# Plant species identification based on leaf venation features using SVM

**Article** *in* TELKOMNIKA Indonesian Journal of Electrical Engineering · March 2020

DOI: 10.12928/TELKOMNIKA.v18i2.14062

**TELKOMNIKA Telecommunication, Computing, Electronics and Control**
Vol. 18, No. 2, April 2020, pp. 726~732
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i2.14062


# Plant species identification based on leaf venation features using SVM

**Agus Ambarwari[1], Qadhli Jafar Adrian[2], Yeni Herdiyeni[3], Irman Hermadi[4]**
[1]Department of Informatics Engineering, Universitas Teknokrat Indonesia, Indonesia
[1]Department of Informatics Management, Politeknik Negeri Lampung, Indonesia
[2]Department of Information System, Universitas Teknokrat Indonesia, Indonesia
[3,4]Department of Computer Science, IPB University, Indonesia

**Article Info**

***Article history:***

Received Sep 7, 2019
Revised Dec 26, 2019
Accepted Feb 6, 2020

***Keywords:***

Feature extraction
Leaf venation
SVM

**ABSTRACT**

The purpose of this study is to identify plant species using leaf venation features. Leaf venation features were obtained through the extraction of leaf venation features. The leaf image segmentation was performed to obtain the binary image of the leaf venation which is then determined the branching point and ending point. From these points, the extraction of leaf venation feature was performed by calculating the value of straightness, a different angle, length ratio, scale projection, skeleton length, number of segments, total skeleton length, number of branching points and number of ending points. So that from the extraction of leaf venation features 19 features were obtained. Identification of plant species was carried out using Support Vector Machine (SVM) with RBF kernel. The learning model was built using 75% of the training data. The testing results using 25% of the data on the training model, obtained an accuracy of 82.67%, with an average of precision of 84% and recall of 83%.



***Corresponding Author:***

Agus Ambarwari,
Department of Informatics Engineering,
Universitas Teknokrat Indonesia,
Jl. Zainal Abidin Pagar Alam No.9-11, Kedaton, Bandar Lampung 35132, Indonesia.
Email: ambarwariagus@teknokrat.ac.id

## 1. INTRODUCTION

Many studies that discuss techniques or methods for identifying plants, show that plant identification is important. The study of the physical form and structure of plant bodies, known as plant morphology [1]. Plant morphology is useful for visually identifying plants, so that plant diversity can be identified, classified and given the right name for each group formed. In addition to describing the shape and composition of the plant body, plant morphology is useful for determining the function of each part in plant life. Furthermore, the origin and composition of the plant body can be known [2]. Information about plant morphology is needed to understand evolution, ecology, geographical distribution, conservation, life cycle, and species definition [3].

Parts of plants that have different characteristics from each plant and commonly used for plant identification are leaves [4-5]. Leaves have an important role for plants to adapt to their environment [6]. The main features on leaves that can distinguish each plant species include shape [7-9], texture [10-13], color [14], and leaf venation [15-18] are called leaf biometrics [19]. Among those features, that have unique diversity and can describe plant characteristics in more detail is leaf venation, although there are some species of plants that have vague patterns that are not very clear [20]. To obtain leaf venation patterns, it can be done using feature extraction techniques through digital image processing [21-22], so that leaf venation features are obtained.

Extraction of leaf venation feature can be done by implementing Self-Invariant Feature Transform (SIFT) [23], Fourier and B-Spline modeling [15]. However, the results of segmentation are less than optimal because they only reach secondary venation. In addition, many parts of leaf venation are not segmented. The method is then improved using the Hessian matrix [24], which is by implementing a vessel measure based on the eigenvalues of the Hessian matrix. The result is that the system can segment leaf venation to tertiary venation. In Prastya's study [25], the binary venation image of the extracted leaves was calculated for the value of straightness, a different angle, length ratio, and scale projection. These values are then used as a marker of leaf venation. Other studies conducted by Ambarwari *et al.* [16], who performed an analysis of leaf venation density features to obtain the most important features, which can distinguish types of leaf venation. These features by Ambarwari *et al.* [26] is used to identify plants based on the type of venation. However, from some of these studies, no one has identified the plant species.

