The 174-dimensional summary feature vector extracted from all 29 valid descriptor maps achieved an accuracy of 0.9778 and a macro F1-score of 0.9785 with logistic regression. These values show that summary statistics calculated from descriptor maps captured substantial class-discriminative visual variation in the controlled seed image dataset. The selected 10-map set also showed high classification performance, with an accuracy of 0.9511 and a macro F1-score of 0.9522, suggesting that a compact descriptor set still contained class-discriminative visual information.

Class-discriminative information was also present in the IG-associated descriptor subsets. The Top 5 IG-associated subset showed an accuracy of 0.9111 and a macro F1-score of 0.9127, and the Top 7 subset showed similar performance. The texture and wavelet-detail subset achieved higher performance than the top IG-associated subsets, with an accuracy of 0.9644 and a macro F1-score of 0.9657. This difference indicates that IG spatial association ranking and summary-level class separation capture different aspects of descriptor information.

Descriptor summary feature classification complements the spatial association analysis presented in Section 3.3. Spatial association analysis quantifies the foreground co-distribution between IG attribution responses and descriptor responses. Descriptor summary classification evaluates class-level separability based on foreground descriptor statistics. Keeping these analyses separate helps distinguish the spatial context of attribution maps from the class-discriminative context of descriptor maps.

4.5 Methodological implications and limitations

The workflow presented in this study was designed to compare CNN attribution maps with predefined image-derived descriptor maps within the seed foreground. By excluding background pixels using foreground masks and aligning attribution maps and descriptor maps in a shared coordinate system, spatial association analysis was restricted to the seed object. This procedure complements visual inspection of attribution heatmaps with numerical summaries of foreground-restricted associations with predefined descriptor responses. Descriptor summary feature classification was also performed as a separate analysis to examine how foreground-level visual variation captured by descriptor maps was related to seed class discrimination.

This approach can be useful for seed image datasets with limited expert annotation [4–10]. Because the descriptor maps were constructed from broad visual properties, including color and intensity, spatial frequency, texture, edge, and shape, they provided reproducible reference maps for comparing attribution patterns within an image-derived visual context. The combined use of spatial association analysis and descriptor summary feature classification allowed the foreground distribution of attribution maps and descriptor-level class separability to be organized separately. In authentication-oriented medicinal seed image analysis, this distinction may help interpret model outputs while keeping descriptor responses within the scope of image-derived visual context.

Several limitations should be considered. All images were collected and processed under controlled imaging and preprocessing conditions, and the stability of the workflow should be evaluated further using external datasets that include different lighting conditions, backgrounds, imaging devices, and seed orientations. The descriptor maps were based on predefined image-derived responses, and their relationships with expert-defined diagnostic traits, morphological measurements, chemical profiles, or genetic information require additional data for more specific evaluation [11–13]. IG attribution can also be affected by the baseline setting and attribution aggregation strategy, while Grad-CAM can show different spatial responses depending on the target layer and activation resolution [18,20,21]. Because descriptor summary feature classification uses foreground-level summary statistics, detailed spatial patterns in descriptor maps should be interpreted together with spatial association analysis.

Future studies can reevaluate this workflow using independent seed image datasets and diverse imaging conditions. When expert annotations or morphology-based measurements are available, the relationship between image-derived descriptor responses and practical discrimination criteria can be compared in greater detail. Applying multiple CNN architectures

and attribution methods would also support a clearer evaluation of the stability of attribution–descriptor spatial associations. These extensions may help apply the foreground-restricted attribution contextualization workflow to seed quality assessment and other plant image analysis tasks beyond medicinal plant seed classification.

