



A deep learning method for differentiating safflower germplasm using optimal leaf structure features



Hoang Thien Van ^{a,*}, Phuong Thuy Khuat ^b, Trang Van ^c, Thai Thanh Tuan ^d, Yong Suk Chung ^{e,*}

^a Faculty of Information Technology, HUTECH University, Ho Chi Minh City, Viet Nam

^b Computer Science Center, University of Science, VNU-HCM, HUTECH University, Ho Chi Minh City, Viet Nam

^c Faculty of Information Technology, Ho Chi Minh City University of Economics and Finance, Ho Chi Minh City, Viet Nam

^d University of Information Technology, VNU-HCM, Ho Chi Minh city, Viet Nam

^e Department of Plant Resources and Environment, Jeju National University, Jeju, Republic of Korea

ARTICLE INFO

Keywords:

Safflower
Plant classification
Gabor filter
Vision transformer
Deep learning

ABSTRACT

Medicinal plants, such as safflower (*Carthamus tinctorius*), are essential in both conventional and modern healthcare. This study evaluates the reliability of leaf image classification for differentiating safflower germplasm accessions, despite the challenge posed by similar leaf structures across varieties. Traditional classification methods can be time-consuming and error prone, underscoring the need for fine-grained classification techniques. To address this, we introduce a novel comprehensive leaf database of safflower varieties that is meticulously curated by experts. This database can be used to support future research. We evaluate state-of-the-art deep learning methods for classifying safflower varieties and propose a novel approach using a Vision Transformer (ViT) model with an optimal leaf structure feature (OLSF). The OLSF, calculated as the average response of a multidirectional Gabor filter bank and optimized with the structural similarity index measure (SSIM), enhances complex leaf features, such as veins, texture, and frequency variations, to enhance the classification performance of deep learning models. The experimental results reveal that the OLSF-ViT model achieves excellent accuracy scores of 100 %, 99.05 %, and 89.65 % on the Folio, UCI Leaves, and JNUSafflower datasets, respectively. These findings demonstrate that leaf image analysis is an effective and affordable tool for investigating the phenotypic diversity among safflower cultivars. This study highlights the potential of OLSF-ViT in automatic plant classification, and the results can be used in plant science, herbal medicine, and biodiversity conservation.

1. Introduction

Safflower (*Carthamus tinctorius* L.) is a medicinal plant that highly valuable in traditional healthcare and modern medicine because of its numerous pharmacological effects that can be used to treat various diseases, including cardiovascular diseases, immune disorders, and inflammation (Mayerhofer et al., 2010; Zhou et al., 2014). Twenty-five species of the *Carthamus* genus have been reported worldwide. Of these, *Carthamus tinctorius* and *Carthamus lanatus* originate in China. Although *Carthamus lanatus* is a rare wild species in the Yunnan province, *Carthamus tinctorius* is widely grown throughout China. Centuries of natural and artificial selection have resulted in distinct *Carthamus* populations, each with a distinctive appearance and chemical properties (Mayerhofer

et al., 2010). Therefore, the accurate identification of safflower varieties is crucial for assessing genetic diversity and preserving biodiversity (Xing et al., 2024; Abdipour et al., 2019). Conventional classification methods can be time-consuming and error prone, highlighting the necessity for automated solutions that incorporate advanced image processing and machine learning techniques (DeeptiBarhate and Dubey, 2023). Studies have classified plant species using leaf images (DeeptiBarhate and Dubey, 2023; Ghosh et al., 2024; Puri et al., 2022), and even differentiated varieties within species, such as chili (Suwarningsih et al., 2022), rice (Koklu et al., 2021), wheat (Ceyhan et al., 2024), grapevines (Diago et al., 2013), and walnuts (Karadeniz et al., 2024). However, classifying safflower leaf varieties remains challenging because limited studies have focused on the automatic

* Corresponding authors.

E-mail addresses: vt.hoang@hutech.edu.vn (H. Thien Van), kt.phuong@csc.hcmus.edu.vn, ktphuong24nct@hutech.edu.vn (P.T. Khuat), trangvtt@uef.edu.vn (T. Van), tuantt@uit.edu.vn (T.T. Tuan), yschung@jejunu.ac.kr (Y. Suk Chung).

classification of safflower varieties. Furthermore, classification remains difficult because the high similarity in leaf shapes among varieties requires fine-grained classification methods. Plant classification can be automated using local filter-based methods and deep-learning-based techniques. Although local filter-based methods such as Gabor filters have been effective in specific cases (Chi et al., 2003; Cope et al., 2010; Ishak et al., 2009; Zhang et al., 2020), these methods exhibit certain limitations such as low generalization, particularly because of diverse and complex datasets. These methods rely heavily on high-quality input images and require complex feature designs and parameter tuning. Deep learning, particularly convolutional neural networks (CNNs), has advanced plant classification (Ceyhan et al., 2024; DeeptiBarhate and Dubey, 2023; Koklu et al., 2021; Suwarningsih et al., 2022). CNNs can extract complex features from images to achieve accurate classification. However, large and diverse datasets are typically required to avoid overfitting. To address this problem, transfer learning has been devised to allow models to use pretrained knowledge from extensive datasets for improving generalization and reduce training time. However, CNNs and transfer learning still cannot accurately identify complex spatial relationships within images. Vision Transformers (ViTs) (Dosovitskiy et al., 2021), which use attention mechanisms, provide a novel approach for effectively capturing spatial details. ViTs have demonstrated strong performance in tasks such as image classification, particularly in plants. However, representing the texture and frequency features using ViT remains challenging. Additionally, local filter-based techniques, such as Gabor filters (Dhakshayani and Surendiran, 2023; Oppong et al., 2022), have been used to enhance the image quality and emphasize key features and improve the classification performance of CNN-based models. Therefore, this study focused on investigating the diversity of safflower varieties using deep learning to classify them based on leaf characteristics and developing classification methods that combine the ViT model with Gabor filter-enhanced leaf images. The major contributions of this study are as follows:

1. A comprehensive database of safflower leaf cultivated and collected by experts that can be used as a reference for cultivar characterization.
2. Assessment of phenotypic diversity in safflower varieties using the state-of-the-art deep learning methods for leaf classification.
3. Proposal of the optimal leaf structure feature (OLSF) technique, which averages responses from a Gabor filter bank and optimizes them using the structural similarity index (SSIM), to enhance the classification performance of the ViT model.
4. Evaluation of state-of-the-art deep learning methods and the OLSF-ViT model on the Folio, UCI Leaf, and safflower datasets for high classification accuracy and reliable plant identification.

The remainder of this paper is organized as follows: Section 2 presents a literature review of the key concepts relevant to the study. Section 3 describes the JNUSafflower dataset used in this study. Section 4 outlines the methodology, including the feature extraction methods and the proposed model. Section 5 analyzes and discusses the results, and Section 6 presents the conclusion to the paper.

2. Related works

Safflower leaf classification is a manual process that relies on expert morphological assessments. However, this method is slow and susceptible to human error (Xing et al., 2024; Abdipour et al., 2019). Digital technology has facilitated the use of image capture and processing for classification. However, early techniques could not handle the complexity of plant features. Thus, the necessity of automated, precise, and rapid classification systems resulted in the implementation of deep learning methods. Therefore, algorithms for leaf classification can be grouped into three approaches as follows: (1) local filter-based techniques (2) deep-learning-based techniques and (3) hybrid techniques.

Local filter-based techniques for plant feature extraction and classification have been widely investigated (Zhang et al., 2020). Chi et al. (Chi et al., 2003) introduced a novel approach using Gabor filter banks to focus on the bark texture features and identify plant species. Ishak et al. (Cope et al., 2010) introduced a novel image analysis technique that combines the Gabor wavelet and gradient field distribution methods for classifying weed types. Cope et al. (Ishak et al., 2009) developed a novel plant classification method based on leaf texture using joint distributions across various scales of the Gabor filter. Chaki and Parekh (Chaki and Parekh, 2012) introduced a novel automated system for plant species recognition using a Gabor filter analysis on leaf images for adjusting filter parameters. Chaki et al. (Chaki et al., 2015) developed a novel plant classification method using a Gabor filter and a grayscale co-occurrence matrix to represent the texture features of plant leaves. This method incorporated a feed forward back-propagation multilayered perceptron (MLP) and a neuro-fuzzy controller classifier. Saleem et al. (Saleem et al., 2019) investigated leaf visual cues and extraction methods to obtain optimized morphological, geometric, and textural features to achieve 97.6 % recognition accuracy on the Flavia dataset. Yang (Yang, 2021) introduced a multiscale triangle descriptor to extract the leaf shape features and used the local binary pattern histogram Fourier (LBP-HF) to compute features and achieve classification accuracies of 99.1 %, 98.4 %, and 95.6 % on the Flavia, Swedish, and MEW2012 datasets, respectively. Lv et al. (Lv and Zhang, 2023) proposed a novel multi-feature fusion approach for plant leaf recognition by combining local binary patterns (LBP) and histograms of oriented gradient (HOG) features. Employing dimensionality reduction through principal component analysis (PCA), a composite feature vector was inputted into an extreme learning machine (ELM) for leaf identification to achieve recognition accuracies of 99.30 % on the Flavia dataset and 99.52 % on the Swedish dataset. Wu et al. (Wu et al., 2024) proposed a novel composite descriptor that combines contour and appearance features for plant species identification. The method incorporates two fusion schemes: (1) low-level feature fusion, in which the local triangle and Speeded-Up Robust Features (SURF) describe leaf images with Hausdorff distance measuring discrepancies, and (2) global feature fusion, in which Fisher vectors encode high-level shape and appearance features with Euclidean distance computing dissimilarity. The method achieved recognition accuracies of 99.83 % and 91.25 % for the Flavia and PlantVillage datasets, respectively.

The application of CNNs, particularly in deep learning, has transformed plant classification. Suwarningsih et al. (Suwarningsih et al., 2022) investigated deep CNN models such as AlexNet, VGG16, Inception-v3, and DenseNet-121 to classify chili varieties based on leaf observations and achieved accuracies from 70.18 % to 78.37 %. DenseNet-121 demonstrated the highest accuracy of 78.37 %. Karadeniz (Karadeniz et al., 2024) introduced a walnut species dataset and proposed a classification method using VGG16, VGG19, and AlexNet for feature extraction with a whale optimization algorithm to select the top features from these models. Lee et al. (Lee et al., 2015) achieved a remarkable accuracy of 99.5 % in classifying 44 plant species using CNNs with features learned without supervision. Dyrmann et al. (Dyrmann et al., 2016) introduced a CNN method to identify plant species in color images and achieved notable 86.2 % classification accuracy across 22 weed and crop species. Lee et al. (Lee et al., 2017) proposed a CNN method for plant classification and demonstrated its capability to learn vein patterns in leaves by distinguishing features directly from raw image data. This study emphasized the hierarchical representation of leaf features and differentiated the contextual information captured by CNNs trained on entire-leaf images from leaf patches. Sun et al. (Sun et al., 2017) demonstrated the effectiveness of a 26-layer deep residual network, ResNet26, in accurately identifying plant species from plant images. Kaya et al. (Kaya et al., 2019) performed a transfer learning analysis for plant classification models using deep neural networks and demonstrated the effectiveness of transfer learning with fine-tuning on various datasets, including Flavia,

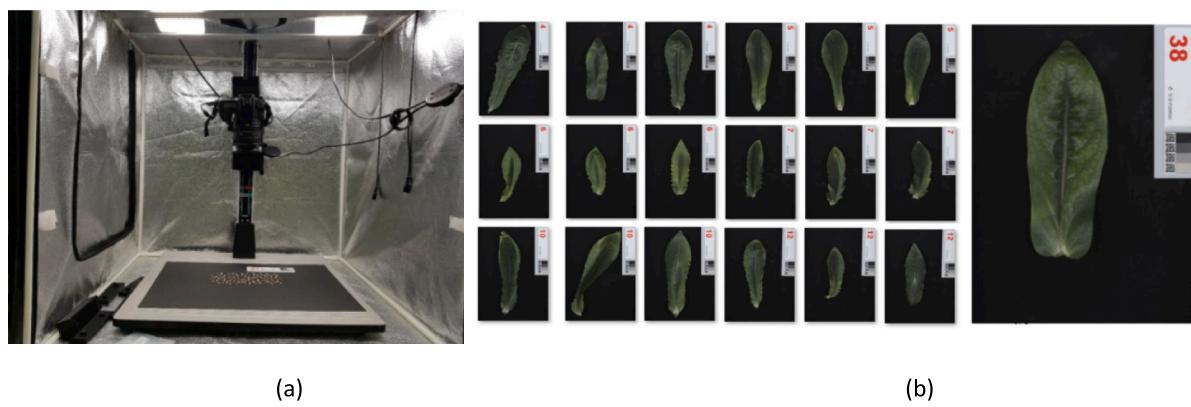


Fig. 1. (a) The capture device and (b) some examples of JNUSafflower dataset.