In this study, identifying plant species using the leaf venation feature. Leaf venation features were obtained through feature extraction. Leaf venation feature extraction will produce several features, including straightness, a different angle, length ratio, scale projection [25], *total skeleton length, number of branching points and number of ending points* [27]. This leaf venation feature was used to identify plant species by classification techniques. The classification technique used is the Support Vector Machine (SVM). In many cases such as pattern recognition, SVM error rates when testing data are significantly better than other methods [28].

## 2. RESEARCH METHOD

### 2.1. Leaf image acquisition

The stages in this study consisted of six stages as shown in Figure 1. Leaf image data used are Flavia dataset [29] with 6 species were taken, namely *Aesculus chinensis*, *Lagerstroemia indica (L.) Pers.*, *Cinnamomum japonicum Sieb.*, *Chimonanthus praecox L.*, *Ilex macrocarpa Oliv.*, and *Koelreuteria paniculata Laxm.* The number of leaves in each species is 50 leaf images. Leaf image samples are shown in Figure 2.

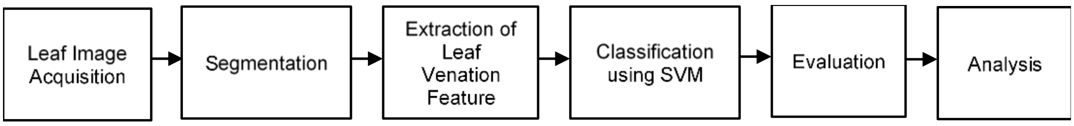


Figure 1. Stages of research

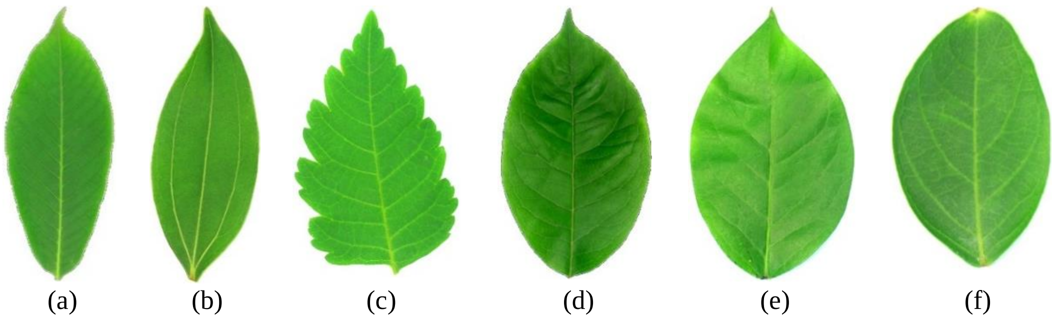


Figure 2. Leaf image data, (a) Aesculus chinensis, (b) Lagerstroemia indica (L.) Pers, (c) Cinnamomum japonicum Sieb, (d) Chimonanthus praecox L, (e) Ilex macrocarpa Oliv, (f) Koelreuteria paniculata laxm

### 2.2. Segmentation

Image segmentation is the process of separating images into homogeneous parts and extracting these parts into objects that will be observed so that the region of interest is obtained [30]. From the acquisition of leaf image data, then image segmentation was performed using the Hessian matrix [24] to obtain the leaf venation shape. Leaf image data from segmentation results in the form of leaf venation binary image data shown in Figure 3.

