IDENTIFYING FAGACEAE SPECIES IN TAIWAN USING LEAF IMAGES

C.-H. Lee, H.-C. Hsu, C.-K. Yang, M.-J. Tsai, Y.-F. Kuo

ABSTRACT. *Fagaceae is the second-largest woody plant family in Taiwan and has considerable economic and ecological value. Identifying Fagaceae species is essential for forest managers and forestry technical personnel. In this study, ten Fagaceae species were distinguished using image processing and machine learning. For each species, 100 leaf images were collected using flatbed scanners. The morphological, marginal, color, and textural traits of the leaves were then quantified. A genetic algorithm was next applied to identify the traits that are essential for species identification. Support vector machine classifiers were developed to identify the species using the selected traits as the inputs. The results indicated that the proposed approach had an accuracy of 92.8%.*

Keywords. *Fagaceae, Image processing, Machine learning, Species identification, Trait selection.*

Fagaceae is an evergreen woody plant family with numerous species that serves a wide range of ecological and economic functions in Taiwan. The seeds of Fagaceae trees supply food to vertebrates in the forest (Kuo, 2003; Lee, 2004). The wood of these trees is used for manufacturing timber products (Ogata, 2008). Recently, the second metabolites in Fagaceae species have been found to have antiulcer, anti-inflammatory, and anti-HIV properties (Chang et al., 2016; Singh, 2018). Being a subtropical island with high mountains, Taiwan has a high level of plant species diversity. Fifty-nine Fagaceae species have been identified in Taiwan (Liao, 1996). The Fagaceae species existing in the wild need to be identified so that these natural resources can be properly protected and used (Cheng, 2005). Leaf morphology serves as a major basis for distinguishing and identifying plant species (Bruschi et al., 2003; Tomlinson et al., 1991). This research aims to identify ten Fagaceae species (fig. 1) using leaf images and image-based approaches.

Genetic marker-based methods have been extensively applied to Fagaceae for examining genetic profiles and phylogeny and for identifying species (Kremer et al., 2012; Coutinho et al., 2014). Fagaceae genera such as *Quercus* and *Castanopsis*, which are rich and endemic in species, have been widely studied. Ueno and Tsumura (2008) used expressed sequence tags and simple sequence repeats (EST-SSR) markers (present in complementary DNA) to identify the genetic diversity in three *Quercus* species. Pencakowski et al. (2018) used EST-SSR and short tandem repeat (present in genomic DNA) analyses to determine the genetic markers for 28 *Quercus* species. Ueno et al. (2009) developed EST-SSR markers for distinguishing the varieties of *Castanopsis sieboldii* and its congener species. Aoki et al. (2014) used EST-SSR markers to investigate the genetic differences between closely related species of *Castanopsis*. Although these studies provided precise discrimination results, marker-based methods are time-consuming and require sophisticated procedures and instruments. Hence, these methods are not preferred by forest managers and forestry technical personnel for field practice.

Image-based approaches, by contrast, are rapid and nondestructive and can be performed using portable devices. With recent advances in digital imaging, some subtle details of leaves (e.g., vein and trichomes) that are difficult to observe with the naked eye can be characterized using image-based approaches. These details can then be used with machine learning classifiers to identify plant species. Harish (2013) recognized six herbal plants (*Syzygium cumin*, *Azadirachta indica*, *Eucalyptus oblique*, *Hibiscus rosa-sinensis*, *Tinospora cordifolia*, and *Piper betle*) in color images by determining the morphological features (e.g., aspect ratio, rectangularity, and perimeter ratio) of the leaves and using support vector machine (SVM) and probabilistic neural network classifiers. Aakif and Khan (2015) quantified the morphological features and Fourier descriptors in color images of leaves from 14 fruit-tree species and identified the species using artificial neural network (ANN) classifiers. Tang et al. (2015) captured the patterns of different types of tea leaves (*Camellia sinensis*) using local binary pattern and gray-level co-occurrence matrix features and then classified the leaf types using back-propagation neural network classifiers. Another study recognized the leaves of herbal plants (*Ocimum tenuiflorum*, *Murraya koenigii*, *Solanum trilobatum*, *A. in-*

Submitted for peer review in January 2019 as manuscript number ITSC 13302; approved for publication as a Research Article by the Information Technology, Sensors, & Control Systems Community of ASABE in June 2019.

The authors are **Cheng-Hao Lee,** Research Assistant, and **Hao-Chun Hsu,** Doctoral Candidate, Department of Biomechatronics Engineering, **Chih-Kai Yang,** Research Fellow, Experimental Forest, College of Bio-Resources and Agriculture, **Ming-Jer Tsai,** Professor, School of Forestry and Resource Conservation, and **Yan-Fu Kuo,** Associate Professor, Department of Biomechatronics Engineering, National Taiwan University, Taipei, Taiwan. **Corresponding author:** Yan-Fu Kuo, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan; phone: +886-2-33665329; e-mail: ykuo@ntu.edu.tw.

Vol. 62(5): 1055-1063 ISSN 2151-0032 https://doi.org/10.13031/trans.13302

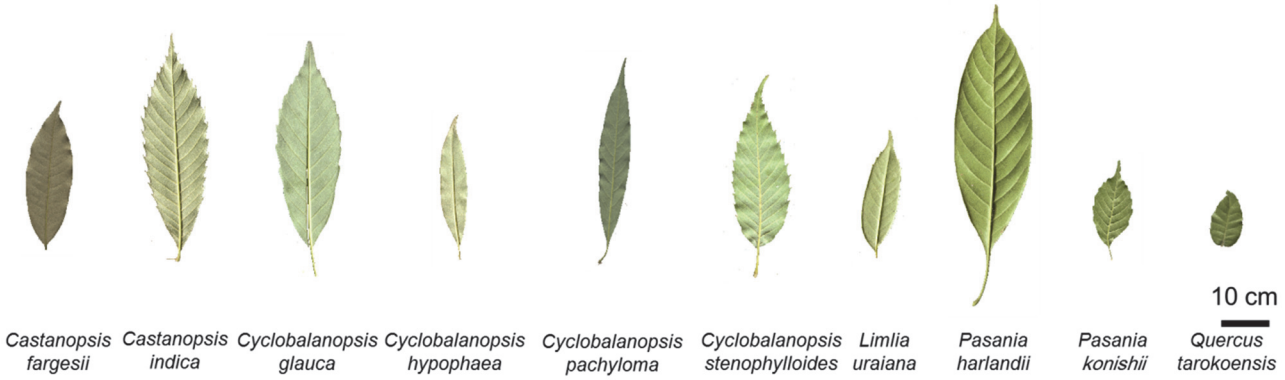

Figure 1. Images of the ten Fagaceae species.

dica, and *H. rosa-sinensis*) using color traits and ANN classifiers (Janani and Gopal, 2013).