PlantVillage, Swedish Leaf, and UCI Leaf. Venkatesh et al. (Venkatesh et al., 2021) introduced a deep learning system using a fine-tuned MobileNet CNN for the automated yield estimation of strawberries and cherries and achieved an accuracy of 98.60 %. CNNs and pretrained models cannot capture complex spatial relationships, which can be captured by ViTs (Dosovitskiy et al., 2021; Conde and Turgutlu, 2021; Van Hieu et al., 2023). ViTs incorporate an attention mechanism to comprehensively capture spatial information, facilitating the detection of complex relationships within images. Dosovitskiy et al. (Dosovitskiy et al., 2021) demonstrated that CNNs are not necessary for image classification tasks, indicating that a pure transformer applied directly to sequences of image patches provide satisfactory outcomes. Conde et al. (Conde and Turgutlu, 2021) introduced a novel multistage ViT framework for fine-grained image classification to achieve a maximum accuracy of 99.8 % without architectural changes, highlighting the potential of ViT in agriculture. Hieu et al. (Van Hieu et al., 2023) introduced PlantKViT, combining ViT with the KNN algorithm, and achieved 93 % accuracy in identifying forest plants in the Danang Forest Plant dataset, outperforming ConvNeXt (89 %) and ResNet-152 (76 %).

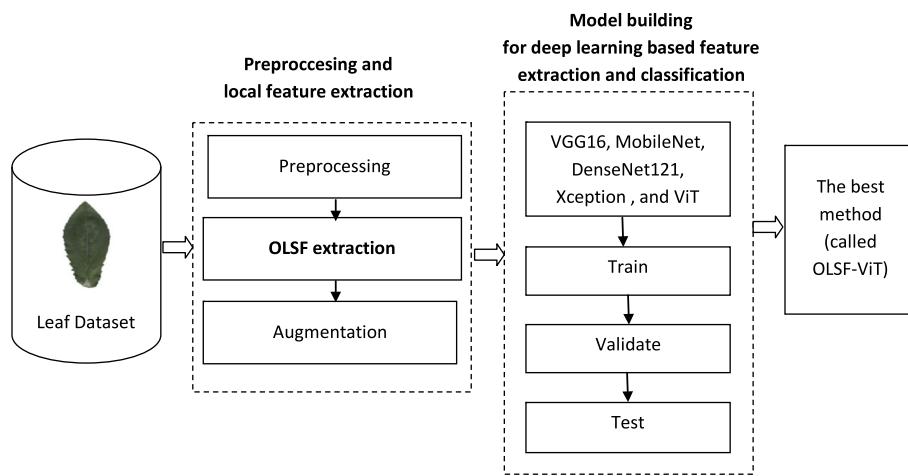
Studies have proposed hybrid approaches for improving plant classification. Wu et al. (Wu et al., 2023) proposed a plant identification method that combines shape and convolutional features. An improved multiscale triangle descriptor (IMTD) captures leaf shape properties, and convolutional features at different levels are analyzed for identification. By combining these complementary features, the method achieved recognition accuracies of 96.28 %, 99.47 %, and 91.29 % for the Flavia, Swedish, and LeafSnap datasets, respectively. Dhakshayani et al. (Dhakshayani and Surendiran, 2023) proposed a GF-CNN, an enhanced CNN that incorporates Gabor filters for maize disease classification. Comparative experiments with both machine learning classifiers and CNNs revealed that the GF-CNN outperformed existing models to achieve 99.25 % accuracy on the maize Plant Village dataset. SO et al. (Oppong et al., 2022) developed a computer vision system that combines CNNs with log-Gabor filters to identify medicinal plants based on leaf textures. Tested on a dataset from the Centre of Plant Medicine Research, Ghana (MyDataset) with 49 species, the system used ten pretrained networks (including AlexNet, GoogLeNet, DenseNet201, Inceptionv3, Mobilenetv2, Resnet18, Resnet50, Resnet101, VGG16, and VGG19) for feature extraction, with DenseNet201 achieving the best accuracy of 87 %. The proposed OTAMNet model, which integrates a log-Gabor layer into DenseNet201, achieved 99 % for Flavia, 100 % for Swedish Leaf, 99 % for MD2020, and 97 % for Folio. Both studies (Dhakshayani and Surendiran, 2023; Oppong et al., 2022) combined

multiple techniques to overcome the challenges of plant classification and encouraged hybrid approaches that use the strengths of spatial and textural analyses. Using these advancements, we proposed the OLSF-ViT method, which combines Gabor filters and ViT model to improve plant classification efficiency. Furthermore, we proposed an optimal application of Gabor filters based on the SSIM metric to ensure that the Gabor filter parameters enhance leaf texture features across all considered datasets. The OLSF technique averages the responses from a Gabor filter bank and selects the optimal responses using the SSIM for each dataset to enhance the ability of the ViT model to classify leaf types effectively.

3. JNUSafflower dataset

The JNUSafflower dataset comprises labeled images of safflower variety leaves cultivated and collected at the Rural Development Administration in Jeonju, Republic of Korea. For imaging, an indoor studio ($800 \times 800 \times 800$ mm) was set up to minimize ambient light interference. An 18 W white LED (5600 K, CN-T96, Plastic, Republic of Korea) was used to ensure consistent illumination and prevent data distortion or damage from shadows. A custom-made white background plate was used to represent the color of the safflower leaves. Camera system settings were adjusted according to the methodology described by (Yu et al., 2023). The camera, Canon EOS D200II (with an EF-S 18–55 mm lens, Canon, Japan), features a 24.1-megapixel CMOS sensor. The initial camera settings were ISO 200, focal length of 35 mm, and exposure time of 1/25. Owing to the convex nature of the lens, which can cause image distortion (thicker at the center than at the edges), we used the built-in distortion-correction function of the camera. The adjusted settings were ISO 100, focal length of 18 mm, and exposure time of 1/13. The background was changed to black to contrast with the safflower leaves. Customized 3D panels were designed to accurately represent the leaf size. Each panel included crop-name tags, four grayscale calibration markers (10 mm each), and a QR code to maintain consistent positioning (<https://www.group8tech.com/grayscale-calibration>).

Images were saved in the JPG format with a resolution of 6024×4024 pixels. Safflower varieties were grown under controlled environmental conditions, including regulated temperatures, irrigation, and fertilizer levels. The experiment was conducted in an experimental garden containing 43 safflower varieties, each randomly positioned in a greenhouse plot and labeled with a variety of codes. Six leaf samples were collected from each variety before and after the bloom stage throughout the growing season. Leaves were selected based on

**Fig. 2.** The proposed framework for leaf recognition.**Fig. 3.** The processing steps for extracting leaf Region of Interest (ROI) images.**Fig. 4.** Samples of safflower image augmentation.

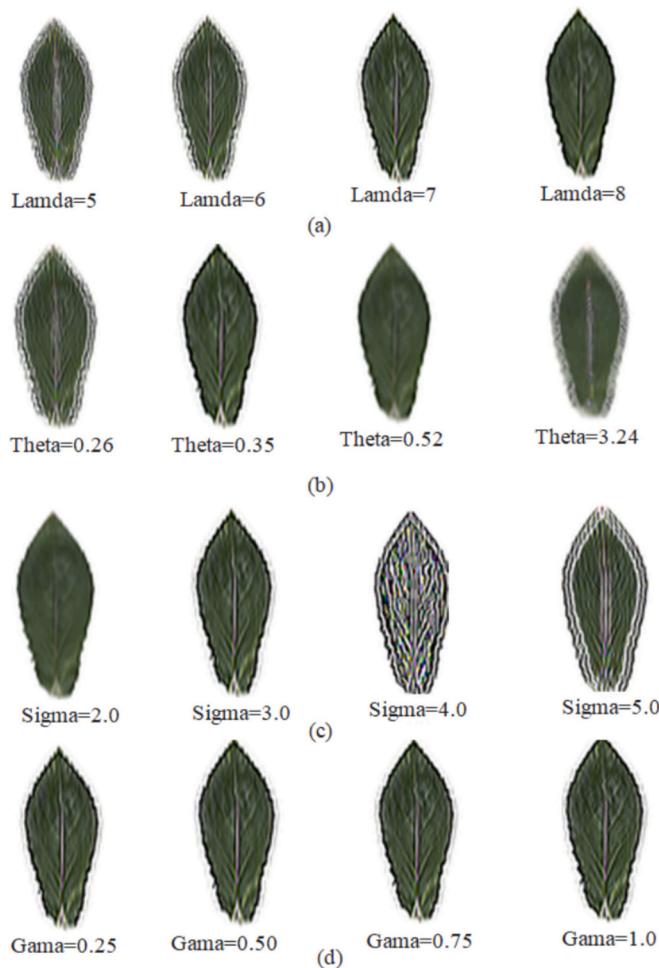


Fig. 5. Sixteen filtered images correspond to sixteen different kernels: (a) Images filtered by the Gabor filter with varying lambda (5, 6, 7, 8), fixing other parameters ($\theta = 0$, $\gamma = 0.5$, $\sigma = 3$, $\Psi = 0$), (b) images filtered by the Gabor filter with varying theta ($\pi/12$, $\pi/9$, $\pi/6$, π), fixing other parameters ($\lambda = 7$, $\gamma = 0.5$, $\sigma = 3$, $\Psi = 0$), (c) images filtered by the Gabor filter with varying sigma (2, 3, 4, 5), fixing other parameters ($\theta = 0$, $\lambda = 7$, $\gamma = 0.5$, $\Psi = 0$), and (d) images filtered by the Gabor filter with varying gamma (0.25, 0.5, 0.75, 1.00), fixing other parameters ($\theta = 0$, $\lambda = 7$, $\sigma = 3$, $\Psi = 0$).

Table 1
Parameter grid.

| Parameter | # Values |
|---------------------|--------------------|
| num_kernels (N) | [4, 8, 16] |
| kernel size | [(15,15), (21,21)] |
| sigma (σ) | [2–5] |
| lambd (λ) | [5–8] |
| gamma (γ) | [0.25, 0.5, 0.75] |
| psi (ψ) | [0] |

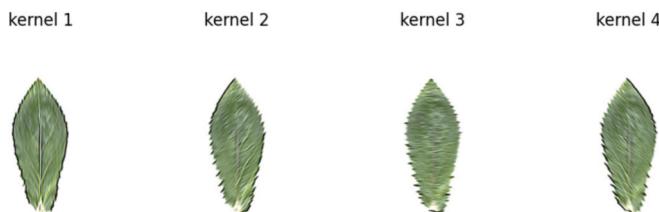


Fig. 6. Using Gabor kernels with 4 different directions to generate 4 filtered images.



Fig. 7. Comparison between the original image and the OLSF image.

standardized criteria such as area, weight, and color. The database initially contained 258 images before augmentation (Fig. 1).

4. Proposed method

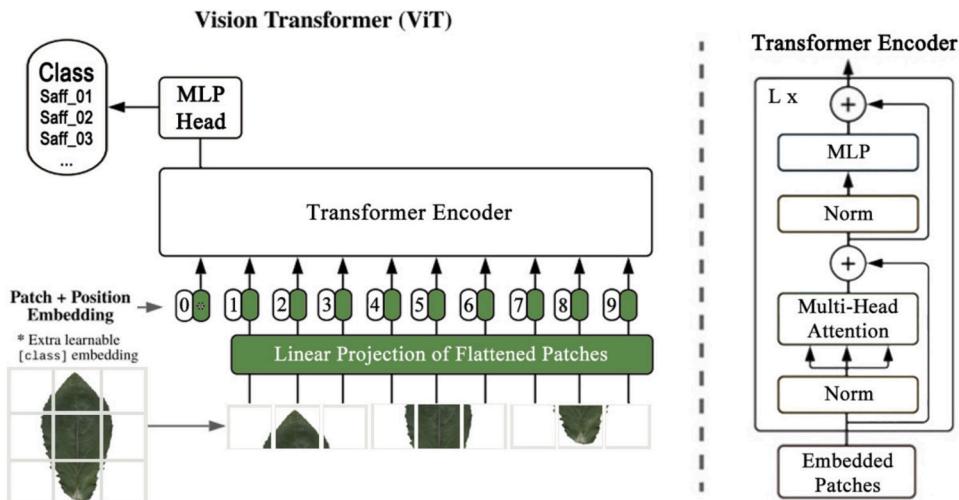
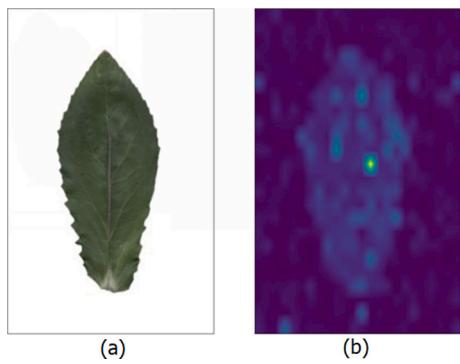
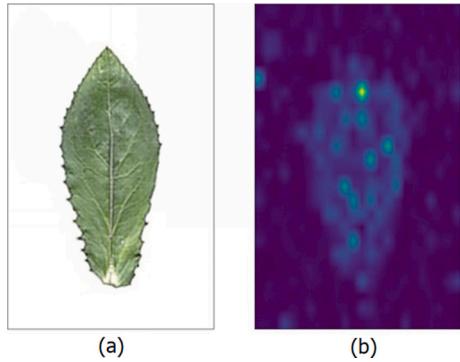
This study focused on the diversity of safflower varieties through deep learning by classifying them based on leaf characteristics and introducing a method for safflower variety distinction. We investigated the state-of-the-art CNN models (VGG16, MobileNet, DenseNet121, and Xception) and the ViT model for safflower leaf classification. The ViT revolutionizes image processing with its attention mechanism and can effectively capturing spatial details across the image spectrum (Dosovitskiy et al., 2021). Gabor filters are highly effective for texture analysis because they can capture both frequency and orientation information to represent the spatial characteristics of an image in a manner that aligns with human visual perception. This enhances the image quality by highlighting features sensitive to local patterns and textures, rendering Gabor filters suitable for tasks such as image classification and object recognition. They complement the local feature extraction abilities of the CNN and ViT models, particularly in extracting frequency-based features across various orientations. Therefore, we proposed a novel method for classifying medicinal plants and safflower varieties by combining a ViT model with OLSF and termed the model OLSF-ViT. The OLSF is computed using a Gabor filter bank, with the optimal response selected based on the SSIM (Li et al., 2020). This method enables the OLSF to capture complex details, such as leaf veins, texture, and frequency variations, enhancing the capability of ViT for deep learning-based leaf recognition. Fig. 2 illustrates the proposed framework for leaf recognition.