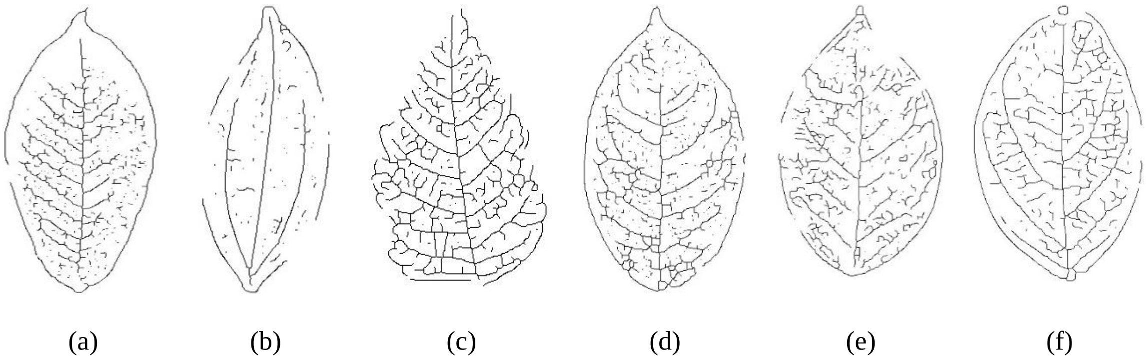


Figure 3. Binary venation leaf image data, (a) Aesculus chinensis, (b) Lagerstroemia indica (L.) Pers., (c) Cinnamomum japonicum Sieb, (d) Chimonanthus praecox L, (e) Ilex macrocarpa Oliv, (f) Koelreuteria paniculata laxm

### 2.3. Extraction of leaf venation feature

The leaf venation image data from the segmentation results, then exraction of the leaf venation feature. Leaf venation feature extraction was obtained from the calculation of the value of straightness, different angle, length ratio and scale projection [25], in order to calculate these values first, the detection of branch points and end points. Illustration of leaf venation feature extraction is shown in Figure 4.

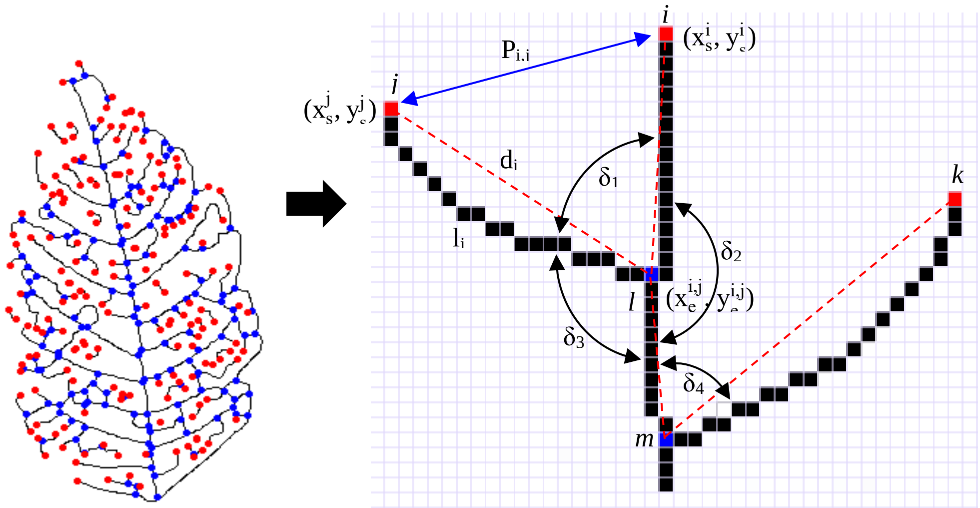


Figure 4. Illustration of leaf venation feature extraction [16]

The notation of Figure 4 is presented below:

x, y = pixel coordinate

$l_j$ = length of the j segment that represented by the number of the pixels

$d_j$ = distance between pixel coordinates $(x_s, y_s)$ and $(x_e, y_e)$

Straightness is a measurement of the alignment value of a segment. From Figure 4, the straightness value is calculated using (1).

$$straightness = \frac{l_j}{d_j} \quad (1)$$

Different angle ($\delta_1$) is a measurement of the angle difference between coincident segments. From Figure 4, different angle values are calculated using the (2).

