■ 1735

Biometric Analysis of Leaf Venation Density Based on Digital Image

Agus Ambarwari*[1], Yeni Herdiyeni[2], Irman Hermadi[3]
[1]Department of Informatics Engineering, Universitas Teknokrat Indonesia
Jl. Zainal Abidin Pagar Alam No.9-11, Kedaton, Bandar Lampung 35132, Indonesia
[2,3]Department of Computer Science, Bogor Agricultural University
Jl. Meranti, Wing 20 Level 5, Darmaga, Bogor 16680, Indonesia
*Corresponding author, e-mail: ambarwariagus@teknokrat.ac.id

Abstract

The density level in the leaf venation type has different characteristics. These different characteristics explain the environment in which plants grow, such as habitat, vegetation, physiology and climate. This research aims to measure of leaf venation density, leaf venation feature analysis and then identifying plants based on venation type. Stages of this research include leaf image data collection, segmentation, vein detection, feature extraction, feature selection, classification, evaluation and ending with analysis. The results of this study indicate that the level of leaf venation density is quite good is the type of venation paralellodromous, acrodromous and pinnate. Based on the selection of features using Boruta Algorithm, obtained 19 most important features that represent the type of leaf venation. This is reinforced by the average of accuracy produced at the time of classification using SVM, which amounted to 77.57%.

Keywords: Biometrics, Feature extraction, Feature selection, Leaf venation density

1. Introduction

Plant morphology is the study of physical form and body structure of plants [1]. It is useful for identifying plants visually so that the vast diversity of plants can be identified and classified and named for each group formed. Plant morphology not only describes the shape and structure of the plant body but also to determine the function of each part in the plant life, then can be known where the origin and the composition of the body that formed [2]. Morphological information is needed in the understanding of life cycle, geographic spread, ecology, evolution, conservation, and defining the species [3].

Parts of plants that have different characteristics between one plant with another plant and often used to identify species are leaves [4]. Leaves have the main features that distinguish each type of plant, among others the form, the structure of the veins (venation) [5] [6], and texture [6]. Among the features, leaf venation has a unique diversity that can describe plant characteristics in more detail, although some plant species do not exhibit clear patterns of venation [7]. The leaf venation network provides an integrative relationship between plant shape, function and climate, including temperature, precipitate and water availability [8].

Associated with the growth of plants, leaf venation has a very important role. Plant growth will experience water release during transpiration. Transpiration itself requires water supply provided by leaf venation. If the growth rate and water release rate of the plant depends on the environment, then only a few types of leaf venation with a certain density can survive in each different environment [8]. Zalenski observed that the venation density of plants in dry habitats was higher than that of plants in mesic habitats (habitat types with adequate moisture or water levels) [9]. Therefore, to know the relationship of physiology and climate where the plant grows, the measurement of the leaf venation density is required. Density is one of the main characteristics of leaf venation because it is directly related to venous function [10].

This study analyzes leaf venation density using plant biometric features. To obtain the leaf venation density features, it is necessary to extract the feature of leaf venation. The results of leaf venation feature extraction will produce some features, such as straightness, different angle, length ratio, scale projection, skeleton length, number of the segment, total skeleton

length, projected leaf area, number of branching points and number of ending points. In some extracted features, doing the calculation of mean, variance and standard deviation were performed [11]. Of all the features that obtain, then the feature selection was performed to get the most important features. The result of feature selection used to classify the leaves based on the venation type. The classification technique used is the Support Vector Machine (SVM). In many cases such as pattern recognition and regression estimation, SVM performance (i.e., error rate at the time of data testing) is significantly better than other methods [12].

As for the benefits of this research that is in addition to texture features and shape of the leaf, leaf venation feature obtained can be used as an additional identifier to identify plants. Then, the results of this study are also expected to facilitate the work of botanists in identifying plants and also predict the environment where the plant grows, of course, with further research.

2. Research Method
2.1. Leaf image dataset
The leaf image dataset used was derived from Computational Intelligence Laboratory, Department of Computer Science Bogor Agricultural University. The number of collected leaf image data as much 271 leaves. This dataset consists of 53 plant species and has grouped according to the type of leaf venation as shown in Figure 1.
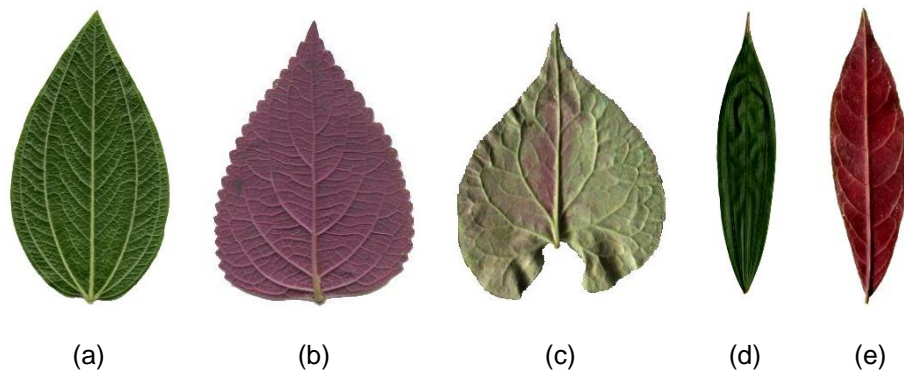