4.1. Preprocessing

Region of interest (ROI) image segmentation is critical for feature extraction. We initially resized the images and applied a threshold to isolate the leaf area and eliminate the background. Subsequently, we used the grab-cut image segmentation technique (Rother et al., 2004) with the number of iterations set to 5 to separate the leaf image from its background, placing the detected leaves at the center of the images with dimensions of 526×526 . Fig. 3 illustrates the extraction process of the ROI images of the safflower leaves.

4.2. Augmentation

Data augmentation increases the diversity of the training dataset, enhancing the performance and resilience of the model. Moreover, augmented test or validation data were used to improve the ability of the model to generalize to real-world scenarios (Bello et al., 2021). Because of the limited sample size of the safflower dataset, both the training and testing sets underwent augmentation, generating 28 additional images per sample. This augmentation process included rotation at angles (15° , 30° , 45° , 60° , 75° , 90° , 180° , and 270°), horizontal and vertical flipping, and resizing with scales (0.5, 0.75, 1.5, and 2), resulting in 24 images for

**Fig. 8.** ViT architecture.**Fig. 9.** (a) The original image and (b) its “attention map last layer” image.**Fig. 10.** (a) The Gabor-filtered image and (b) its “attention map last layer” image.

training and 10 images for testing, in each sample (See Fig. 4).

4.3. Extracting OLSF

The Gabor filter (Cope et al., 2010; Chaki and Parekh, 2012), being a convolution filter, combines a Gaussian component and a sinusoidal term. The Gaussian part provides weighting factors while the sinusoidal aspect adds directionality to the filter. The Gabor filter is defined as follows:

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left\{-\frac{\dot{x}^2 + \gamma^2 \dot{y}^2}{2\sigma^2}\right\} \exp\left\{i\left(2\pi \frac{\dot{x}}{\lambda} + \psi\right)\right\},$$

$$\dot{x} = x \cos \theta + y \sin \theta, \dot{y} = -x \sin \theta + y \cos \theta \quad (1)$$

where $i = \sqrt{-1}$, λ is the wavelength of the sine component, θ is the orientation of the function, ψ is the phase offset, σ is the standard deviation of the Gaussian envelop, and γ is spatial aspect ratio. Numerous digital filters (kernels) can be defined by varying Gabor parameters (Fig. 5).

We investigated deep learning methods for recognizing safflower varieties. However, effective fine-grained classification requires image preprocessing to enhance image quality. Furthermore, deep learning approaches exhibit considerable limitations in representing frequency features in various orientations. Gabor filters are designed to be sensitive to both orientation and frequency, rendering the filters suitable for identifying the unique patterns and textures in plant leaves. Therefore, we devised a novel deep learning-based approach for more effective safflower variety classification by applying Gabor filters to enhance image quality. The selection of Gabor filter parameters is crucial because they directly influence the ability of the filter to capture relevant features in the image. These parameters were chosen based on preliminary experiments for maximizing the detection of important textural features in plant leaves, such as vein patterns and edge details. In these experiments, we systematically varied each parameter within a defined range and observed the effect on feature clarity and completeness in a controlled set of plant leaf images. Conventional Gabor filter parameter selection, which is typically based on trial and error, frequently results in suboptimal performance. To overcome this problem, the OLSF method was proposed for automatically selecting the optimal parameters for Gabor filters using the SSIM metric (Li et al., 2020). The SSIM is a full-reference image structure similarity evaluation metric that is used to assess similarity based on brightness, contrast, and structure. The SSIM values ranged from 0 to 1, with higher values indicating lower image distortion (Li et al., 2020). This range ensures that leaf veins are emphasized without disrupting the structural integrity of the leaves across the dataset, highlighting critical leaf texture patterns and considerably improving the classification accuracy of deep learning models. Furthermore, certain local feature enhancement methods (e.g., such as LBP and HOG) can alter or obscure the shape and structure of the original image. Therefore, these methods do not utilize the strengths of deep learning models in learning both prominent local and global features from the original images for classification. By contrast, OLSF enhances leaf vein quality and simultaneously preserves the texture and

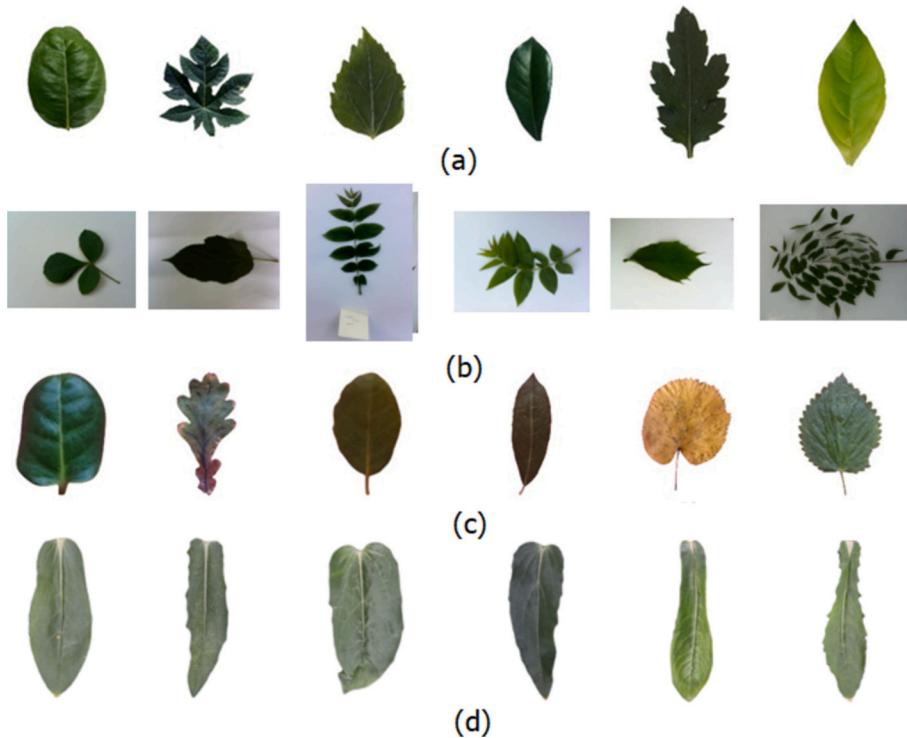


Fig. 11. Sample images from (a) Folio, (b) LeafSnap, (c) UCI Leaf and (d) JNUSafflower dataset.

Table 2
Dataset properties and parameters.

| Dataset | # of classes | # Samples | Color | # Avg images/class | Training set | Testing set |
|-----------|--------------|-----------|-------|--------------------|--------------|-------------|
| Folio | 32 | 637 | RGB | 20 | 510 | 127 |
| UCI Leaf | 40 | 443 | RGB | 11 | 355 | 88 |
| Leafsnap | 184 | 7719 | RGB | 42 | 6109 | 1610 |
| Safflower | 43 | 258 | RGB | 6 | 4128 | 860 |

Table 3
Comparative performance of deep learning models on JNUSafflower dataset.

| Methods | Accuracy (%) | F1 score (%) | Training time (second) | Test time (second/sample) |
|-------------|--------------|--------------|------------------------|---------------------------|
| VGG16 | 38.02 | 34.00 | 11,376 | 0.008 |
| Xception | 44.54 | 43.00 | 27,332 | 0.012 |
| MobileNet | 61.28 | 61.00 | 22,276 | 0.008 |
| DenseNet201 | 51.51 | 50.00 | 17,340 | 0.008 |
| ViT | 88.02 | 87.65 | 26,841 | 0.027 |

structure of the original image, which improves the learning effectiveness of deep learning methods compared with using the original images. OLSF can be computed as follows:

$$f_{OLSF} = \text{Average}_{\theta_i=\frac{\pi(i-1)}{N}, i=1,2,\dots,N} f^*g(x, y, \lambda, \theta_i, \psi, \sigma, \gamma) \quad (2)$$

where f_{OLSF} is the OLSF image obtained by convoluting image f using the Gabor filter with N directions. The parameters (N , λ , ψ , σ , γ , and size) of the Gabor filters are determined using the training dataset by the following processing steps:

Table 4
Comparative performance of deep learning models on Folio dataset.

| Methods | Accuracy (%) | F1 score (%) | Training time (second) | Test time (second/sample) |
|-------------|--------------|--------------|------------------------|---------------------------|
| VGG16 | 92.19 | 92.00 | 8540 | 0.114 |
| Xception | 93.75 | 94.00 | 12,084 | 0.114 |
| MobileNet | 95.31 | 95.00 | 4540 | 0.085 |
| DenseNet201 | 98.44 | 98.00 | 5864 | 0.088 |
| ViT | 98.85 | 98.36 | 21,583 | 0.172 |

1. Initialization of the parameter grid: Parameters, such as the number of kernels, kernel size, sigma, lambda, gamma, and psi, are defined (Table 1).
2. Creation of the Gabor filter bank: A set of Gabor filters is generated based on the initialized parameters.
3. The filters from the Gabor filter bank are applied to the image using formula (2), and the filtered images are converted to grayscale.
4. For each filtered image, the SSIM values of the filtered image is calculated in comparison to the original image. SSIM can be calculated as follows:

$$\text{SSIM}(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\delta_{xy} + C_2}{\delta_x^2 + \delta_y^2 + C_2} \quad (3)$$

Table 5
Comparative performance of deep learning models on UCI dataset.

| Methods | Accuracy (%) | F1 score (%) | Training time (second) | Test time (second/sample) |
|-------------|--------------|--------------|------------------------|---------------------------|
| VGG16 | 87.74 | 87.00 | 2184 | 0.023 |
| Xception | 88.68 | 88.00 | 3408 | 0.020 |
| MobileNet | 93.40 | 92.00 | 1716 | 0.017 |
| DenseNet201 | 95.28 | 95.00 | 2328 | 0.019 |
| ViT | 98.11 | 97.48 | 6576 | 0.027 |

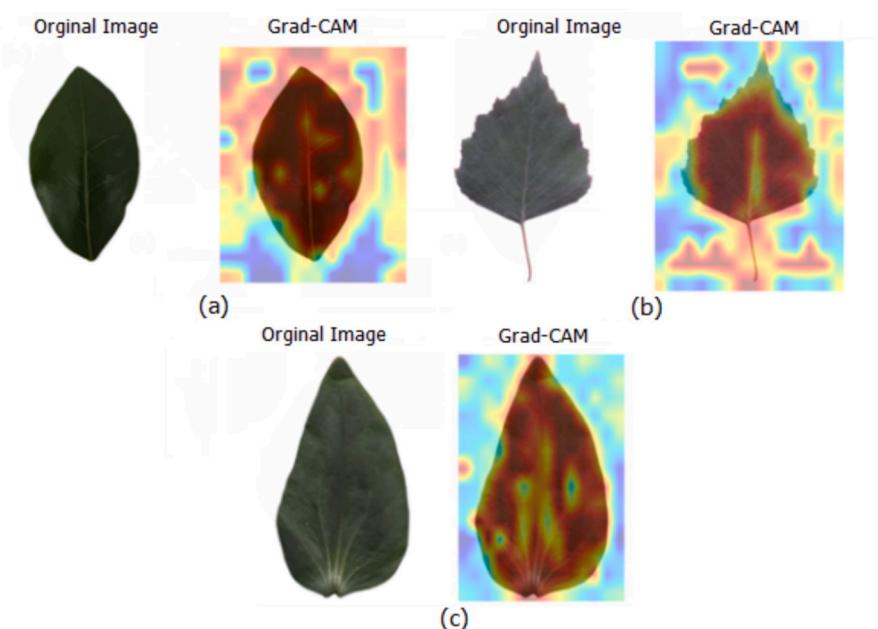


Fig. 12. Feature visualization of Layer-wise Relevance Propagation (LRP) from the ViT model using GradCAM on (a) Folio, (b) UCI, and (c) Safflower leaves.

where μ_x and μ_y represent the average values of the original image and the filtered image, respectively. Furthermore, δ_x^2 and δ_y^2 denote the variances of the original image and the filtered image, respectively. Here, C_1 and C_2 are constants used to prevent systematic errors when the denominator is zero. Furthermore, δ_{xy} indicates the covariance of the real image and the filtered image.

5. Comparison of the SSIM and updating of optimal parameters: If the current average SSIM is higher than the previously obtained best SSIM, then the optimal parameter set is updated.
6. Iteration over all parameter combinations: Steps 2 through 5 are repeated for each parameter combination in the grid.
7. Output of optimal parameters and corresponding SSIM: Upon completing iterations over all combinations, the best parameter set and its corresponding SSIM value are returned.