$$\delta_1 = |\alpha_i - \alpha_j| \quad (2)$$

Length ratio ($R_i$) is measured by comparing the length of each segment with the maximum length of the segment in a leaf venation image. From Figure 4, the length ratio value is calculated using (3)

$$R_i = \frac{l_i}{\max(\vec{l})} \tag{3}$$

Scale Projection ($P_{i,j}$) is a measurement of projection length between coincident segments. In Figure 4, the projected length between segments i and j is calculated using (4) and (5).

$$P_{i,j} = \frac{\vec{i} \cdot \vec{j}}{(\max(|i|,|j|))^2} \tag{4}$$

$$P_{i,j} = \frac{(x_e^i - x_s^i)(x_e^j - x_s^j) + (y_e^i - y_s^i)(y_e^j - y_s^j)}{(\max(d_i,d_j))^2} \tag{5}$$

This leaf venation feature extraction produces several features, including straightness, different angle, length ratio, scale projection, skeleton length, segment, total skeleton, number of branching points and number of ending points.

### 2.4. SVM classification

The model was built using the SVM classifier with two kernel functions, the RBF and Linear kernels. In the RBF kernel function, parameters C and gamma ($\gamma$). C is a parameter to determine the amount of penalty due to an error in the classification of training data, while $\gamma$ is a parameter controlling the width of the Gaussian function variant. The parameter C value that was tested was [1, 10, 100, 1000] and the parameter value $\gamma$ that was tested was [1e-1, 1e-2, 1e-3, 1e-4]. The search for the best parameter values is done using a grid search with 5-fold cross-validation in the training data. The best pair of C and $\gamma$ values are obtained based on the greatest accuracy value at the time of training the data. The stages of the SVM classification process are shown in Figure 5.

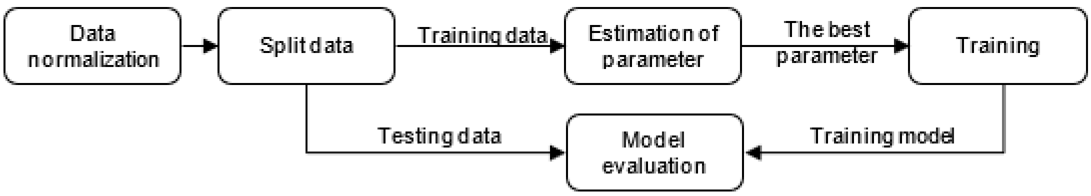


Figure 5. Stages of the SVM classification process

## 3. RESULTS AND ANALYSIS

The Flavia dataset [28] used in this method was chosen based on the extraction of the leaf venation feature. The Flavia dataset consists of 32 species, there are some species which the leaf venation features cannot be extracted. This is because the pattern of leaf venation was not clearly visible, so only leaf venation that has a clear pattern was chosen. Then from this dataset, extraction of the leaf venation feature was performed. Data resulting from the extraction of leaf venation feature consist of 19 features, including mean, variance, standard deviation (of straightness, a different angle, length ratio, scale projection, and length), total skeleton, endpoint, branch point, and segment. The value of these features has a different range of values so that it is normalized using Min-Max normalization. The data is then divided into training data and testing data with 75% portion of training data (225 data) and 25% of testing data (75 data).

In the training data, the best parameter search was performed to build a training model by applying a grid search with 5-fold cross-validation. The best parameter search results are done by pairing the parameters that have been prepared. The best parameters are determined based on the highest accuracy. The parameter search results using the grid search are shown in Table 1.

Based on Table 1 obtained the highest accuracy of 77.8% using the RBF kernel. The best parameters were selected based on the accuracy, that is the parameter values C = 1000 and gamma ($\gamma$) = 0.1. This combination of kernel and parameters was used to build the SVM learning model. Then the results of tests conducted on the model that has been built are displayed in the form of a confusion matrix. The results of the confusion matrix on the testing data against the SVM learning model are shown in Figure 6.