In recent years, convolutional neural networks (CNNs) have emerged as a powerful tool for species identification. Compared with conventional image-based methods, CNN approaches do not require handcrafted features (Wäldchen and Mäder, 2018). Instead, CNNs can learn essential characteristics for differentiating between species directly from raw images. Lee et al. (2015) identified 44 plant species using a CNN and gained insight into the discriminative features learned by the CNN. Lee et al. (2017) distinguished 76 plants using color leaf images, leaf patches, and hybrid fusion of global and local CNNs. Grinblat et al. (2016) identified three legume species and unveiled discerning vein patterns by replacing a conventional feature extraction module with a deep CNN. Yang et al. (2019) distinguished three very similar *Cinnamomum* species using leaf images, leaf patches, deep CNNs, and score fusion. CNNs are multilayer perceptrons composed of millions of neurons. Thus, CNN approaches usually require hundreds of training images. Otherwise, overfitting may occur with CNN models. Collection of the training images is usually time-consuming. For studies with limited images, conventional image-based approaches can still be an option.

In this research, ten native Fagaceae species in Taiwan were identified using leaf images, image processing, and machine learning. The specific objectives of this study were to (1) acquire images of the ten Fagaceae species, (2) define and quantify the morphological, color, Fourier descriptor, and fractal dimension traits of the leaves, (3) apply a genetic algorithm (GA) to select traits essential for species identification, and (4) develop SVM classifiers to distinguish the sampled Fagaceae species.

Materials and Methods

Leaf Sample Collection and Image Acquisition

Leaves of ten Fagaceae species were collected from the main campus of National Taiwan University (NTU) and from the Xiaping nature education area of the NTU experimental forest (table 1). Five to seven individual trees were sampled for each species. For each tree, a branch at the outer canopy was harvested and preserved in a zipper bag to prevent water loss. Only mature and sound leaves without feeding scars were selected and separated from the branch for scanning. Images of the fresh leaf samples were then acquired using flatbed color scanners (V37, Epson, Nagano, Japan; 5100 × 7019 pixel image size) at a resolution of 600 dpi. Several leaves were scanned simultaneously to reduce the scanning time (approx. 60 s per scan). Images were acquired of the lower surfaces of the leaves so that the textural patterns (e.g., veins) of the leaves could be determined. Black cloths were used to shelter the leaves from exposure to stray light. The scanned images were stored in TIF format. The whole process of leaf image acquisition was accomplished within 24 h after the branch was harvested. The collection and scanning were performed between September 2015 and April 2017.

Table 1. Locations of leaf sample collection.

Species	Location
Castanopsis fargesii	NTU forest, Xiaping nature education area
Castanopsis indica	NTU forest, Xiaping nature education area
Cyclobalanopsis glauca	NTU campus
Cyc. hypophaea	NTU forest, Xiaping nature education area
Cyc. pachyloma	NTU forest, Xiaping nature education area
Cyc. stenophylloides	NTU campus
Limlia uraiana	NTU forest, Xiaping nature education area
Pasania harlandii	NTU forest, Xiaping nature education area
Pasania konishii	NTU forest, Xiaping nature education area
Quercus tarokoensis	NTU forest, Xiaping nature education area

Leaf Segmentation and Image Quality Improvement

Each scanned image contained several leaves (fig. 2a). Image processing algorithms were applied to partition the scanned images into separate leaf images, segment the leaf areas from the background, and reduce the noise in the images. First, a scanned image was converted from the RGB color space to the YUV color space. This image was then converted into a binary image on the Y channel using Otsu's method (Otsu, 1979). Connected-component labeling (Haralock and Shapiro, 1991) was then employed to detect objects in the binary image. Objects with pixel sizes smaller than a certain threshold were regarded as noise (i.e., sparkles) and were removed. The resulting image served as an image mask. Subsequently, the scanned image was masked to obtain the leaves in the scanned image. The leaf images were then obtained as rectangles, each of which framed a leaf (fig. 2b). Each rectangle was augmented with 30 pixels

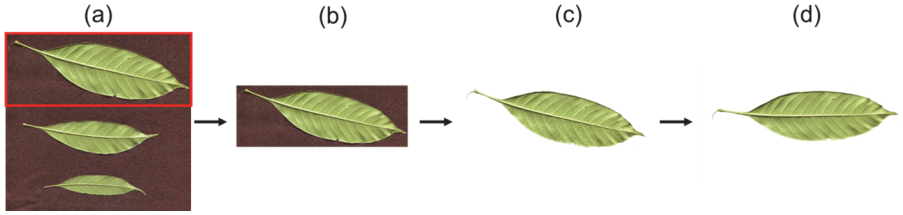

Figure 2. (a) Scanned image with multiple leaves, (b) leaf image with background, (c) image after background removal, and (d) aligned image.

on each side to ensure that the entire leaf was completely retained in the leaf image. Graph cut (Boykov and Kolmogorov, 2004) was applied to the leaf images for automatically segmenting the leaves from the background (fig. 2c). The smooth cost, differential threshold, maximum iteration, and constant weights for the graph cut were set as 50, 0.01, 10, and 0.3, respectively. Connected-component labeling was applied to the foreground leaf image, and then principal component analysis (PCA) was implemented on each labeled leaf image to identify the long axis of the leaf. The leaf image was then rotated so that the long axis aligned with the horizontal axis (fig. 2d). The algorithms were performed using MATLAB (The Mathworks, Natick, Mass.). A total of 100 leaf images were collected for each species.

Leaf Vein Identification

The leaf veins were categorized as primary and higher-order veins (fig. 3). The vein image of a leaf was the combination of the two types of veins. The primary vein of a leaf image was identified using the following procedure. First, the leaf image was converted from the original RGB color space to the YUV color space. This image was then converted to a grayscale image using the Y channel. The contrast of the grayscale image was enhanced using contrast stretching. The primary vein was then identified in the grayscale image using Otsu's thresholding (fig. 3b). The higher-order vein was identified by applying a Hessian-based filter (Frangi et al., 1998; Salem et al., 2007) to the grayscale image (fig. 3c). Finally, the vein image was obtained by applying the binary OR operation to the primary and higher-order vein images (fig. 3d).