The parameter ranges listed in Table 1 were selected based on the image sizes and resolutions of the experimental datasets. The kernel size was selected to be smaller than the smallest detail of the image. The number of kernels refers to the number of Gabor filters applied in various orientations. Typical kernel sizes for medium-resolution images, such as (15,15) and (21,21), provide an excellent balance between capturing image details and maintaining computational efficiency. The wavelength (λ) was varied between 5 and 8 to ensure that the Gabor filter effectively captured the relevant texture features for leaf classification. Although smaller λ values can capture fine details, they can also amplify noise and less relevant textural information, which can hinder subsequent feature extraction and classification. By contrast, larger λ values, although effective in reducing noise, can smooth out crucial fine textures that are crucial for accurate classification. The standard deviation (σ) range was set between 2 and 5 to balance noise reduction with feature preservation based on empirical observations that indicated a tradeoff between these two factors. Finally, the aspect ratio (γ) was set at 0.25, 0.5, and 0.75 to enable the filter to detect both elongated structures, such as veins, and isotropic patterns, such as spots, reflecting the diverse morphological characteristics in the leaf datasets. Fig. 6 reveals four filtered images corresponding to kernels with different orientations. Fig. 7 illustrates a comparison between the original image and OLSF, revealing a clear texture but remaining insensitive to changes in lighting conditions. The OLSF image not only contains rich local structural

information such as spatial frequency, spatial location, and direction selectivity but also preserves the structural details of the original image.

4.4. OLSF-ViT for leaf recognition

ViT (Dosovitskiy et al., 2021) represents an advancement in computer vision and is a transformative approach to tasks such as image classification, object detection, and semantic segmentation. Unlike conventional convolutional networks, ViT is based on a transformer architecture. In the ViT model, the input image is divided into non overlapping patches, typically of size 16×16 , during the patch embedding phase. These patches are flattened and transformed into a sequence of patch embeddings. To retain the spatial information of the image, positional encoding is added to each patch embedding. This encoding is a positional index that helps the model understand the arrangement of patches in the original image. Next, the sequence of patch embeddings, with positional encodings, is inputted into a transformer encoder. Each layer of the transformer encoder is composed of two primary components: multi-headed self-attention (MSA) and a multi-layer perceptron (MLP). The MSA component allows the model to focus on different regions of the image simultaneously by applying the attention mechanism across multiple heads. This result enables the ViT to capture both local and global dependencies within the image. The MLP, composed of fully connected layers and activated using the Gaussian error linear unit, processes the attention mechanism output. Each transformer encoder layer also includes layer normalization and residual connections to ensure that the input to each layer is normalized and that residuals are added back to the output to improve gradient flow during training. After processing through L transformer layers, the final patch representations are passed to a classifier, which generates the classification output. Fig. 8 depicts the architecture of the ViT model.

In this study, we propose the OLSF-ViT method, which combines Gabor filters with a ViT model for enhancing leaf image classification. The OLSF-ViT method first preprocesses images to isolate leaf regions and subsequently applies the OLSF technique. The filtered images are then inputted into the ViT model for training and testing. We used the ViT-B/16 model with ImageNet-21 K pretrained weights and fine-tuned it on a leaf classification dataset. The model includes 12 transformer encoder layers with 12 attention heads each. Fig. 9 depicts the original image and its corresponding “last-layer attention map,” and Fig. 10

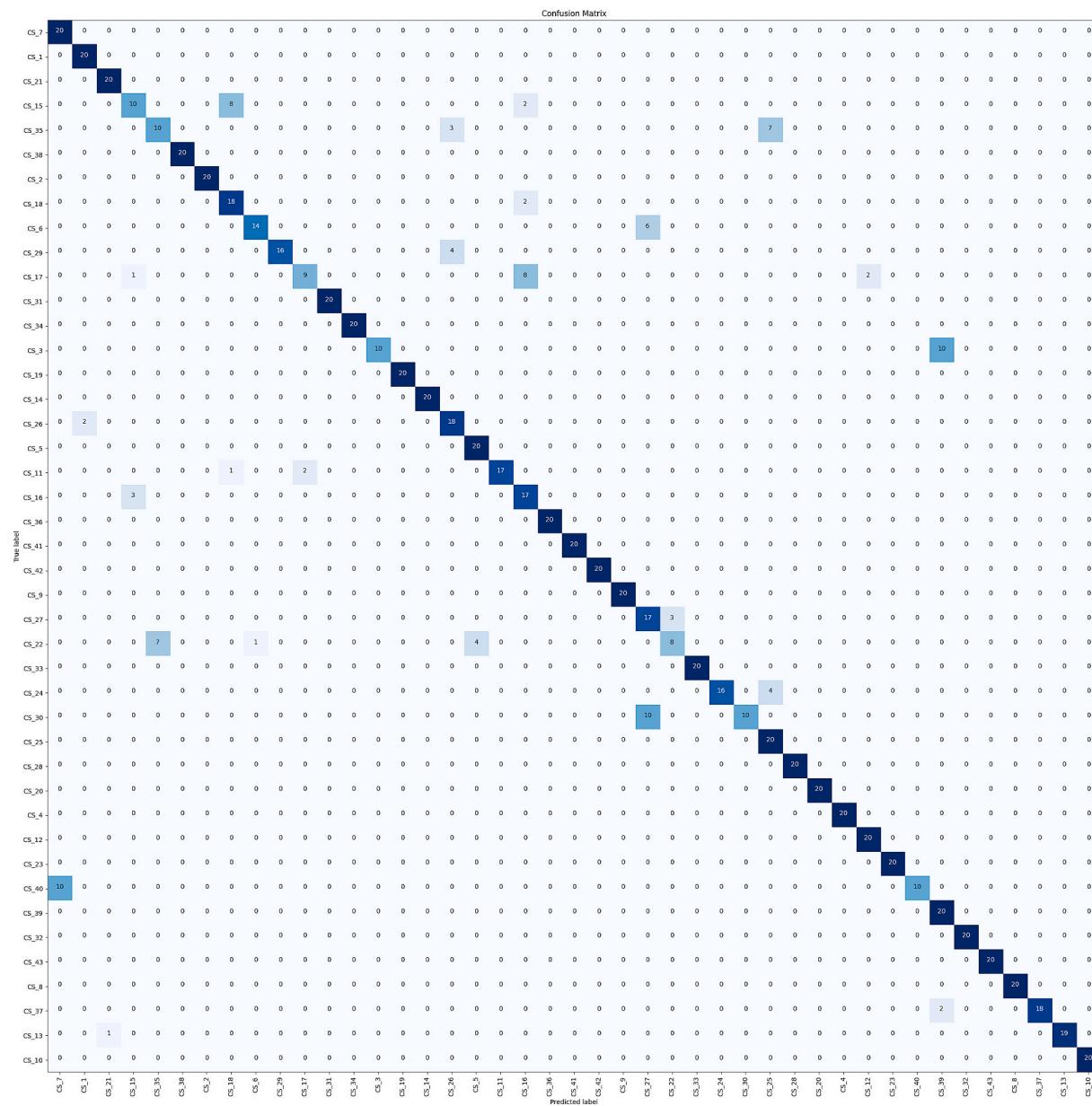


Fig. 13. Confusion matrix of the ViT model on the Safflower dataset.

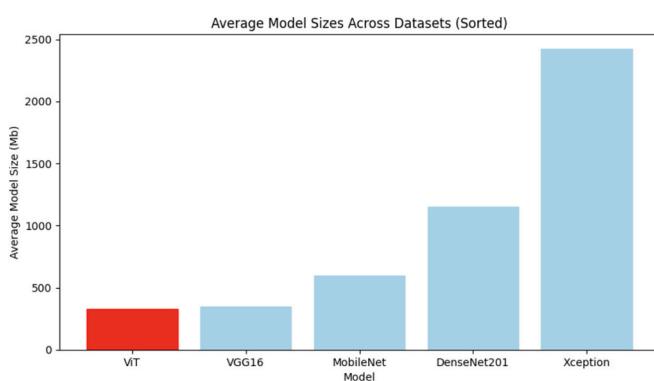


Fig. 14. Comparative of average deep learning model sizes (Mb) across datasets.

displays the OLSF image and its attention map. The comparison between these maps indicates that ViT incorporates more high-value regions from the OLSF image to improve classification performance.

5. Experimental results

The objective of these experiments was to evaluate the phenotypic diversity of safflower varieties by assessing the performance of deep learning-based classification methods on the proposed safflower leaf dataset. Additionally, we compared the effectiveness of the proposed OLSF-ViT algorithm against state-of-the-art models, such as ViT, VGG16, Xception, MobileNet, and DenseNet121, in identifying safflower varieties and other medicinal plant species. The experimental models were trained for 50 epochs by incorporating an early stopping criterion. This criterion was configured with a patience value of five epochs and a stopping threshold of 0.0, with accuracy as the performance metric. The K-fold cross-validation method was used to prevent overfitting and improve generalization. With k set to 4, each subset was used as the validation set, whereas the remaining k - 1 subsets were used for

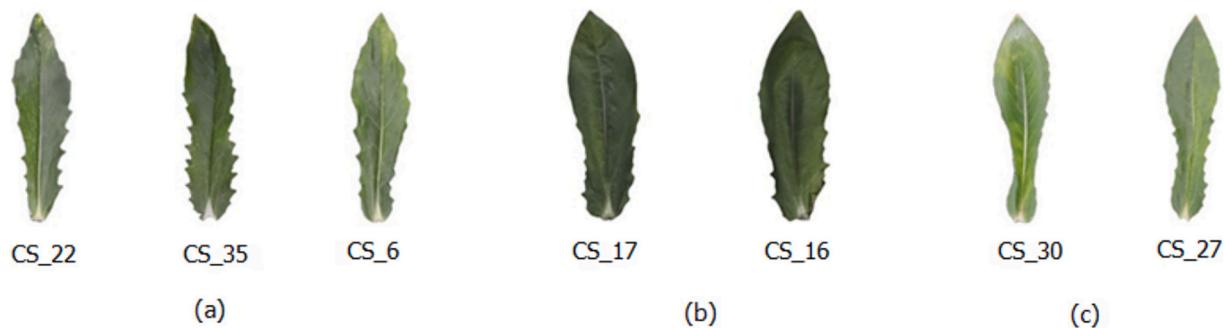


Fig. 15. Examples of misclassification: (a) Sample CS_22 misclassified as CS_35 and CS_6, (b) Sample CS_17 misclassified as CS_16, and (c) Sample CS_30 misclassified as CS_27.

Table 6
Optimal parameter sets for Gabor filters across all datasets.

| Dataset | Best parameters (number of kernels, kernel size, sigma, lambda, gamma, psi) | # SSIM values |
|-----------|---|---------------|
| Folio | (16, (15, 15), 3, 7, 0.25, 0) | 0.795 |
| UCI | (16, (15, 15), 3, 6, 0.75, 0) | 0.742 |
| Safflower | (4, (15, 15), 3, 7, 0.75, 0) | 0.774 |
| Leafsnap | (16, (15, 15), 3, 7, 0.75, 0) | 0.753 |

training. Computational experiments were conducted on the Google Colab platform using T4 GPUs to facilitate model development.

5.1. Datasets

In this study, the JNUSafflower dataset was used for safflower variety classification, whereas the Folio (Pudaruth, 2015), UCI Leaf (Silva and Maral, 2014), and LeafSnap (Kumar et al., 2012) datasets were used to classify the plant species. Fig. 11 reveals the sample images from these datasets.

The Folio dataset, designed to evaluate model performance with

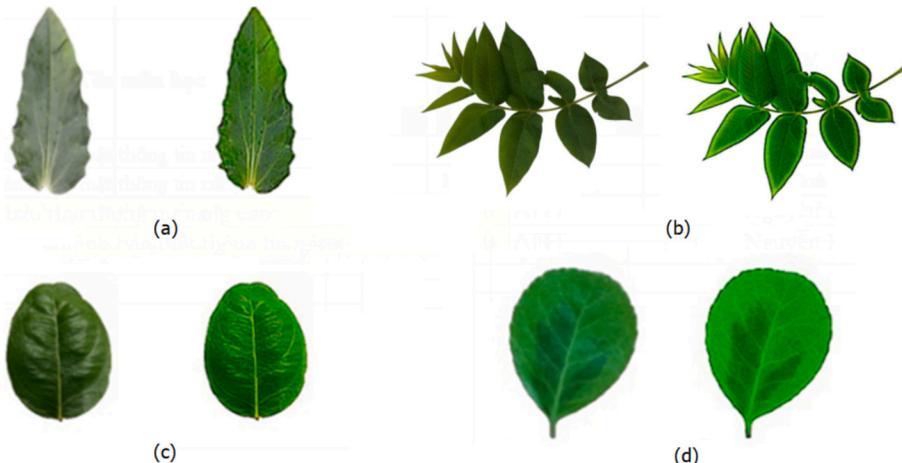


Fig. 16. Original images and their corresponding OLSF images across (a) Safflower, (b) LeafSnap, (c) Folio, and (d) UCI datasets.