Table 1. Search results of parameters using the grid search

| Kernel | C | Gamma (γ) | Accuracy |
|---|---|---|---|
| rbf | 1 | 0.1 | 0.600 |
| rbf | 1 | 0.01 | 0.413 |
| rbf | 1 | 0.001 | 0.413 |
| rbf | 1 | 0.0001 | 0.413 |
| rbf | 10 | 0.1 | 0.742 |
| rbf | 10 | 0.01 | 0.596 |
| rbf | 10 | 0.001 | 0.413 |
| rbf | 10 | 0.0001 | 0.413 |
| rbf | 100 | 0.1 | 0.751 |
| rbf | 100 | 0.01 | 0.738 |
| rbf | 100 | 0.001 | 0.600 |
| rbf | 100 | 0.0001 | 0.413 |
| **rbf** | **1000** | **0.1** | **0.778** |
| rbf | 1000 | 0.01 | 0.751 |
| rbf | 1000 | 0.001 | 0.738 |
| rbf | 1000 | 0.0001 | 0.600 |
| linear | 1 | - | 0.711 |
| linear | 10 | - | 0.764 |
| linear | 100 | - | 0.760 |
| linear | 1000 | - | 0.764 |

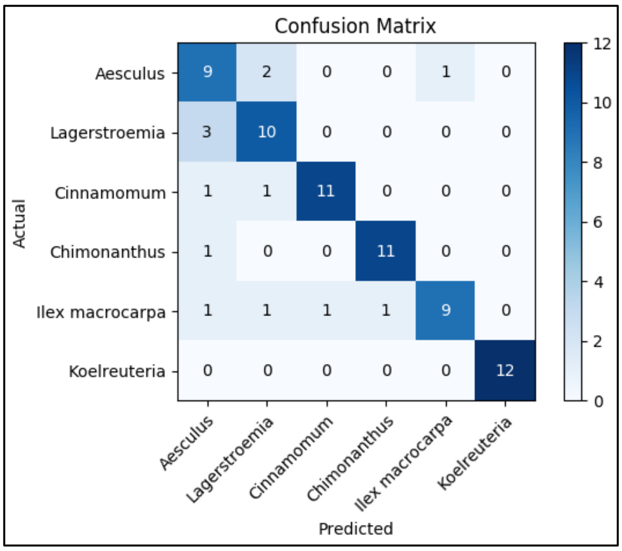


Figure 6. Confusion matrix

Based on the confusion matrix Figure 6 it can be seen that of the six species tested only Koelreuteria paniculata Laxm. species, where all the data is correctly classified. Whereas the other species on average have misclassified between 2 to 3 misclassified data. Overall, the average accuracy obtained in testing data is 82.67%. This accuracy value is far better than identifying leaves based on the type of leaf venation done by Ambarwari *et al.* [16]. In the research of Ambarwari *et al.* [16] the highest accuracy was obtained at 77.57%. One reason for the small accuracy is the amount of data that is not the same in each category. In addition, the type of leaf venation consists of several different species, so it is possible have different features of leaf venation. Detailed results of the SVM classification are shown in Table 2.

Besides accuracy, to see the performance of a classifier we need precision and recall values. Based on Table 2 it can be seen that the species of Aesculus chinensis and Lagerstroemia indica (L.) Pers. which has precision and recall values below 80%. This means that the data of the two species affect the decreases of accuracy. Even so, the average precision and recall obtained were quite high, namely 84% for the value of precision and 83% for the value of recall. This result is very small compared to the study of Wang *et al.* [29], who used the leaf shape feature for plant identification. However, the leaf venation feature can be used to obtain information related to plant anatomy such as the location of plants growt and environmental conditions [16].