Morphological Traits

Six morphological traits were quantified from the leaf images (fig. 2d): blade area (mm^2), blade length (mm), blade width (mm), petiole length (mm), aspect ratio, and blade-petiole length ratio. For morphological trait quantification, a leaf image was first binarized using Otsu's method. Morphological opening (Gonzalez and Woods, 2008) was then applied to the leaf image using a disk-structuring element with a size of 40 × 40 pixels. After morphological opening, connected-component labeling was applied to identify the largest component as the blade region and the second-largest component as the petiole region. PCA was then applied to the blade and petiole regions for identifying the principal and secondary axes of the two regions. The principal axis was defined as the direction with the largest spatial variance of pixels. The secondary axis was perpendicular to the principal axis. The blade length and width were defined as the length of the principal axis and secondary axis, respectively, across the blade region. The petiole length was defined as the length of the principal axis across the petiole region. The aspect ratio was defined as the ratio of the blade length to the blade width. The blade-petiole length ratio was defined as the ratio of the blade length to the petiole length. The blade area was defined as the summation of the pixels in the blade region.

Color Traits

Two color traits were quantified, namely the hue values for the blade and petiole regions. The blade and petiole regions were converted from the original RGB color space to the HSV color space. The mean hue values (in degrees) of the blade and petiole regions were calculated as color traits.

Fourier Descriptors

The Fourier descriptors (Rohlf and Archie, 1984) were quantified from the contours of the blade regions. Fourier descriptors are a set of sine and cosine harmonics at various frequencies. The descriptors were used to depict the contour of a blade by aligning the sine and cosine harmonics to the

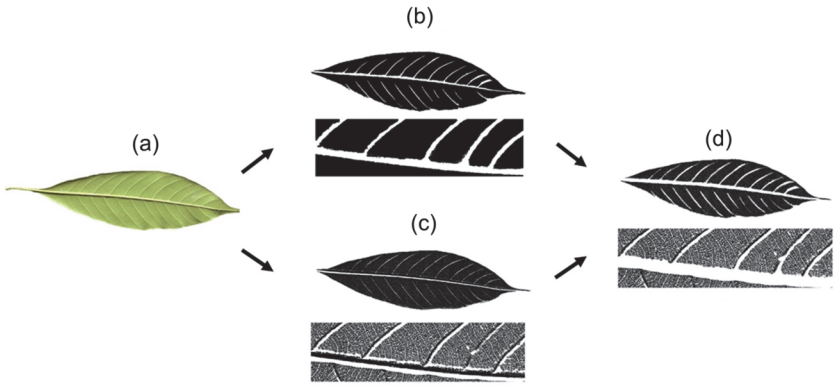

Figure 3. Leaf vein identification: (a) original leaf image, (b) primary vein image, (c) higher-order vein image, and (d) unified image of the primary and higher-order vein images.

blade. The descriptors at different frequencies represent the levels of serration in the blade margin. For calculating the descriptors, the blade contour was first represented as a series of sequentially connected points in a Cartesian coordinate system. The coordinates of the connected points in each dimension were then converted into Fourier descriptors using discrete Fourier transform (Harris, 1978). The first 100 descriptors were quantified for characterizing the margin. The leaf images were two-dimensional, and a sine and cosine harmonic existed in each dimension. Four hundred coefficients of the descriptors were collected as the traits.

Fractal Dimension Features

The fractal dimension features were quantified from the vein images (fig. 3d). The fractal dimension features described the vein patterns of the leaves. These features were calculated using a box-counting algorithm (Du et al., 2013). First, a vein image was padded to an image of 6000 × 6000 pixels with zero padding, and images comprising checkerboard-allocated boxes (fig. 4) were then generated. The box lengths were set as 1, 2, 4, and 8 pixels. The generated images served as masks on the vein image. For each box length, the number of boxes covered by non-zero pixels (N_r) in the vein image was counted. The fractal dimension (D) was then determined as follows:

$$D = \frac{\log(N_r)}{\log(1/r)} \quad (1)$$

where r is the ratio of the box length to the image length. The fractal dimension was calculated for all four box lengths.

Species Identification

Soft-margin SVM classifiers with radial basis function (RBF) kernels were developed to identify the Fagaceae species from the aforementioned traits. The classifiers were developed using the open-source library LIBSVM (Chang and Lin, 2011). The margin and kernel parameters of the classifiers were determined using grid search and ten-fold cross-validation (CV; Arlot and Celisse, 2010). The performance of the classifiers was also evaluated using ten-fold CV. In the CV, the leaf samples were randomly partitioned into ten groups. Nine groups were used for developing the classifiers, and the remaining group was used for evaluating the performance of the classifiers. The process was repeated ten times, with each of the groups used once for evaluation.

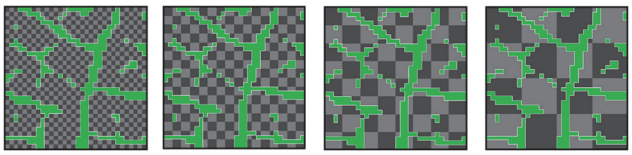

Figure 4. Illustration of box counting. The vein area is represented in green, and the checkerboards are represented in gray and black. The box lengths of the checkerboards are (a) 1, (b) 2, (c) 4, and (d) 8 pixels.

Trait Selection

Attribute selection was applied to determine the traits essential for species identification. Excluding insignificant traits improves the efficiency of model development and reduces the risk of overfitting. A genetic algorithm (Huang and Wang, 2006), which implemented an iterative process involving three steps, was used for performing attribute selection. First, the initial attribute candidates were randomly generated. SVM classifiers were then developed using the candidates as inputs. The SVM classifiers were evaluated using a fitness function. The fitness function was set as the weighted summation of the accuracy, sensitivity, and specificity of the developed classifiers. The weights for the accuracy, sensitivity, and specificity were set as 0.4, 0.3, and 0.3, respectively. Attributes with large fitness values were retained using roulette wheel selection (Holland, 1992). Crossover and mutation were then applied to modify the retained candidates. The crossover and mutation rates were set as 0.8 and 0.03, respectively. The aforementioned procedure continued until the number of iterations reached 300 or the fitness value did not change.

Results and Discussion

Variation in Morphological Traits

Morphological discrepancies were observed among the leaves of the ten species. Table 2 lists the means and standard deviations (SD) of the morphological traits for the ten species. The interspecific differences were significantly larger than the intraspecific differences for all the morphological traits (ANOVA, $p < 0.01$). Thus, the morphological traits may help to distinguish the species.

Variation in Color Traits

Color discrepancies were observed in the blades and pet-

Table 2. Statistics of morphological traits. Lowercase letters denote the grouping results of Scheffé's multiple comparison test.