Table 7
K-fold cross-validation accuracy of the methods using OLSF images on JNUSafflower dataset (%).

| Methods | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Mean accuracy | STD-DEV | Confidence interval (CI) |
|------------------|--------|--------|--------|--------|---------------|---------|--------------------------|
| OLSF-VGG16 | 43.02 | 47.44 | 53.95 | 38.37 | 45.70 | 6.63 | [35.15, 56.25] |
| OLSF-Xception | 47.44 | 46.51 | 46.51 | 42.55 | 45.75 | 2.18 | [42.28, 49.22] |
| OLSF-MobileNet | 58.60 | 55.34 | 55.81 | 43.25 | 53.25 | 6.82 | [42.40, 64.10] |
| OLSF-DenseNet201 | 51.16 | 57.44 | 61.63 | 47.20 | 54.36 | 6.42 | [44.14, 64.58] |
| OLSF-ViT | 87.44 | 85.35 | 78.37 | 78.14 | 82.33 | 4.78 | [74.72, 89.94] |

Table 8
K-fold cross-validation accuracy of the OLSF-ViT method on datasets.

| Dataset | Fold 1 | Fold 2 | Fold 3 | Fold 4 | Mean | STD-DEV | Confidence interval (CI) |
|--------------|--------|--------|--------|--------|-------|---------|--------------------------|
| Folio | 99.20 | 100.00 | 100.00 | 98.80 | 99.50 | 00.60 | [98.55, 100] |
| UCI Leaf | 97.80 | 100.00 | 97.50 | 96.60 | 98.00 | 01.40 | [95.77, 100] |
| JNUSafflower | 87.44 | 85.35 | 78.37 | 78.14 | 82.33 | 04.78 | [74.72, 89.94] |

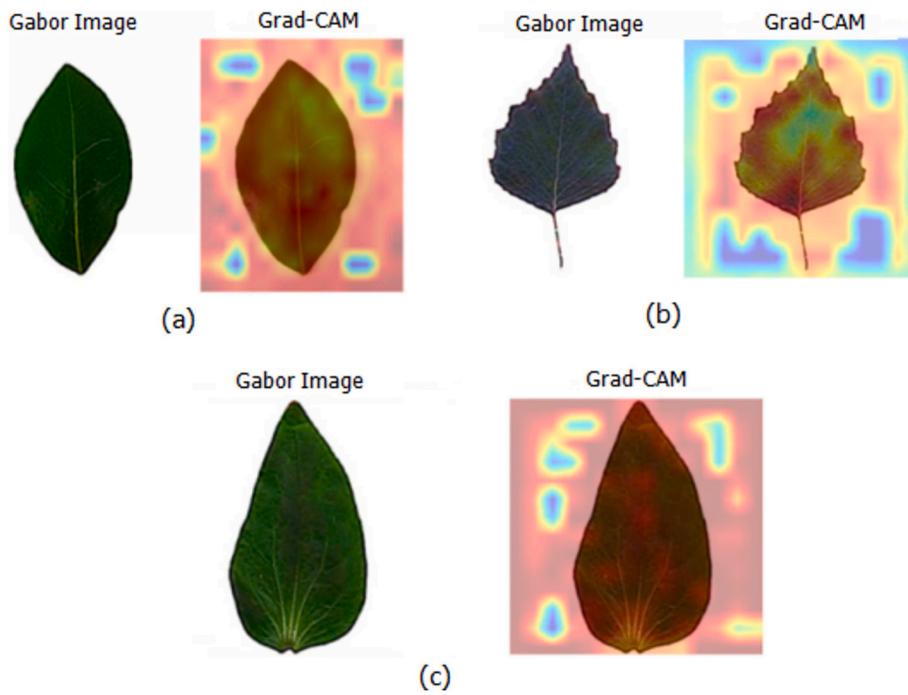


Fig. 17. Feature visualization of Layer-wise Relevance Propagation (LRP) from the OLSF-ViT model using GradCAM on (a) Folio, (b) UCI Leaf, and (c) Safflower leaves.

Table 9
Comparison of performance among models on datasets with OLSF images.

| Dataset | OLSF-VGG16 | OLSF-Xception | OLSF-MobileNet | OLSF-DenseNet201 | OLSF-ViT |
|-----------------------------|------------|---------------|----------------|------------------|----------|
| Folio Accuracy (%) | 96.09 | 94.53 | 96.88 | 98.44 | 100 |
| F1 Score (%) | 96.00 | 94.00 | 97.00 | 98.00 | 100 |
| UCI Leaf Accuracy (%) | 96.22 | 87.74 | 95.28 | 96.23 | 99.05 |
| F1 Score (%) | 96.00 | 87.00 | 95.00 | 96.00 | 98.67 |
| Safflower Leaf Accuracy (%) | 39.76 | 47.67 | 61.28 | 56.86 | 89.65 |
| F1 Score (%) | 37.00 | 45 | 59.00 | 55.00 | 89.19 |

limited data, typically includes approximately 20 images per class. By contrast, the UCI Leaf dataset contains lower-resolution images with fewer images per class, which poses a challenge because of its limited data volume. The LeafSnap dataset was collected outdoors using mobile phones and included 184 classes with each class averaging 42 samples. This dataset is complex and features blurry, grainy, shadowed, and under-exposed images. The safflower dataset underwent data augmentation to create a larger dataset, resulting in 24 images for training and 10 images for testing, in each sample. The training dataset contained 4128 images, whereas the test dataset contained 860 images. The Folio, UCI, and LeafSnap datasets were divided into 80 % for training and 20 % for testing. Table 2 lists the experimental dataset parameters.

5.2. Evaluation metrics

We evaluated the proposed method using several performance met-

rics, including accuracy (A), precision (P), recall (R), and F1-score (F1), providing an assessment of the model's classification capabilities. The calculation of these metrics for multiclass classification using macro-averaging is described by the following equations:

$$\text{AvgA} = \frac{1}{\text{classes}} \sum_i^{\text{classes}} A(i), A(i) = \frac{TP(i) + TN(i)}{TP(i) + TN(i) + FP(i) + FN(i)} \quad (4)$$

$$\text{AvgR} = \frac{1}{\text{classes}} \sum_i^{\text{classes}} R(i), R(i) = \frac{TP(i)}{FN(i) + TP(i)} \quad (5)$$

$$\text{AvgP} = \frac{1}{\text{classes}} \sum_i^{\text{classes}} P(i), P(i) = \frac{TP(i)}{TP(i) + FP(i)} \quad (6)$$

$$\text{AvgF}_1 = \frac{1}{\text{classes}} \sum_i^{\text{classes}} F_1(i), F_1(i) = 2 \times \frac{P(i) \times R(i)}{P(i) + R(i)} \quad (7)$$

where, with class i , TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

5.3. Result of experiment 1: proving the ViT is the best model for evaluating phenotypic diversity in safflower varieties and plant species classification compared with VGG16, Xception, MobileNet, and DenseNet201

We implemented five models, namely VGG16, Xception, MobileNet, DenseNet201, and ViT, on the original datasets. The experimental results are presented in Tables 3, 4, and 5, respectively.

Fig. 12 depicts the feature visualization of layer-wise relevance propagation (LRP) from the ViT model using GradCAM for (a) Folio, (b) UCI, and (c) safflower leaves. The classification accuracy of the deep learning models for safflower leaves ranged from 38.02 % to 88.02 %, with the ViT model achieving the highest accuracy of 88.02 %. Fig. 13 depicts the confusion matrix of the ViT model for the safflower dataset. These findings demonstrated that the classification of safflower varieties based on leaf features yields high accuracy, emphasizing the geometric diversity of safflower leaf characteristics. Furthermore, the ViT model achieved the highest accuracy for the Folio (98.85 %) and UCI (98.11 %)

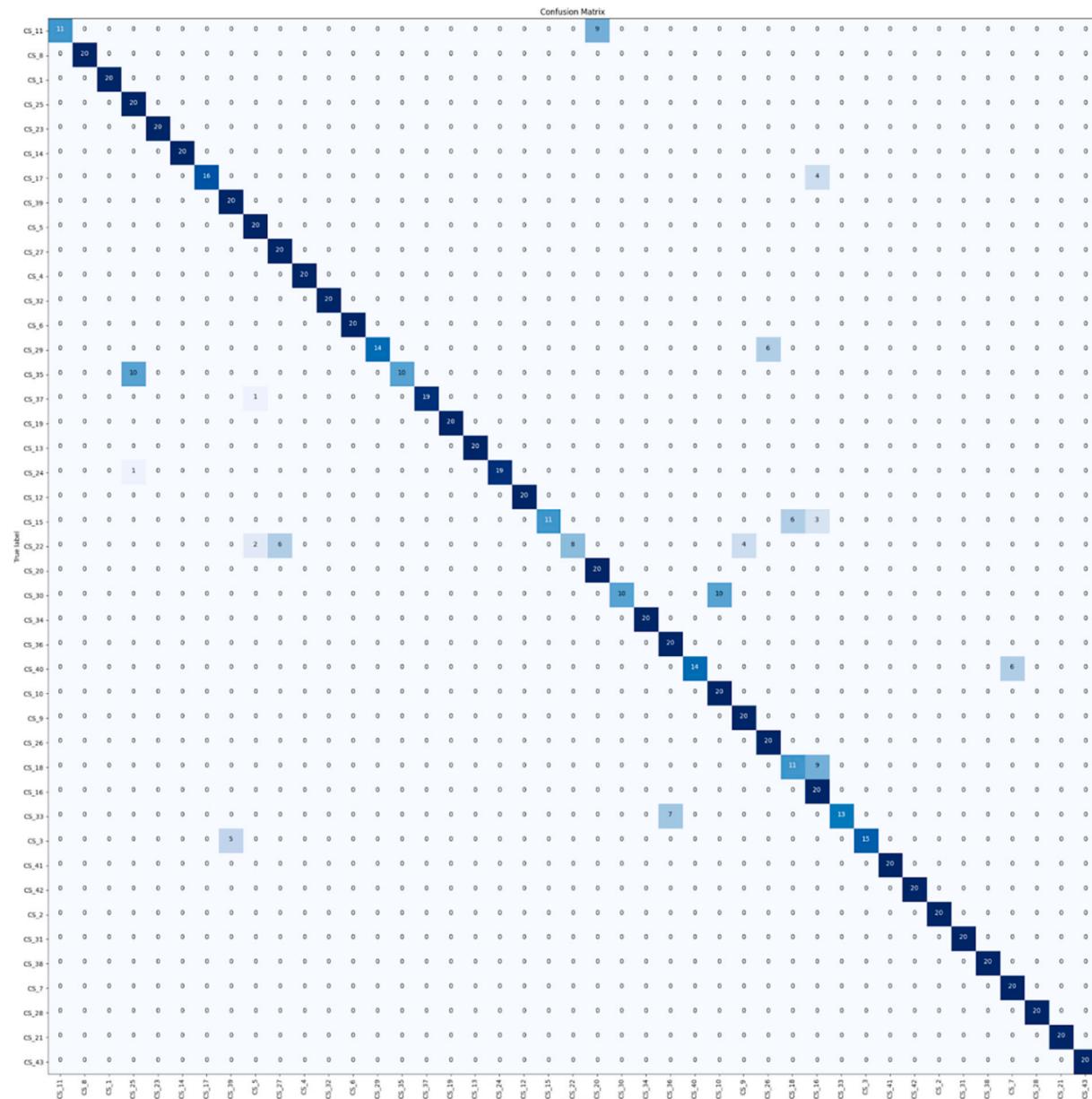


Fig. 18. Confusion matrix of the result of the OLSF-ViT model on the JNUSafflower dataset.

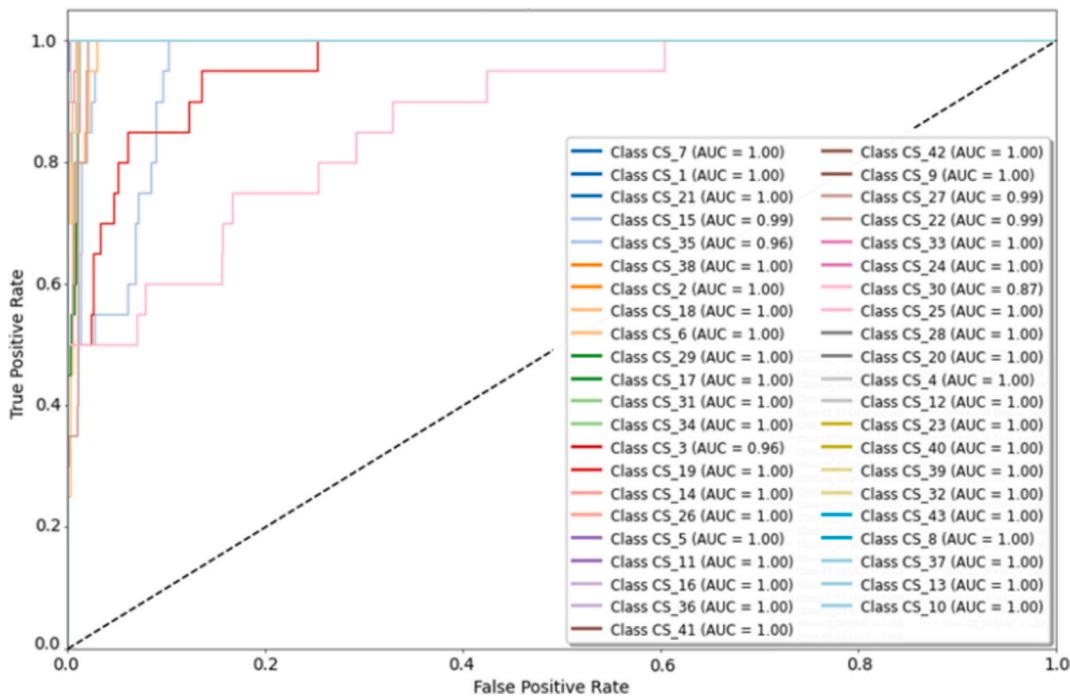


Fig. 19. AUC ROC of the ViT model on the JNUSafflower dataset.