5. Conclusions

This study presented a foreground-restricted descriptor contextualization workflow for comparing CNN attribution maps with predefined image-derived descriptor maps in medicinal plant seed image classification. Under controlled imaging and preprocessing conditions, three CNN backbones achieved near-ceiling classification performance, and zero-baseline absolute IG maps from ConvNeXt-Small were used as the primary attribution reference. By aligning seed foreground masks, attribution maps, and descriptor maps in a shared coordinate system, the workflow quantified attribution–descriptor spatial association only within the seed object. Among the tested descriptor maps, LAB L, Brightness, and FFT low-pass showed the highest positive foreground-restricted associations with IG attribution maps, indicating that attribution responses were mainly co-distributed with lightness-related, intensity-related, and low-frequency image responses within the seed foreground.

Descriptor summary feature classification further showed that predefined descriptor maps contained class-discriminative visual information at the foreground level. The full descriptor set achieved high classification performance, and selected descriptor subsets also captured class-relevant visual variation. These results support the complementary use of attribution–descriptor spatial association and descriptor summary classification: the former summarizes spatial co-distribution between attribution and descriptor responses, while the latter evaluates

descriptor-level class separability. The proposed workflow provides a reproducible way to organize CNN attribution patterns within image-derived visual contexts and may support more structured interpretation of CNN responses in authentication-oriented medicinal seed image analysis, seed quality assessment, and related plant image analysis tasks.

Acknowledgements

Credit authorship contribution statement

Su-Min Chin: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft. **Yu-Jin Jeon:** Conceptualization. **Yea-Jin Park:** Investigation, Data curation. **Chi-Won Choi:** Investigation. **Min-Woo Kim:** Investigation. **Min-Hyeok Jang:** Investigation. **Dae-Hyun Jung:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing.

Funding

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

Declaration of competing 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.

Data Availability

The raw and processed images, foreground masks, sample metadata, fixed train/test split, and associated documentation are publicly available in Zenodo at <https://doi.org/10.5281/zenodo.21537568>. The analysis code is available at <https://github.com/sgn08178-bit/medicinal-seed-attribution-descriptors> and archived in Zenodo at <https://doi.org/10.5281/zenodo.21537321>.

References

1. Li, L., Zhang, Q. & Huang, D. A Review of Imaging Techniques for Plant Phenotyping. *Sensors* **14**, 20078–20111 (2014).
2. Li, Y. *et al.* A Review of Computer Vision Technologies for Plant Phenotyping. *Comput. Electron. Agric.* **176**, 105672 (2020).
3. Gehan, M. A. *et al.* PlantCV v2: Image Analysis Software for High-Throughput Plant Phenotyping. *PeerJ* **5**, e4088 (2017).
4. Dell'Aquila, A. Towards New Computer Imaging Techniques Applied to Seed Quality Testing and Sorting. *Seed Sci. Technol.* **35**, 519–538 (2007).
5. Tanabata, T., Shibaya, T., Hori, K., Ebana, K. & Yano, M. SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis. *Plant Physiol.* **160**, 1871–1880 (2012).
6. Halcro, K. *et al.* The BELT and phenoSEED Platforms: Shape and Colour Phenotyping of Seed Samples. *Plant Methods* **16**, 49 (2020).
7. Loddo, A. *et al.* An effective and friendly tool for seed image analysis. *Vis. Comput.* **39**, 335–352 (2023).
8. ElMasry, G. *et al.* Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant Methods* **15**, 24 (2019).
9. Granitto, P. M., Navone, H. D., Verdes, P. F. & Ceccatto, H. A. Automatic identification of weed seeds by color image processing. *Comput. Electron. Agric.* **47**, 15–25 (2005).
10. Jin, C. *et al.* Application of Deep Learning for High-Throughput Phenotyping of Seed: A Review. *Artif. Intell. Rev.* <https://doi.org/10.1007/s10462-024-11079-5> (2025)
[doi:10.1007/s10462-024-11079-5](https://doi.org/10.1007/s10462-024-11079-5).
11. Ichim, M. C. The DNA-Based Authentication of Commercial Herbal Products Reveals Their Globally Widespread Adulteration. *Front. Pharmacol.* **10**, 1227 (2019).