Table 2. Detailed results of the SVM classification

| Species | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| *Aesculus chinensis* | 0.60 | 0.75 | 0.67 | 12 |
| *Lagerstroemia indica (L.) Pers.* | 0.71 | 0.77 | 0.74 | 13 |
| *Cinnamomum japonicum Sieb.* | 0.92 | 0.85 | 0.88 | 13 |
| *Chimonanthus praecox L.* | 0.92 | 0.92 | 0.92 | 12 |
| *Ilex macrocarpa Oliv.* | 0.90 | 0.69 | 0.78 | 13 |
| *Koelreuteria paniculata Laxm.* | 1.00 | 1.00 | 1.00 | 12 |
| Average / total | 0.84 | 0.83 | 0.83 | 75 |

## 4. CONCLUSION

Identification of plant species based on leaf venation features was carried out using the SVM classifier. The application of RBF kernel with parameters C and gamma ($\gamma$) in SVM, is able to classify plants with an accuracy of 82.67%. In addition to accuracy, performance on SVM is also measured by an average of precision and recall which is 84% for precision and 83% for recall. The use of leaf venation as a feature in the identification of plant species can be used as an alternative when the leaves of plants have similar shapes and textures. As for suggestions on studies related to plant identification, the process of identification using the feature of leaf venation requires quite a long time, especially at the segmentation stage. The duration of leaf segmentation is due to the manually determined threshold value, meaning that each leaf threshold value is different. On the other hand, information regarding leaf venation features can be obtained. In addition, if the focus of research is on identifying and not caring about plant conditions, another method that can be applied for plant identification is the convolutional neural network (CNN). In terms of speed is superior, but the specifications of the devices used are also higher.

## ACKNOWLEDGEMENTS

The authors thank the Directorate of Research and Community Service, the Directorate General of Research and Technology Strengthening and Development of the Ministry of Research, Technology and Higher Education of the Republic of Indonesia that providing research grants through the 2018 Research and Community Service Program.

## REFERENCES

[1] R. F. Evert and S. E. Eichhorn, “Raven Biology of Plants,” 8th ed. New York (US): Peter Marshall, 2013.
[2] G. Tjitrosoepomo, “Morfologi Tumbuhan,” Yogyakarta (ID): Gadjah Mada University Press, 2016.
[3] B. Douaihy *et al.*, “Morphological versus molecular markers to describe variability in Juniperus excelsa subsp. excelsa (Cupressaceae),” *AoB PLANTS*, 2012.
[4] T. L. Le, D. T. Tran, and V. N. Hoang, “Fully automatic leaf-based plant identification, application for Vietnamese medicinal plant search,” *in The Fifth Symposium on Information and Communication Technology - SoICT ’14*, Hanoi, Viet Nam, pp. 146–154, 2014.
[5] S.S. Chakkaravarthy, G. Sajeevan, E. Kamalanaban and K.V. Kumar, "Automatic leaf vein feature extraction for first degree veins," *In Advances in Signal Processing and Intelligent Recognition Systems*, pp. 581-592, 2016.
[6] J. Yang *et al.*, “Leaf form-climate relationships on the global stage: an ensemble of characters: Global leaf form and climate relationships,” *Global Ecology and Biogeography*, vol. 24, no. 10, pp. 1113–1125, Oct. 2015.
[7] S. Singh and M. S. Bhamrah, “Leaf Identification Using Feature Extraction and Neural Network,” *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 10, no. 5, pp. 134–140, 2015.
[8] A. Bakhshipour and A. Jafari, “Evaluation of support vector machine and artificial neural networks in weed detection using shape features,” *Computers and Electronics in Agriculture*, vol. 145, pp. 153–160, February 2018.
[9] A.A. Patil and K.S. Bhagat, "Plant identification by leaf shape recognition: a review," *International Journal of Engineering Trends and Technology*, vol. 35, no. 8, 2016.
[10] J. S. Cope, P. Remagnino, S. Barman, and P. Wilkin, “Plant texture classification using gabor co-occurrences,” in *Advances in Visual Computing: 6th International Symposium, ISVC 2010*, Berlin (DE), pp. 669-667, 2010.
[11] D. Tomar and S. Agarwal, “Leaf Recognition for Plant Classification Using Direct Acyclic Graph Based Multi-Class Least Squares Twin Support Vector Machine,” *International Journal of Image and Graphics*, vol. 16, no. 3, 2016.
[12] F. R. F. Padao and E. A. Maravillas, “Using Naïve Bayesian method for plant leaf classification based on shape and texture features,” *2015 International Conference on Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management (HNICEM)*, pp. 1-5, 2015.