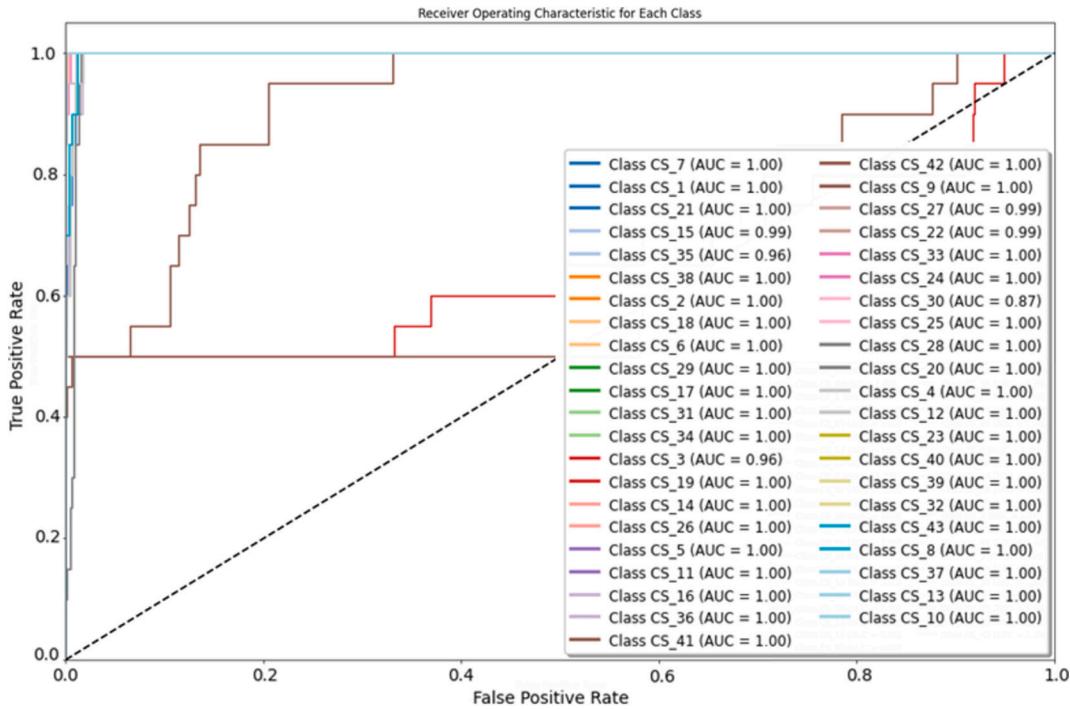


Fig. 20. AUC ROC of the OLSF-ViT model on the JNUSafflower dataset.

datasets. This consistent outperformance compared with other convolutional neural network architectures (VGG16, Xception, MobileNet, and DenseNet201) highlights the viability and generalizability of the ViT model for various medicinal leaf classification problems. As presented in Tables 3, 4, and 5, the ViT model achieved a longer training time than the other models. However, the testing time per sample was comparable to that of the other methods. Notably, the ViT model was the smallest in size, as illustrated in Fig. 14, which resulted in an efficient inference time and space usage. Therefore, the ViT model is highly

effective for classifying both safflower varieties and medicinal plant species and provides a tradeoff between high performance and computational efficiency.

In the ViT model, some samples were misclassified because of their similar shapes. For example (Fig. 15), sample CS_22 was misidentified as CS_6 or CS_35, CS_17 was occasionally misidentified as CS_16, and CS_30 was misclassified as CS_27. Therefore, clarifying the detailed vein patterns to distinguish similarly shaped leaves is crucial. In the next section, we present the experiments that evaluate the effectiveness of the OLSF

Table 10

Comparison of recognition accuracy of the ViT method using Gabor filtered images with different parameter sets on the Safflower leaf dataset.

| Name | Parameter set (number of kernels, kernel size, sigma, lambda, gamma, psi) | Accuracy (%) | F1 score (%) |
|---|---|--------------|--------------|
| Set 1 | (8, (15, 15), 3, 7, 0.5, 0) | 87.91 | 87.46 |
| Set 2 | (4, (15, 15), 3, 7, 0.5, 0) | 85.23 | 8434 |
| Optimal Parameter Set Determined Using SSIM (The parameter set of OLSF) | (4, (15, 15), 3, 7, 0.75, 0) | 89.65 | 89.19 |

feature extraction method in enhancing the accuracy of deep learning algorithms.

5.4. Result of experiment 2: demonstrating the effectiveness of the OLSF method in improving model accuracy for safflower varieties and medicinal plant species classification

In experiment 2, we conducted an extended analysis using deep learning methods (VGG16, Xception, MobileNet, DenseNet201, and ViT) on datasets enhanced using the OLSF method. The parameter grid outlined in Table 1 was used to identify the optimal parameter set for the OLSF method. Table 6 presents the best parameter set results after calculating and comparing the SSIM of the filters with different parameter sets by exhaustively testing all parameter combinations generated from Table 1. These parameters were used to create the corresponding OLSF image datasets. Fig. 16 revealsthe visual effects of applying Gabor filters with the optimal parameter sets to sample images from the safflower, LeafSnap, Folio, and UCI Leaf datasets.

Tables 7 and 8 present the k-fold cross-validation accuracy of the deep learning methods using OLSF images on the JNUSafflower dataset and the OLSF-ViT method on the three datasets. The proposed algorithm achieved the highest performance on the JNUSafflower dataset, with an average accuracy (mean) of 82.33 % and a confidence interval of [74.72 %, 89.94 %], demonstrating strong generalization and zero overfitting. The low standard deviation (STD-DEV) of the proposed method across datasets (from 0.6 % to 4.78 %) indicated stable results across various

Table 11

Comparison of the performance of the ViT model and the OLSF-ViT model on Folio dataset.

| Dataset | Accuracy (%) | F1 score (%) | Trainning time (second) | Test time (second/sample) |
|----------|--------------|--------------|-------------------------|---------------------------|
| ViT | 98.85 | 98.36 | 21,583 | 0.1720 |
| LBP-ViT | 99.21 | 99.21 | 21,136 | 0.9548 |
| HOG-ViT | 99.20 | 99.20 | 25,584 | 0.6944 |
| OLSF-ViT | 100.00 | 100.00 | 19,695 | 0.1650 |

Table 12

Comparison of the performance of the ViT model and the OLSF-ViT model on UCI Leafdataset.

| Dataset | Accuracy (%) | F1 score (%) | Trainning time (second) | Test time (second/sample) |
|----------|--------------|--------------|-------------------------|---------------------------|
| ViT | 98.11 | 97.48 | 6576 | 0.027 |
| LBP-ViT | 98.11 | 97.81 | 5244 | 0.8518 |
| HOG-ViT | 99.06 | 98.81 | 5416 | 0.5510 |
| OLSF-ViT | 99.05 | 98.67 | 5144 | 0.061 |

Table 13

Comparison of the performance of the ViT model and the OLSF-ViT model on JNUSafflower dataset.

| Dataset | Accuracy (%) | F1 score (%) | Trainning time (second) | Test time (second/sample) |
|---------------------|--------------|--------------|-------------------------|---------------------------|
| HOG – SVM | 29.19 | 27.00 | 3721 | 0.003 |
| HOG - Random Forest | 26.74 | 23.00 | 222 | 0.355 |
| ViT | 88.02 | 87.65 | 26,841 | 0.027 |
| LBP-ViT | 85.12 | 84.71 | 32,973 | 0.8548 |
| HOG-ViT | 84.30 | 83.43 | 28,497 | 0.5575 |
| OLSF-ViT | 89.65 | 89.19 | 26,662 | 0.059 |

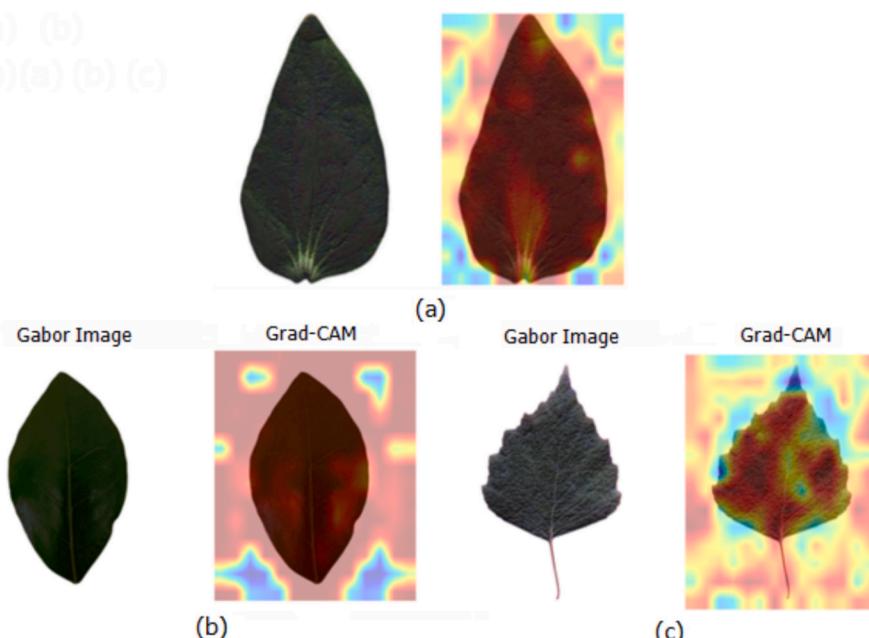


Fig. 21. Feature visualization of Layer-wise Relevance Propagation (LRP) from the LBP-ViT model using GradCAM on (a) Folio, (b) UCI Leaf, and (c) Safflower leaves.

Table 14

Comparison of the performance of the ViT model and the OLSF-ViT model on LeafSnap dataset.

| Dataset | Accuracy (%) | F1 score (%) | Trainning time (second) | Test time (second/sample) |
|----------|--------------|--------------|-------------------------|---------------------------|
| ViT | 95.34 | 96.72 | 22,809 | 0.0183 |
| OLSF-ViT | 95.22 | 96.68 | 24,731 | 0.0496 |

folds. This result confirmed a reasonable data division for k-fold validation and a minimal impact from small changes in the training data. For species classification, the model achieved near-perfect precision on the Folio dataset and high performance on the UCI dataset. In the safflower dataset, the accuracy was lower because of the challenge posed by the similarity in leaf structures among different varieties.

Fig. 17 reveals the feature visualization of LRP from the OLSF-ViT model using GradCAM for (a) Folio, (b) UCI, and (c) Safflower leaves. Table 9 lists the experimental results. The classification accuracy of the deep learning models on the OLSF images of the JNUSafflower dataset ranges from 39.76 % to 89.65 %, with the OLSF-ViT model achieving the highest accuracy of 89.65 % (Figs. 18, 19, and 20). As presented in Tables 3, 4, 5, and 9, the accuracy of the deep learning algorithms improved when processing OLSF images from the JNUSafflower, UCI Leaf, and Folio datasets. This improvement highlights the effectiveness of the OLSF features in enhancing the performance of deep learning models for plant classification.

To verify the results of selecting the best parameter set using the SSIM, we conducted additional experiments with other parameter sets, namely one set with eight directions and one set with gamma = 0.5, while keeping the remaining parameters unchanged. The results in Table 10 revealed that the ViT method with the image set filtered by the optimal parameter set based on SSIM achieved the highest accuracy compared with the manually selected parameter sets on the safflower leaf dataset. This phenomenon demonstrates the effectiveness of the OLSF filter in enhancing leaf features for safflower variety identification. We performed feature extraction using local binary patterns and a HOG and created LBP and HOG datasets for the images in the JNUSafflower, Folio, and UCI Leaf datasets.

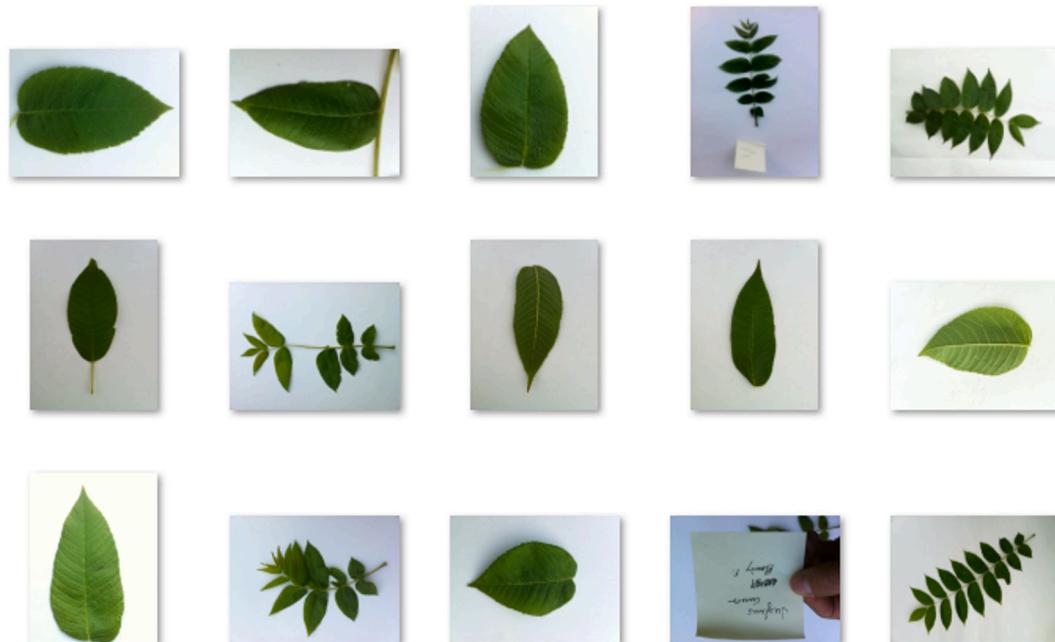


Fig. 22. Examples of the species *Juglans cinerea* illustrating various errors in leaf collection from the LeafSnap dataset, including single leaves, branches with multiple leaves, overlapping leaves, and occluded leaves.