Species	Blade Width (mm)	Blade Length (mm)	Blade Area (mm^2)	Petiole Length (mm)	Blade-Petiole Length Ratio	Aspect Ratio
Cas. fargesii	34.29 ±6.02 a	118.70 ±18.81 a	2751.96 ±885.97 a	9.62 ±3.40 ab	13.49 ±4.22 a	3.49 ±0.27 a
Cas. indica	53.58 ±8.61 b	161.26 ±22.89 b	5641.22 ±1622.93 b	13.71 ±5.23 cd	13.10 ±4.23 a	3.03 ±0.29 b
Cyc. glauca	32.99 ±5.31 a	101.62 ±14.61 c	2187.90 ±613.32 c	12.11 ±7.90 bc	11.73 ±6.44 a	3.10 ±0.28 bc
Cyc. hypophaea	15.39 ±2.06 c	68.78 ±12.07 d	713.39 ±205.81 de	9.89 ±3.00 ab	7.54 ±2.46 bc	4.46 ±0.51 d
Cyc. pachyloma	21.41 ±4.67 de	113.35 ±22.44 a	1594.78 ±635.79 f	15.05 ±3.75 d	7.84 ±2.06 bc	5.34 ±0.61 e
Cyc. stenophylloides	38.91 ±8.78 f	113.52 ±21.18 a	2759.84 ±1063.42 a	19.75 ±3.97 e	5.85 ±1.11 b	2.95 ±0.32 b
Lim. uraiana	24.87 ±3.24 dg	74.39 ±8.51 d	1203.43 ±261.84 df	8.42 ±1.54 a	9.17 ±2.23 c	3.01 ±0.32 b
Pas. harlandii	42.96 ±9.81 h	140.29 ±30.94 e	4055.57 ±1800.59 g	26.87 ±7.64 g	5.53 ±1.66 b	3.30 ±0.44 ac
Pas. konishii	27.22 ±5.84 g	65.15 ±11.96 d	1061.84 ±421.61 df	12.99 ±6.62 cd	6.23 ±2.96 b	2.41 ±0.26 f
Que. tarokoensis	18.39 ±2.88 ce	32.08 ±5.26 f	435.56 ±134.27 e	3.14 ±1.55 h	12.04 ±4.72 a	1.75 ±0.18 g
p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
F-value	362.45	443.55	298.65	171.71	74.92	719.21

ioles (fig. 5). The hue of the blades ranged between 40° and 100°, and the hue of the petioles ranged between 30° and 90°. The blades of *Cyc. stenophylloides* and the petioles of *Cyc. pachyloma* were associated with a high degree of greenness because their hue values were relatively large. By contrast, the blades and petioles of *Cas. fargesii* were associated with a high degree of redness because their hue values were relatively small. Figure 1 shows sample images of *Cas. fargesii* and *Cyc. pachyloma* leaves. The blade and petiole colors varied considerably from one species to another. The blades and petioles of *Cas. fargesii* were associated with a strong degree of brownness (fig. 5). By contrast, the blades and petioles of *Cyc. pachyloma* were associated with a strong degree of greenness. The color differences were precisely quantified using image processing.

Variation in Leaf Shape

PCA was applied to the Fourier descriptors for visualizing the shape discrepancies of the leaves among the species. The first four principal components (PCs) accounted for 97.79% of the total variance. Hence, the blade shape variations corresponding to only the first four PCs were analyzed. Figure 6 illustrates the leaf shapes reconstructed using inverse PCA and the altered PC values (mean and mean ±SD). For the leaves shown in figure 6, the left-side tips are the base and the right-side tips are the apex. PC1 was primarily associated with the size. The blade regions with small PC1 values had a relatively large size. By contrast, the blade regions with large PC1 values had a relatively small size. PC2 was primarily associated with the ovate-oblong shape. The leaves with large PC2 values were acute at the base, whereas the leaves with small PC2 values were obtuse at the base. PC3 was related to the obovate-oblong shape. The leaves with small PC3 values had a relatively large blade width. Conversely, the leaves with large PC3 values had slender blades. PC4 was primarily associated with left-right asymmetry. The leaves with small PC4 values curved outward to the right, whereas the leaves with large PC4 values curved outward to the left.