[13] A. Ambarwari, Y. Herdiyeni and T. Djatna, "Combination of relief feature selection and fuzzy K-nearest neighbor for plant species identification," *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 315-320, 2016.

[14] B. VijayaLakshmi and V. Mohan, "Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture, shape, and color features," *Computers and Electronics in Agriculture*, vol. 125, pp. 99–112, July 2016.

[15] R. M and Y. Herdiyeni, "Shape and vein extraction on plant leaf images using fourier and B-Spline modeling," *AFITA International Conference, the Quality Information for Competitive Agricultural Based Production System and Commerce*, pp. 306–310, 2010.

[16] A. Ambarwari, Y. Herdiyeni, and I. Hermadi, "Biometric Analysis of Leaf Venation Density Based on Digital Image," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 16, no. 4, pp. 1735–1744, August 2018.

[17] G.L. Grinblat, L.C. Uzal, M.G. Larese and P.M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp.418-424, September 2016.

[18] J. W. Tan, S. Chang, S. Binti Abdul Kareem, H. J. Yap and K. Yong, "Deep Learning for Plant Species Classification using Leaf Vein Morphometric," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018.

[19] H.F. Eid and A. Abraham, "Plant species identification using leaf biometrics and swarm optimization: A hybrid PSO, GWO, SVM model," *International Journal of Hybrid Intelligent Systems*, 14(3), pp.155-165, 2017.

[20] A. Wahyumiyanto, I. K. E. Purnama, and Christyowidiasmoro, "Identifikasi tumbuhan berdasarkan minutiae tulang daun menggunakan SOM kohonen," Institut Teknologi Sepuluh Nopember, Surabaya (ID), 2011.

[21] Z. Wang, H. Li, Y. Zhu and T. Xu, "Review of plant identification based on image processing," *Archives of Computational Methods in Engineering*, vol. 24, pp.637-654, July 2017.

[22] G. Dhingra, V. Kumar and H.D. Joshi, "Study of digital image processing techniques for leaf disease detection and classification," *Multimedia Tools and Applications*, vol. 77, pp.19951-20000, 2018.

[23] M.G. Larese and P.M. Granitto, "Finding local leaf vein patterns for legume characterization and classification," *Machine Vision and Application*, vol. 27, pp.709-720, 2016.

[24] A. Salima, Y. Herdiyeni, and S. Douady, "Leaf vein segmentation of medicinal plant using hessian matrix," in *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Depok (ID), pp. 275–279. 2015.

[25] A. D. Prastya, "Ekstraksi fitur venasi daun tumbuhan obat berbasis geometri," Institut Pertanian Bogor, Bogor (ID), 2016.

[26] A. Ambarwari, Y. Herdiyeni, and I. Hermadi, "Identification of Venation Type Based on Venation Density using Digital Image Processing," *Jurnal Teknoinfo*, vol. 12, no. 2, pp. 87–92, 2018.

[27] J. Bühler *et al.*, "Phenovein - A tool for leaf vein segmentation and analysis," *Plant Physiology*, vol. 169, pp. 2359–2370, 2015.

[28] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, June 1998.

[29] Z. Wang, X. Sun, Y. Zhang, and Z. Ying, "Leaf recognition based on PCNN," *Neural Comput & Applic*, vol. 27, no. 4, pp. 899–908, 2016.

[30] R. Gonzalez, R. Woods, and S. Eddins, "Digital Image Processing Using MATLAB," New Jersey (US): Pearson Prentice Hall, 2004.