Table 15

Related studies summary - plant classification accuracy.

| | Datasets | Best accuracy (%) | Method |
|--|--|-----------------------------------|--|
| Elhariri et al., 2014 | UCI Leaf | 92.65 | Linear Discriminant Analysis |
| Mishra et al., 2023 | UCI Leaf | 95.70 | ResNet50, PCA and LightGBM |
| Gu et al., 2021 | Folio | 97.90 | VGG16 |
| Kanda et al., 2021 | Folio | 98.75 | Deep Learning and Logistic regression |
| AnubhaPearline and Sathish Kumar, 2022 | Folio Flavia Swedish datasets | 96.38 99.58 100.00 | Bilateral network combined with machine learning classifier (MoDeNet + MLR) |
| Lv and Zhang, 2023 | Flavia Swedish Leaf | 99.30 99.52 | Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG), Principal Component Analysis (PCA), and Extreme Learning Machine (ELM) |
| Oppong et al., 2022 | Flavia PlantVillage | 99.83 91.25 | SURF and PCA |
| Wu et al., 2023 | PlantVillage Swedish LeafSnap | 96.28 99.47 91.29 | IMTD and VGG16 |
| Our proposed method | LeafSnap Folio UCI Leaf Safflower Leaf | 95.22 100.00 99.05 89.65 | Optimal Leaf Structure Features and Vision Transformer (OLSF-ViT) |

Fig. 21 depicts the feature visualization of LRP from the LBP-ViT model using GradCAM for (a) Folio, (b) UCI Leaf, and (c) Safflower leaves. Tables 11, 12, and 13 present the experimental results of the ViT method on the Folio, UCI Leaf, and JNUSafflower datasets using the original images or features extracted by LBP, HOG, and OLSF. The LBP-ViT and HOG-ViT methods achieved higher accuracy than ViT on the Folio and UCI Leaf datasets but performed worse on the JNUSafflower dataset. The OLSF-ViT method achieved the highest accuracy on the Folio dataset (100 %) and nearly matched HOG-ViT (99.06 %) with

99.05 % accuracy on the UCI Leaf dataset. For the JNUSafflower dataset, the proposed method achieved the highest accuracy (89.65 %) compared with ViT (88.02 %), LBP-ViT (85.12 %), HOG-ViT (84.30 %), and conventional machine learning methods, such as HOG-SVM (29.19 %) and HOG-random forest (26.74 %). These results confirmed the effectiveness of the OLSF method in enhancing the detailed leaf features and improving the performance of ViT in leaf classification. Table 14 presents the results for the LeafSnap dataset, where the OLSF-ViT method achieved a slightly lower accuracy (95.22 %) than that of ViT (95.34 %). However, the LeafSnap dataset contained many errors in leaf collection within the species, including single leaves, branches with multiple leaves, overlapping leaves, and occluded leaves (Fig. 22). This result highlights a limitation of the OLSF method and enhances features in non-leaf regions, which can affect classification when images exhibit considerable shape variations within the same class. Regarding the processing time, although the recognition processing time per sample and training time for the proposed method were longer than those for ViT, they were shorter than those for LBP-ViT and HOG-ViT. The balance between high accuracy and moderate processing time renders the proposed method a promising approach for plant leaf identification applications.

5.5. Discussion

This study evaluated the phenotypic diversity of safflower varieties by assessing the effectiveness of deep learning-based classification methods on a specialized safflower leaf dataset. The experimental results demonstrated that the leaf-based classification of safflower varieties can achieve high accuracy, with the ViT model attaining 88.02 % accuracy. However, some samples exhibited similar shapes within the same varieties, necessitating the enhancement of vein details to improve recognition accuracy. To address this challenge, we proposed an OLSF that accentuates the complex leaf details, including veins, textures, and frequency variations, to enhance the performance of deep learning models. The accuracy of the deep learning models on the OLSF-enhanced safflower dataset ranged from 39.76 % to 89.65 %, with the OLSF-ViT model achieving the highest accuracy of 89.65 %. Additionally, the OLSF-ViT model achieved accuracies of 100 %, 99.05 %, and 95.22 % for the Folio, UCI Leaf, and LeafSnap datasets, respectively. Compared with the existing literature summarized in Table 15, these results represent advancements in plant classification. Table 15 shows that although previous studies have applied image processing and deep learning techniques to plant classification with commendable accuracy, the superior performance of the proposed method highlights the advantage of integrating the OLSF approach with ViT over conventional methods. The model's enhanced feature extraction allows the identification of distinct leaf structures that are critical for differentiating similar-looking varieties. However, the proposed method exhibits certain limitations when applied to datasets collected under challenging conditions, including errors in leaf collection within species such as single leaves, branches with multiple leaves, overlapping leaves, and occluded leaves. Although the OLSF-ViT approach involves a longer processing time than ViT because of the OLSF extraction step, this additional time is manageable, especially considering the notable improvement in accuracy. The results of the proposed approach can be extended to botanical, agricultural, and medicinal plant research. In botany and agriculture, the OLSF-ViT model can support biodiversity conservation efforts by providing a reliable tool for accurate plant

identification and aiding in the documentation and preservation of plant species. In medicinal plant research, this model can be used to identify and classify plants with therapeutic properties. By examining various plant species and their morphological traits, researchers can identify new sources of medicinal compounds that support innovation in natural medicine and the development of new treatments.

6. Conclusions

This study advances the accurate identification of safflower varieties to assess their genetic diversity and supports biodiversity conservation using deep-learning-based on leaf characteristics. We evaluated state-of-the-art deep learning methods and introduced a novel approach, OLSF-ViT, which combines the ViT model with an OLSF. This feature optimally applies Gabor filters with the SSIM metric to enhance the leaf texture across datasets. The OLSF-ViT model achieved excellent accuracy scores of 100 %, 99.05 %, 95.22 %, and 89.65 % on the Folio, UCI Leaves, LeafSnap, and JNUSafflower datasets, respectively. This study highlights the potential of OLSF-ViT for automatic plant classification and details the applications in plant science, herbal medicine, and biodiversity preservation. Additionally, the success of this method in leaf classification indicates its adaptability to other plant structures such as flowers, stems, roots, and fruits. Consequently, numerous comprehensive plant identification tools can be developed to enhance decision-making processes and advance our understanding of plant diversity and ecological significance. These findings can preserve the genetic diversity of cultivated plants and enhance the genetic foundation of current breeding programs.

CRediT authorship contribution statement

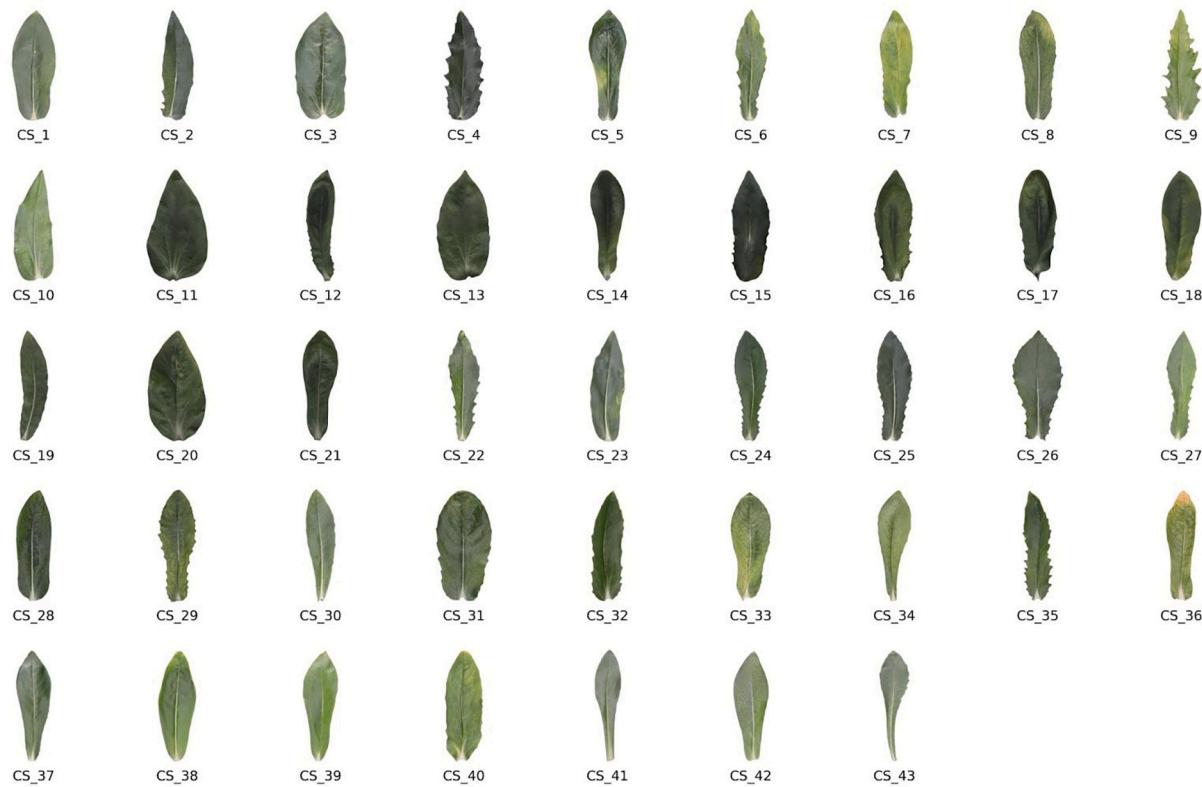
Hoang Thien Van: Writing – review & editing, Writing – original draft, Validation, Supervision, Conceptualization, Software, Methodology. **Phuong Thuy Khuat:** Writing – review & editing, Visualization, Validation, Software. **Trang Van:** Writing – review & editing, Validation, Software. **Thai Thanh Tuan:** Data curation, Conceptualization, Visualization, Validation. **Yong Suk Chung:** Writing – review & editing, Resources, Investigation, Formal analysis, Data curation, Conceptualization, Supervision.

Data availability

Van Thien, Hoang; Yong Suk, Chung (2024), “JNUSafflower”, Mendeley Data, V1, doi: [10.17632/4rs446vg8t.1](https://doi.org/10.17632/4rs446vg8t.1). Direct URL to the data: <https://data.mendeley.com/datasets/4rs446vg8t/1>.

Acknowledgments

This research was supported by a grant from the Standardization and Integration of Resource Information for Seed-Clusters in the Hub-Spoke Material Bank program (Project No. PJ01587004), Rural Development Administration, Republic of Korea. This study was part of a collaborative project between the Ho Chi Minh City University of Technology (HUTECH University), Vietnam, and the Subtropical Horticultural Research Institute (SHRI), Jeju National University, Republic of Korea. The authors express their gratitude to HUTECH University and SHRI-Jeju National University for their support in this study.