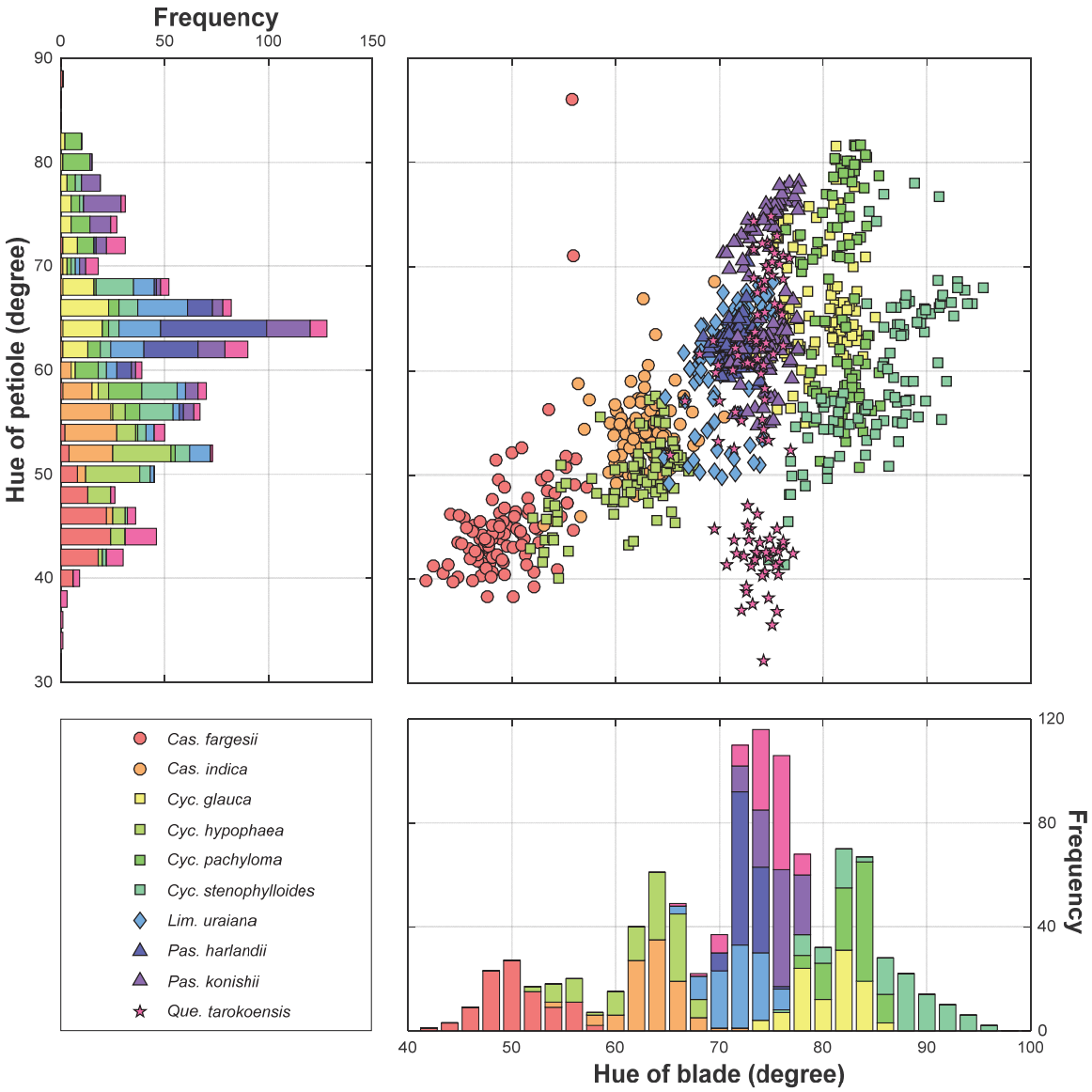

Figure 5. Color variations in the blades and petioles of the ten Fagaceae species.

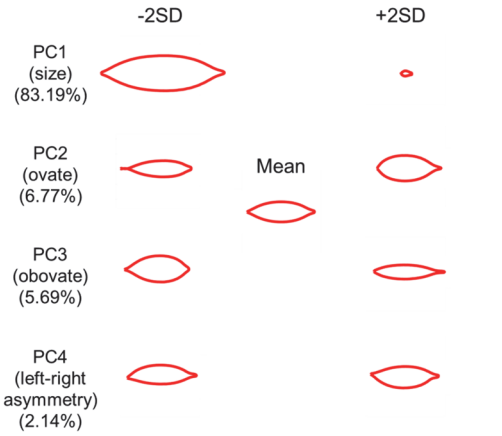

Figure 6. Major variations in blade shape.

Figure 7 shows the distribution, correlation, and histogram of the Fourier descriptor PCs. Discrepancies were observed in the Fourier descriptor PCs of the species. *Que. tarokoensis* and *Pas. konishii* had positive PC1 values. By contrast, *Pas. harlandii* and *Cas. indica* had mostly negative PC1 values. This observation indicated that the blade sizes of *Pas. harlandii* and *Cas. indica* were usually larger than those of *Que. tarokoensis* and *Pas. konishii* (fig. 1). *Cyc. stenophylloides* had positive PC2 values, which indicated that its blades were ovate (fig. 1). *Cas. indica* mostly had negative PC3 values, which indicated that its blades were obovate (fig. 1). *Pas. konishii*, *Que. tarokoensis*, and *P. harlandii* had neutral PC4 values, which indicated that the blades of these species were left-right symmetrical (fig. 1).

Variation in Fractal Dimension Features

Discrepancies were observed in the fractal dimension features of the species (fig. 8). A small fractal dimension at a certain box length indicated that the vein at the frequency corresponding to the box length was sparse. *Cyc. stenophylloides* was associated with the smallest fractal dimension at box lengths of 4 and 8 pixels and with a medium fractal dimension at box lengths of 1 and 2 pixels. This observation

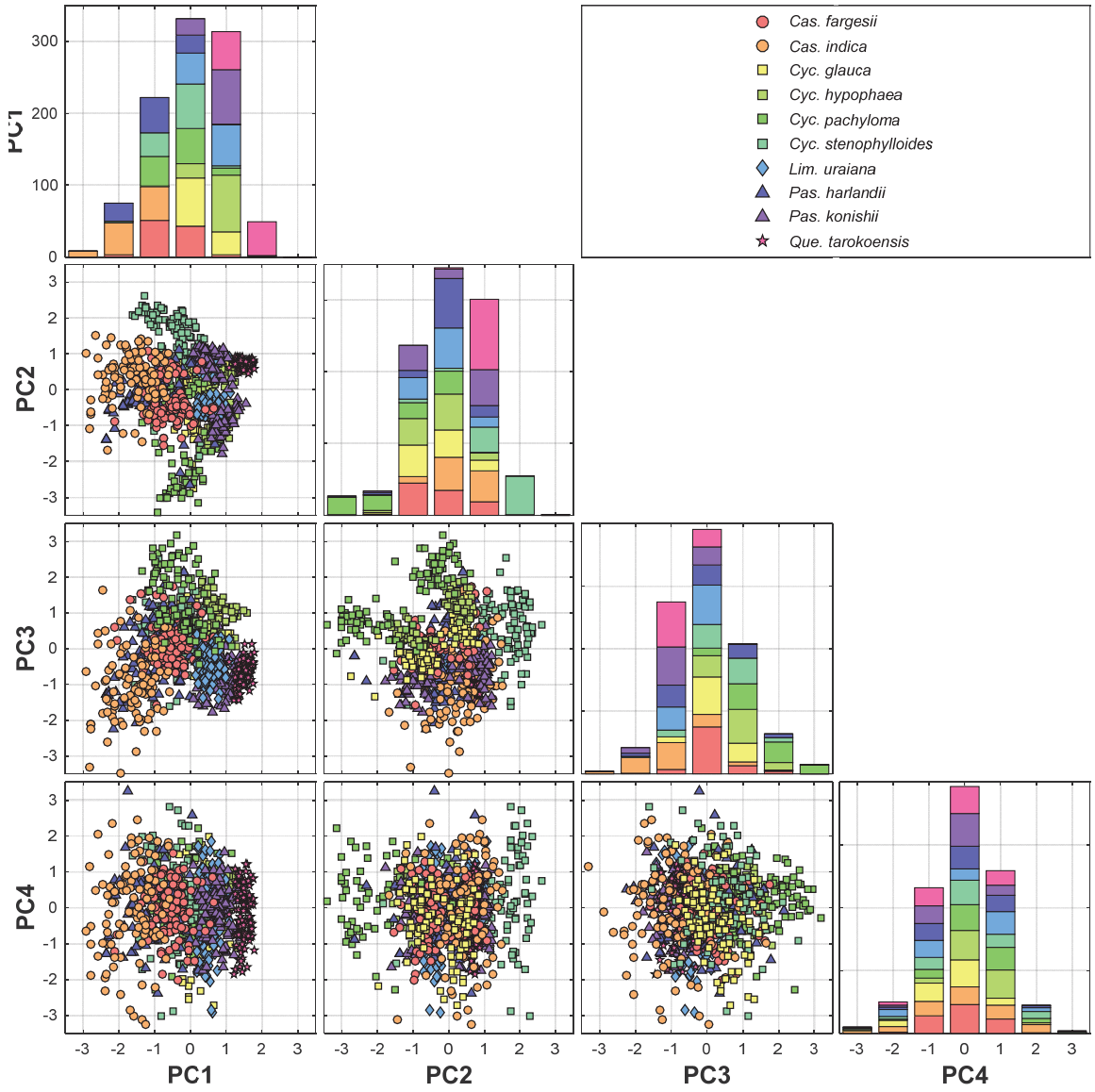

Figure 7. Distributions of the PCs of the Fourier descriptors for the ten Fagaceae species.