- 805 12. Nithanial, S., Vassou, S. L., Poovitha, S., Raju, B. & Parani, M. Identification of species
806 adulteration in traded medicinal plant raw drugs using DNA barcoding. *Genome* **60**, 139–
807 146 (2017).
- 808 13. Raclariu, A. C., Heinrich, M., Ichim, M. C. & de Boer, H. Benefits and limitations of
809 DNA barcoding and metabarcoding in herbal product authentication. *Phytochem. Anal.*
810 **29**, 123–128 (2018).
- 811 14. Gulzar, Y., Hamid, Y., Soomro, A. B., Alwan, A. A. & Journaux, L. A Convolution Neural
812 Network-Based Seed Classification System. *Symmetry* **12**, 2018 (2020).
- 813 15. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep
814 Convolutional Neural Networks. *Commun. ACM* **60**, 84–90 (2017).
- 815 16. Ribeiro, M. T., Singh, S. & Guestrin, C. Why Should I Trust You? Explaining the
816 Predictions of Any Classifier. in 97–101 (2016). doi:10.18653/v1/N16-3020.
- 817 17. Too, E. C., Yujian, L., Njuki, S. & Yingchun, L. A comparative study of fine-tuning deep
818 learning models for plant disease identification. *Comput. Electron. Agric.* **161**, 272–279
819 (2019).
- 820 18. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. in vol.
821 70 3319–3328 (2017).
- 822 19. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via
823 Gradient-Based Localization. *Int. J. Comput. Vis.* **128**, 336–359 (2020).
- 824 20. Adebayo, J. *et al.* Sanity Checks for Saliency Maps. in vol. 31 (2018).
- 825 21. Hooker, S., Erhan, D., Kindermans, P.-J. & Kim, B. A Benchmark for Interpretability
826 Methods in Deep Neural Networks. Preprint at
827 <https://doi.org/10.48550/ARXIV.1806.10758> (2018).
- 828 22. Cooley, J. W. & Tukey, J. W. An algorithm for the machine calculation of complex
829 Fourier series. *Math. Comput.* **19**, 297–301 (1965).

- 830 23. Mallat, S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet
831 Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693 (1989).
- 832 24. Manjunath, B. S. & Ma, W. Y. Texture features for browsing and retrieval of image data.
833 *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 837–842 (1996).
- 834 25. Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution Gray-Scale and Rotation
835 Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal.*
836 *Mach. Intell.* **24**, 971–987 (2002).
- 837 26. Felzenszwalb, P. F. & Huttenlocher, D. P. Distance transforms of sampled functions.
838 *Theory Comput.* **8**, 415–428 (2012).
- 839 27. Zahn, C. T. & Roskies, R. Z. Fourier Descriptors for Plane Closed Curves. *IEEE Trans.*
840 *Comput.* **C-21**, 269–281 (1972).
- 841 28. Qin, X. *et al.* U2-Net: Going Deeper with Nested U-Structure for Salient Object
842 Detection. *Pattern Recognit.* **106**, 107404 (2020).
- 843 29. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE*
844 *Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, Miami, FL,
845 2009). doi:10.1109/CVPR.2009.5206848.
- 846 30. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in
847 770–778 (2016). doi:10.1109/CVPR.2016.90.
- 848 31. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural
849 Networks. in vol. 97 6105–6114 (2019).
- 850 32. Liu, Z. *et al.* A ConvNet for the 2020s. in 11976–11986 (2022).
851 doi:10.1109/CVPR52688.2022.01167.
- 852 33. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. Preprint at
853 <https://doi.org/10.48550/arXiv.1711.05101> (2019).
- 854 34. Spearman, C. The proof and measurement of association between two things. *Am. J.*

855 *Psychol.* **15**, 72–101 (1904).

856 35. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
857 Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300
858 (1995).

859 36. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).

860 37. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**,
861 2825–2830 (2011).

862 38. Van Der Walt, S. *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

863 39. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

864 40. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python.
865 *Nat. Methods* **17**, 261–272 (2020).

866

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [04SupplementaryInformation.docx](#)