Appendix A. Leaf image list of 43 SAFFLOWER (*Carthamus tinctorius* L.) varieties

Appendix B. Information on SAFFLOWER (*Carthamus tinctorius* L.) varieties

| No | IT/temporary number with prefix | Resource name | Origin |
|-------|---------------------------------|------------------|--------|
| CS_1 | K185908 | PI 407610 | TUR |
| CS_2 | K185229 | PI 292000 | ISR |
| CS_3 | IT202728 | Local | UZB |
| CS_4 | IT300330 | 가시총화 | KOR |
| CS_5 | K186681 | XJ-040 | CHN |
| CS_6 | K186725 | 청수총화 | KOR |
| CS_7 | IT202729 | Tashkent-51 | UZB |
| CS_8 | IT209544 | UZB-1998-806,674 | KAZ |
| CS_9 | IT209507 | UZB-1998-806,638 | UZB |
| CS_10 | IT209513 | UZB-1998-806,644 | AFG |
| CS_11 | K184480 | PI 181866 | SYR |
| CS_12 | K185025 | PI 253899 | SYR |
| CS_13 | K185104 | PI 262423 | AUS |
| CS_14 | IT202721 | Local | KAZ |
| CS_15 | IT202729 | Tashkent-51 | UZB |
| CS_16 | IT209520 | UZB-1998-806,651 | ARM |
| CS_17 | IT209532 | UZB-1998-806,663 | TKM |
| CS_18 | IT209541 | UZB-1998-806,672 | KAZ |
| CS_19 | IT209542 | UZB-1998-806,673 | UZB |
| CS_20 | IT209560 | UZB-1998-806,691 | AZE |
| CS_21 | IT209563 | UZB-1998-806,694 | PAK |
| CS_22 | IT209534 | UZB-1998-806,665 | KAZ |
| CS_23 | IT209538 | UZB-1998-806,669 | KAZ |
| CS_24 | K171322 | 가시총화 | KOR |
| CS_25 | K186073 | PI 407624 | TUR |
| CS_26 | IT209552 | UZB-1998-806,683 | IRN |
| CS_27 | IT209557 | UZB-1998-806,688 | IRN |
| CS_28 | K185781 | PI 369849 | RUS |
| CS_29 | K184495 | PI 181866 | SYR |
| CS_30 | K184931 | PI 251291 | JOR |
| CS_31 | IT321102 | 가시총화 | KOR |
| CS_32 | IT321105 | 가시총화 | KOR |

(continued on next page)

(continued)

| No | IT/temporary number with prefix | Resource name | Origin |
|-------|---------------------------------|-----------------|--------|
| CS_33 | K184983 | PI 253538 | ARM |
| CS_34 | K185012 | PI 253759 | IRQ |
| CS_35 | IT300331 | 가시총화 | KOR |
| CS_36 | K185021 | PI 253895 | SYR |
| CS_37 | K185033 | BJ-1233 | AFG |
| CS_38 | K185222 | PI 286199 | KWT |
| CS_39 | K185307 | PI 304595 | AFG |
| CS_40 | K185719 | PI 312275 | HUN |
| CS_41 | IT252163 | MMR-JYH-2010-98 | MMR |
| CS_42 | K185881 | PI 406001 | IRN |
| CS_43 | K185895 | PI 406015 | IRN |

References

- Abdipour, Moslem, Younessi-Hmazekhanlu, Mehdi, Ramazani, Seyyed Hamid Reza, Hassanomidi, Amir, 2019. Artificial neural networks and multiple linear regression as potential methods for modeling seed yield of safflower (*Carthamus tinctorius* L.). *Ind. Crop. Prod.* 127, 185–194. ISSN 0926-6690. <https://doi.org/10.1016/j.indcrop.2018.10.050>.
- AnubhaPearline, S., Sathiesh Kumar, V., 2022. Performance analysis of real-time plant species recognition using bilateral network combined with machine learning classifier. *Ecol. Inform.* 67, 101492. ISSN 1574-9541. <https://doi.org/10.1016/j.ecoinf.2021.101492>.
- Bello, I., et al., 2021. Revisiting ResNets: improved training and scaling strategies. *Adv. Neural Inf. Proces. Syst.* 27, 22614–22627. <https://doi.org/10.48550/arXiv.2103.07579>.
- Ceyhan, M., Kartal, Y., Özkan, K., et al., 2024. Classification of wheat varieties with image-based deep learning. *Multimed. Tools Appl.* 83, 9597–9619. <https://doi.org/10.1007/s11042-023-16075-5>.
- Chaki, J., Parekh, R., 2012. Plant leaf recognition using Gabor filter. *Int. J. Comput. Appl.* 56 (10).
- Chaki, J., Parekh, R., Bhattacharya, Samar, 2015. Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recogn. Lett.* 58, 61–68. ISSN 0167-8655. <https://doi.org/10.1016/j.patrec.2015.02.010>.
- Chi, Zheru, Houqiang, Li, Wang, Chao, 2003. Plant species recognition based on bark patterns using novel Gabor filter banks. In: International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003, Nanjing, 2, pp. 1035–1038. <https://doi.org/10.1109/ICNNSP.2003.1281045>.
- Conde, M.V., Turgutlu, K., 2021. Exploring Vision Transformers for Fine-grained Classification [arXiv:2106.10587](https://arxiv.org/abs/2106.10587).
- Cope, J.S., Remagnino, P., Barman, S., Wilkin, P., 2010. Plant texture classification using Gabor co-occurrences. In: Bebis, G., et al. (Eds.), Advances in Visual Computing. ISVC 2010. Lecture Notes in Computer Science, vol. 6454. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17274-8_65.
- DeeptiBarhate, Sunil Pathak, Dubey, Ashutosh Kumar, 2023. Hyperparameter-tuned batch-updated stochastic gradient descent: plant species identification by using hybrid deep learning. *Ecol. Inform.* 75, 102094. ISSN 1574-9541. <https://doi.org/10.1016/j.ecoinf.2023.102094>.
- Dhakshayani, J., Surendiran, B., 2023. GF-CNN: an enhanced deep learning model with Gabor filters for maize disease classification. *SN Comput. Sci.* 4, 538. <https://doi.org/10.1007/s42979-023-01988-7>.
- Diago, María P., Fernandes, A.M., Millan, B., Tardaguila, J., Melo-Pinto, P., 2013. Identification of grapevine varieties using leaf spectroscopy and partial least squares. *Comput. Electron. Agric.* 99, 7–13. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2013.08.021>.
- Dosovitskiy, et al., Jun. 03, 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv. <https://doi.org/10.48550/arXiv.2010.11929>.
- Dyrmann, Mads, Karstoft, Henrik, SkovMidtby, Henrik, 2016. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. ISSN 1537-5110. <https://doi.org/10.1016/j.biosystemseng.2016.08.024>.
- Elhariri, E., El-Bendary, N., Hassanien, A.E., 2014. Plant classification system based on leaf features. In: 2014 9th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, pp. 271–276. <https://doi.org/10.1109/ICCES.2014.7030971>.
- Ghosh, S., Singh, A., Kumar, S., 2024 Jul 1. HPB3C-3PG algorithm: a new hybrid global optimization algorithm and its application to plant classification. *Ecol. Inform.* 81, 102581.
- Gu, J., Yu, P., Lu, X., Ding, W., 2021. Leaf species recognition based on VGG16 networks and transfer learning. In: 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, pp. 2189–2193. <https://doi.org/10.1109/IAEAC50856.2021.9390789>.
- Ishak, Asnor Juraiza, Hussain, Aini, Mustafa, Mohd Marzuki, 2009. Weed image classification using Gabor wavelet and gradient field distribution. *Comput. Electron. Agric.* 66 (1), 53–61. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2008.12.003>.
- Kanda, P.S., Xia, K., Sanusi, O.H., 2021. A deep learning-based recognition technique for plant leaf classification. *IEEE Access* 9, 162590–162613. <https://doi.org/10.1109/ACCESS.2021.3131726>.
- Karadeniz, A.T., Başaran, E., Çelik, Y., 2024. Classification of walnut dataset by selecting CNN features with whale optimization algorithm. *Multimed. Tools Appl.* 83, 77061–77076. <https://doi.org/10.1007/s11042-024-18586-1>.
- Kaya, A., Keceli, A.S., Catal, C., Yalic, H.Y., Temucin, H., Tekinerdogan, B., 2019. Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* 158, 20–29. <https://doi.org/10.1016/j.compag.2019.01.041>.
- Koklu, Murat, Cinar, Ilkay, Taspinar, Yavuz Selim, 2021. Classification of rice varieties with deep learning methods. *Comput. Electron. Agric.* 187, 106285. ISSN 0168-1699. <https://doi.org/10.1016/j.compag.2021.106285>.
- Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V., 2012. Leafsnap: a computer vision system for automatic plant species identification. In: Computer Vision-ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part II 12. Springer, Berlin Heidelberg, pp. 502–516.
- Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: plant identification with convolutional neural networks. In: 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, pp. 452–456. <https://doi.org/10.1109/ICIP.2015.7350839>.
- Lee, Sue Han, Chan, Chee Seng, Mayo, Simon Joseph, Remagnino, Paolo, 2017. How deep learning extracts and learns leaf features for plant classification. *Pattern Recogn.* 71, 1–13. ISSN 0031-3203. <https://doi.org/10.1016/j.patcog.2017.05.015>.
- Li, H.-A., et al., 2020. Medical image coloring based on Gabor filtering for internet of medical things. *IEEE Access* 8, 104016–104025. <https://doi.org/10.1109/ACCESS.2020.2999454>.
- Lv, Zhimin, Zhang, Zhibin, 2023. Research on plant leaf recognition method based on multi-feature fusion in different partition blocks. *Digit. Signal Process.* 134, 103907. ISSN 1051-2004. <https://doi.org/10.1016/j.dsp.2023.103907>.
- Mayerhofer, R., Archibald, C., Bowles, V., Good, A.G., 2010 Apr. Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. *Genome* 53 (4), 266–276. <https://doi.org/10.1139/g10-002>. PMID: 20616858.
- Mishra, V., Sharma, V., Mishra, U., 2023. A hybrid approach for leaf classification using machine learning and deep learning. In: 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, pp. 1589–1593. <https://doi.org/10.1109/ICCPCT58312023.2023.10245548>.
- Oppong, S.O., Twum, F., Hayfron-Acuah, J.B., Missah, Y.M., 2022 Sep 27. A novel computer vision model for medicinal plant identification using log-Gabor filters and deep learning algorithms. *Comput. Intell. Neurosci.* 2022, 1189509. <https://doi.org/10.1155/2022/1189509>. PMID: 36203732; PMCID: PMC9532088.
- Pudaruth, S., 2015. Folio. In: UCI Machine Learning Repository. <https://doi.org/10.24432/CSTW3F>.
- Puri, D., Kumar, A., Virmani, J., et al., 2022. Classification of leaves of medicinal plants using laws' texture features. *Int. J. Inf. Technol.* 14, 931–942. <https://doi.org/10.1007/s41870-019-00353-3>.
- Rother, Carsten, Kolmogorov, Vladimir, Blake, Andrew, 2004. "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* 23 (3), 309–314. <https://doi.org/10.1145/1015706.1015720> (August 2004).
- Saleem, G., Akhtar, M., Ahmed, N., Qureshi, W.S., 2019. Automated analysis of visual leaf shape features for plant classification. *Comput. Electron. Agric.* 157, 270–280.
- Silva, Pedro, Maral, Andr, 2014. UCI Leaf Dataset. <https://doi.org/10.24432/C53C78>.
- Sun, Y., Liu, Y., Wang, G., Zhang, H., May 2017. Deep learning for plant identification in natural environment. *Comput. Intell. Neurosci.* 2017, 7361042. <https://doi.org/10.1155/2017/7361042>.
- Suwarningsih, W., Khotimah, P.H., Rozie, A.F., Arisal, A., Riswantini, D., Nugraheni, E., Munandar, D., Kirana, R., February 2022. Ide-cabe: chili varieties identification and classification system based leaf. *Bull. Electr. Eng. Inform.* 11 (1), 445–453. ISSN: 2302-9285. <https://doi.org/10.11591/eei.v11i1.3276>.
- Van Hieu, N., Hien, N.L.H., Van Huy, L., Tuong, N.H., Thoa, P.T.K., 2023. PlantKViT: a combination model of vision transformer and KNN for forest plants classification. *JUCS: J. Univ. Comput. Sci.* 29 (9). <https://doi.org/10.3897/jucs.94657>.
- Venkatesh, N.Y., Hegde, S.U., S, S, 2021. Fine-tuned MobileNet classifier for classification of strawberry and cherry fruit types. In: 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1–8. <https://doi.org/10.1109/ICCI50826.2021.9402444>.

- Wu, Hao, Fang, Lincong, Yu, Qian, Yuan, Jingrong, Yang, Chengzhan, 2023. Plant leaf identification based on shape and convolutional features. *Expert Syst. Appl.* 219, 119626. ISSN 0957-4174. <https://doi.org/10.1016/j.eswa.2023.119626>.
- Wu, Hao, Fang, Lincong, Yu, Qian, Yang, Chengzhan, 2024. Composite descriptor based on contour and appearance for plant species identification. *Eng. Appl. Artif. Intell.* 133 (Part C), 108291. ISSN 0952-1976. <https://doi.org/10.1016/j.engappai.2024.108291>.
- Xing, Zhenyu, Zhang, Zhenguo, Wang, Yunze, Xu, Peng, Guo, Quanfeng, Zeng, Chao, Shi, Ruimeng, 2024. SDC-DeepLabv3+: lightweight and precise localization algorithm for safflower-harvesting robots. *Plant Phenomics*. <https://doi.org/10.34133/plantphenomics.0194>.
- Yang, Chengzhan, 2021. Plant leaf recognition by integrating shape and texture features. *Pattern Recogn.* 112, 107809. ISSN 0031-3203. <https://doi.org/10.1016/j.patcog.2020.107809>.
- Yu, J.K., Chang, S., Han, G.D., et al., 2023. Implication of high variance in germplasm characteristics. *Sci. Rep.* 13, 515. <https://doi.org/10.1038/s41598-023-27793-z>.
- Zhang, Shanwen, Huang, Wenzhun, Huang, Yu-an, Zhang, Chuanlei, 2020. Plant species recognition methods using leaf image: overview. *Neurocomputing* 408, 246–272. ISSN 0925-2312. <https://doi.org/10.1016/j.neucom.2019.09.113>.
- Zhou, X., Tang, L., Xu, Y., Zhou, G., Wang, Z., Jan. 2014. Towards a better understanding of medicinal uses of *Carthamus tinctorius* L. in traditional Chinese medicine: a phytochemical and pharmacological review. *J. Ethnopharmacol.* 151 (1), 27–43. <https://doi.org/10.1016/j.jep.2013.10.050>.