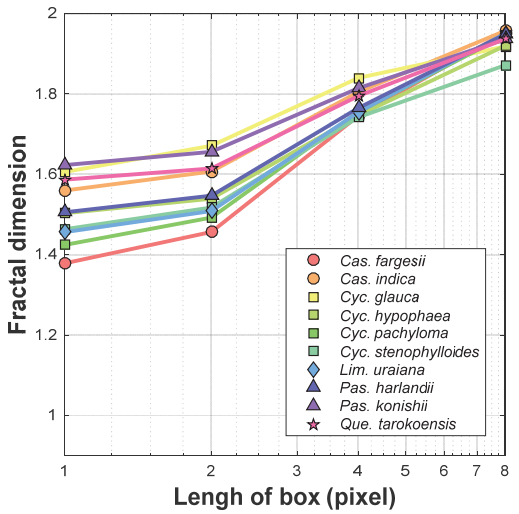

Figure 8. Fractal dimension features of the species.

indicated that *Cyc. stenophylloides* had a larger number of higher-order veins than lower-order veins. By contrast, *Cas. indica* was associated with the largest fractal dimension at a box length of 8 pixels and a medium fractal dimension at box lengths of 1 and 2 pixels. This observation indicated that *Cas. indica* had a higher number of lower-order veins than higher-order veins.

Figure 9 shows vein patches with the maximum and minimum fractal dimension values at box lengths of 1, 2, 4, and 8 pixels. Each vein patch contains a second-order vein (white line) that crosses the patch. The density of the higher-order veins varied between the species. *Pas. konishii* was associated with a high density of higher-order veins at box lengths of 1 and 2 pixels (fig. 9). The third-order and fourth-order veins of *Pas. konishii* could be clearly observed. By contrast, *Cas. fargesii* was associated with a low density of higher-order veins.

Species Identification Using Various Traits

The SVM classifiers that used the morphological traits, color traits, Fourier descriptor traits, fractal dimension features, and all the traits as inputs had CV accuracies of 65.2%, 65.0%, 70.6%, 89.5%, and 73.7%, respectively (fig. 10). The classifier using fractal dimension features as inputs outperformed all the other models. Thus, the venation-related traits of the leaves comprised the most essential information for identifying the Fagaceae species. Moreover, the classifier using all the traits as inputs had a mediocre accuracy, which indicated that a certain level of overfitting existed in this classifier. This observation also provided evidence for the necessity of using trait selection as a remedy against overfitting.

Species Identification Using Selected Traits

Traits essential for species identification were selected using the GA. The selected trait set contained four of the six morphological features, one of the two color features, 196 of the 400 Fourier descriptors, and all four fractal dimension features. The developed SVM classifier achieved an overall accuracy of 92.8% (fig. 11). The classifier using the selected features as inputs outperformed the classifiers using the fractal dimension and all the features as inputs.

Conclusion

In this study, ten Fagaceae species were identified using image processing and machine learning. Leaf images of the ten Fagaceae species were collected using flatbed scanners. The morphological, color, Fourier descriptor, and fractal dimension traits of the leaves were quantified using image processing algorithms. GA was then applied to select essential traits for species identification. An optimum set comprising 205 traits was determined. SVM classifiers were developed to identify the species using the selected traits as inputs. The proposed approach had a CV accuracy of 92.8%. In contrast to genetic marker-based methods, the proposed approach does not require sophisticated equipment. Thus, the proposed method can be used for identifying Fagaceae species in the field.

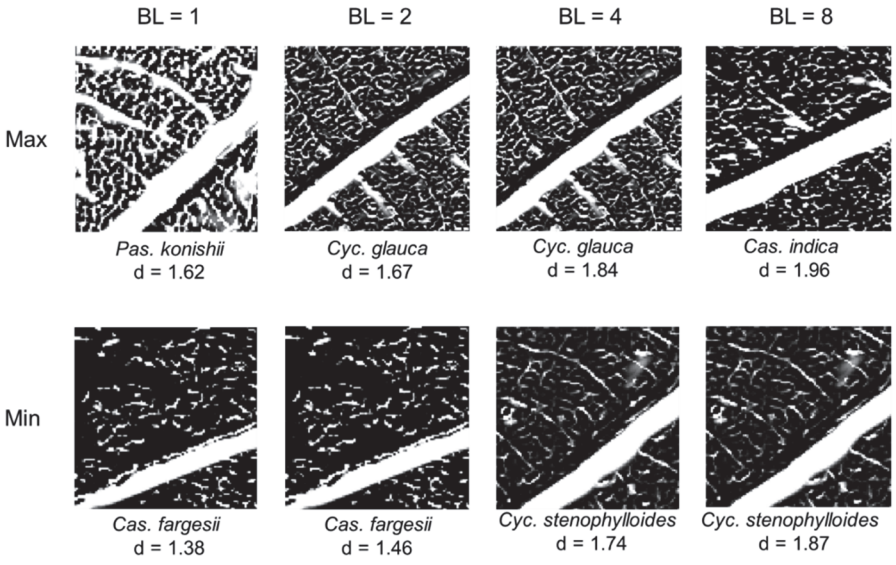

Figure 9. Vein images of patches with the maximum and minimum fractal dimension values (d) at box lengths (BL) of 1, 2, 4, and 8 pixels.

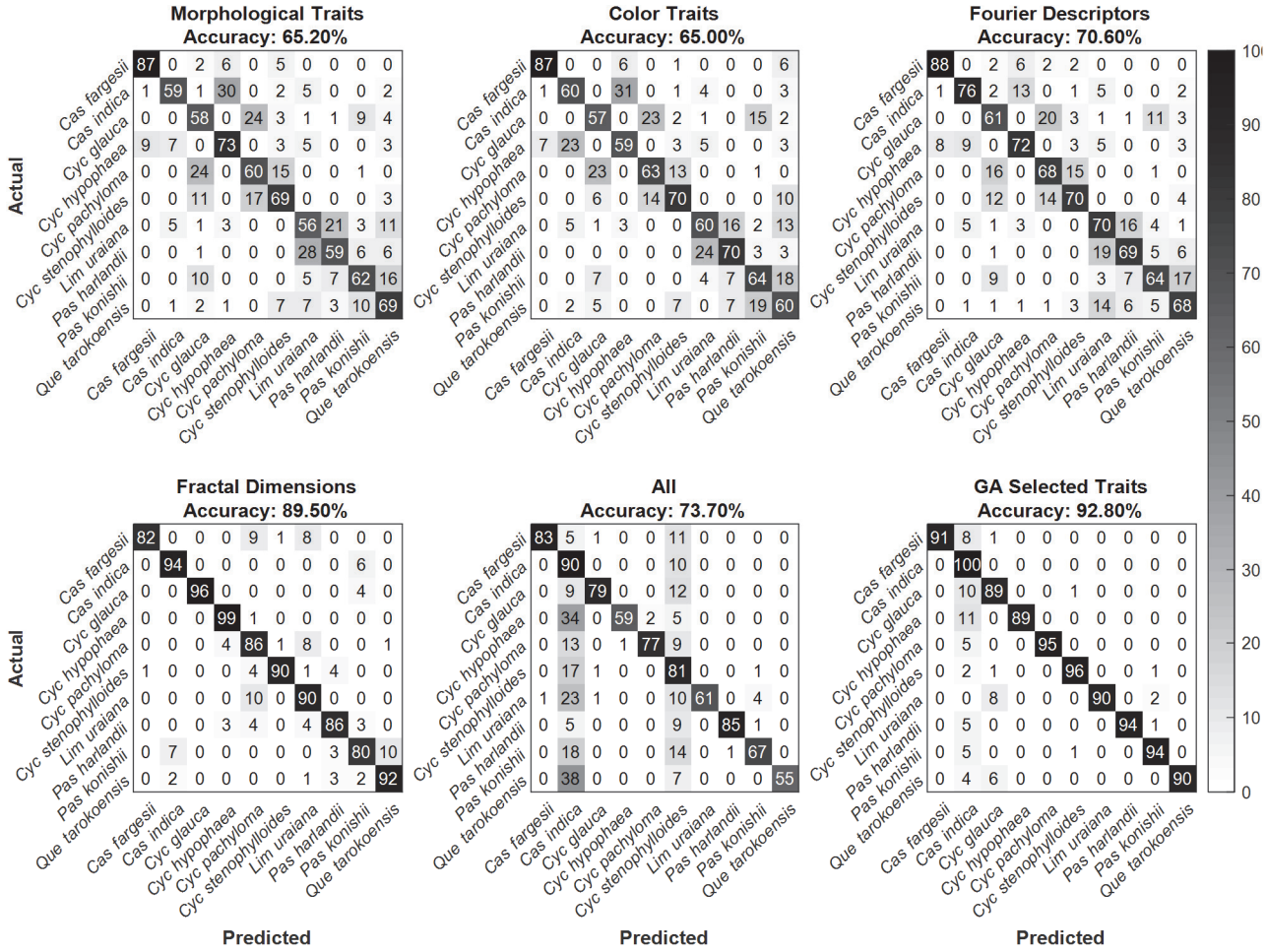

Figure 10. Confusion matrices of the SVM classifiers using the morphological traits, color traits, Fourier descriptor traits, fractal dimension features, all the traits, and the GA-selected traits as inputs.

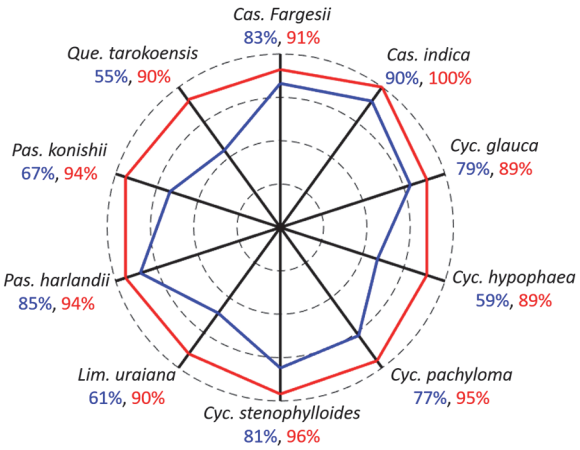

Figure 11. Identification accuracies of the SVM classifiers using the GA-selected traits (red) and all the traits (blue).

ACKNOWLEDGEMENTS

This research was supported by the Taiwan Ministry of Science and Technology under Grant No. 107-2313-B-002-013-MY3. We thank Fang-Hua Chu and Wei-Ting Liou for their invaluable suggestions for this research.

REFERENCES

Aakif, A., & Khan, M. F. (2015). Automatic classification of plants based on their leaves. *Biosyst. Eng., 139*, 66-75. https://doi.org/10.1016/j.biosystemseng.2015.08.003

Aoki, K., Ueno, S., Kamijo, T., Setoguchi, H., Murakami, N., Kato, M., & Tsumura, Y. (2014). Genetic differentiation and genetic diversity of *Castanopsis* (Fagaceae), the dominant tree species in Japanese broadleaved evergreen forests, revealed by analysis of EST-associated microsatellites. *PLoS One, 9*(1), e87429. https://doi.org/10.1371/journal.pone.0087429

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Stat. Surv., 4*, 40-79. https://doi.org/10.1214/09-SS054

Boykov, Y., & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Machine Intel., 26*(9), 1124-1137. https://doi.org/10.1109/TPAMI.2004.60

Bruschi, P., Vendramin, G. G., Bussotti, F., & Grossoni, P. (2003). Morphological and molecular diversity among Italian populations of *Quercus petraea* (Fagaceae). *Ann. Botany, 91*(6), 707-716. https://doi.org/10.1093/aob/mcg075

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intel. Syst. Tech., 2*(3), 1-27. https://doi.org/10.1145/1961189.1961199

Chang, F.-R., Wang, C.-H., Cheng, Y.-B., & Wu, Y.-C. (2016). Chemical constituents and bioactivities from the leaves and

twigs of *Lithocarpus synbalanos*. *Planta Med., 82*(S 01), P242. https://doi.org/10.1055/s-0036-1596388
Cheng, Y.-P., Hwang, S.-Y., & Lin, T.-P. (2005). Potential refugia in Taiwan revealed by the phylogeographical study of *Castanopsis carlesii* Hayata (Fagaceae). *Molecular Ecol., 14*(7), 2075-2085. https://doi.org/10.1111/j.1365-294X.2005.02567.x
Coutinho, J. P., Carvalho, A., & Lima-Brito, J. (2014). Genetic diversity assessment and estimation of phylogenetic relationships among 26 Fagaceae species using ISSRs. *Biochem. System. Ecol., 54*, 247-256. https://doi.org/10.1016/j.bse.2014.02.012
Du, J.-X., Zhai, C.-M., & Wang, Q.-P. (2013). Recognition of plant leaf image based on fractal dimension features. *Neurocomputing, 116*, 150-156. https://doi.org/10.1016/j.neucom.2012.03.028
Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. (1998). Multiscale vessel enhancement filtering. *Proc. Intl. Conf. Medical Image Computing and Computer-Assisted Intervention* (MICCAI '98) (pp. 130-137). Berlin, Germany: Springer. https://doi.org/10.1007/BFb0056195
Gonzalez, R. C., and Woods, R. E. (2008). *Digital image processing* (3rd Ed.). Upper Saddle River, NJ: Pearson Education.
Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric., 127*, 418-424. https://doi.org/10.1016/j.compag.2016.07.003
Haralock, R. M., & Shapiro, L. G. (1991). *Computer and robot vision.* Boston, MA: Addison-Wesley Longman.
Harish, B. S., Hedge, A., Venkatesh, O., Spoorthy, D. G., & Sushma, D. (2013). Classification of plant leaves using morphological features and Zernike moments. *Proc. Intl. Conf. Advances in Computing, Communications, and Informatics* (pp. 1827-1831). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICACCI.2013.6637459
Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE, 66*(1), 51-83. https://doi.org/10.1109/PROC.1978.10837
Holland, J. H. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/1090.001.0001
Huang, C.-L., & Wang, C.-J. (2006). A GA-based feature selection and parameters optimizationfor support vector machines. *Expert Syst. Appl., 31*(2), 231-240. https://doi.org/10.1016/j.eswa.2005.09.024
Janani, R., & Gopal, A. (2013). Identification of selected medicinal plant leaves using image features and ANN. *Proc. Intl. Conf. Advanced Electronic Systems* (pp. 238-242). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICAES.2013.6659400
Kremer, A., Abbott, A. G., Carlson, J. E., Manos, P. S., Plomion, C., Sisco, P., ... Vendramin, G. G. (2012). Genomics of Fagaceae. *Tree Genetics Genomes, 8*(3), 583-610. https://doi.org/10.1007/s11295-012-0498-3
Kuo, C.-C., & Lee, L.-L. (2003). Food availability and food habits of Indian giant flying squirrels (*Petaurista philippensis*) in Taiwan. *J. Mammal., 84*(4), 1330-1340. https://doi.org/10.1644/BOS-039
Lee, P. F., Ding, T. S., Hsu, F. H., & Geng, S. (2004). Breeding bird species richness in Taiwan: Distribution on gradients of elevation, primary productivity, and urbanization. *J. Biogeography, 31*(2), 307-314. https://doi.org/10.1046/j.0305-0270.2003.00988.x
Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recog., 71*, 1-13. https://doi.org/10.1016/j.patcog.2017.05.015
Lee, S. H., Chan, C. S., Wilkin, P., & Remagnino, P. (2015). Deep-plant: Plant identification with convolutional neural networks. *Proc. Intl. Conf. Image Processing (ICIP)* (pp. 452-456). Piscataway, NJ: IEEE. https://doi.org/10.1109/ICIP.2015.7350839
Liao, J.-C. (1996). Fagaceae. In *Flora of Taiwan* (2nd Ed., Vol. 2, pp. 51-123). Taipei, Taiwan: National Taiwan University.
Ogata, K., Fujii, T., Abe, H., & Baas, P. (2008). Identification of the timbers of southeast Asia and the western Pacific. *Holzforschung, 62*(6), 765-765. https://doi.org/10.1515/HF.2008.132
Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cyber., 9*(1), 62-66. https://doi.org/10.1109/TSMC.1979.4310076
Pencakowski, B. M., Tokarski, M., Jonkisz, A., Czosnykowska-Lukacka, M., Lenard, E., & Małodobra-Mazur, M. (2018). DNA profiling of oaks (*Quercus* spp.). *Archiwum Medycyny Sądowej i Kryminologii/Arch [Archives of Forensic Medicine and Criminology], 68*(1), 1-9. https://doi.org/10.5114/amsik.2018.75942
Rohlf, F. J., & Archie, J. W. (1984). A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *System. Zool., 33*(3), 302-317. https://doi.org/10.2307/2413076
Salem, N. M., Salem, S. A., & Nandi, A. K. (2007). Segmentation of retinal blood vessels based on analysis of the Hessian matrix and clustering algorithm. *Proc. 15th European Signal Processing Conf.* (pp. 428-432). Piscataway, NJ: IEEE.
Singh, A. K., Singh, S. K., Singh, P. P., Srivastava, A. K., Pandey, K. D., Kumar, A., & Yadav, H. (2018). Biotechnological aspects of plants metabolites in the treatment of ulcer: A new prospective. *Biotech. Reports, 18*, e00256. https://doi.org/10.1016/j.btre.2018.e00256
Tang, Z., Su, Y., Er, M. J., Qi, F., Zhang, L., & Zhou, J. (2015). A local binary pattern based texture descriptors for classification of tea leaves. *Neurocomputing, 168*, 1011-1023. https://doi.org/10.1016/j.neucom.2015.05.024
Tomlinson, P. T., Dickson, R. E., & Isebrands, J. G. (1991). Acropetal leaf differentiation in *Quercus rubra* (Fagaceae). *American J. Bot., 78*(11), 1570-1575. https://doi.org/10.1002/j.1537-2197.1991.tb11436.x
Ueno, S., & Tsumura, Y. (2008). Development of ten microsatellite markers for *Quercus mongolica* var. *crispula* by database mining. *Conserv. Genetics, 9*(4), 1083-1085. https://doi.org/10.1007/s10592-007-9462-4
Ueno, S., Aoki, K., & Tsumura, Y. (2009). Generation of expressed sequence tags and development of microsatellite markers for *Castanopsis sieboldii* var. *sieboldii* (Fagaceae). *Ann. Forest Sci., 66*(5), 1-21. https://doi.org/10.1051/forest/2009037
Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Arch. Comput. Methods Eng., 25*(2), 507-543. https://doi.org/10.1007/s11831-016-9206-z
Yang, H.-W., Hsu, H.-C., Yang, C.-K., Tsai, M.-J., & Kuo, Y.-F. (2019). Differentiating between morphologically similar species in genus *Cinnamomum* (Lauraceae) using deep convolutional neural networks. *Comput. Electron. Agric., 162*, 739-748. https://doi.org/10.1016/j.compag.2019.05.003