(a)   (b)   (c)   (d)   (e)

Figure 1. The leaf venation types (a) Acrodromous, (b) Actinodromous, (c) Campylodromous, (d) Parallelodromous, (d) Pinnate

2.2. Methodology
In general, the process stages in this study shown in Figure 2. Explanations for each stage are described in subsequent chapters.
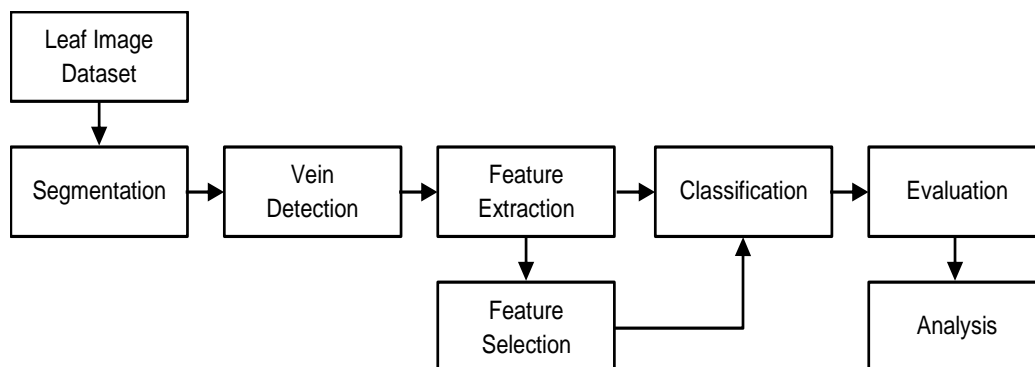

Figure 2. Methodology

2.3. Segmentation

Segmentation was done using the Hessian matrix [13] to obtain leaf venation images and thresholding to get leaf shape images. The leaf image data of segmentation result that is in the form of binary image data. Figure 3 illustrates the segmentation process.

Figure 3. Segmentation process to get (a) leaf shape, (b) leaf venation

2.4. Vein Detection

Vein detection is the stage to determine the branch-point (bifurcation) and end-point (extreme point) on the leaf venation image. The points are stored in the form of coordinate values (x, y), then will be formed an image segment consisting of two pairs of points as shown in Figure 4.
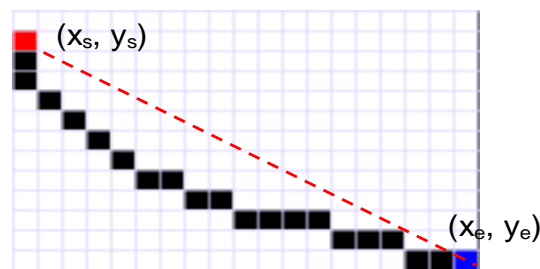
Figure 4. A segment with two pairs of point coordinates

The basic idea of branch-point and end-point determination is that when a pixel has only one or two neighboring pixels, the pixel is an end-point (illustrated as in Figure 5a). If a pixel has three or more neighboring pixels, then the pixel is a branch-point (illustrated as in Figure 5b).
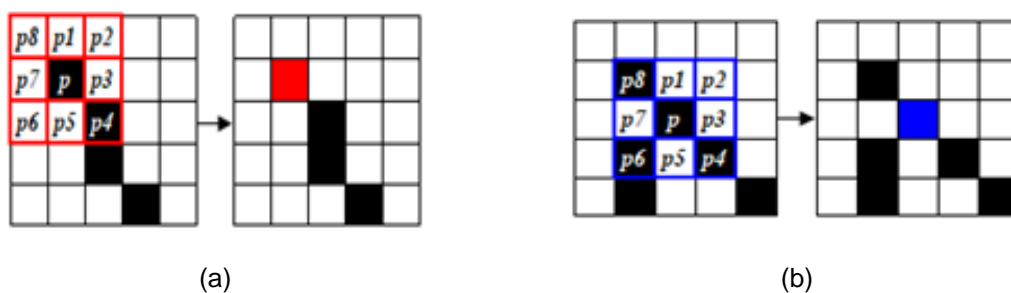
Figure 5. (a) illustration of end-point determination, (b) illustration of branch-point determination

2.5. Feature Extraction

The leaf venation feature extraction obtained from the calculation of straightness, different angle, length ratio and scale projection. To calculate the value of the feature, it takes the position of the corresponding point coordinates (one segment) of the segment extraction process as shown in Figure 4. Each segment in the leaf venation has a different length, the length of this segment is called the skeleton length. Figure 6 shows the illustration of extraction of leaf venation features
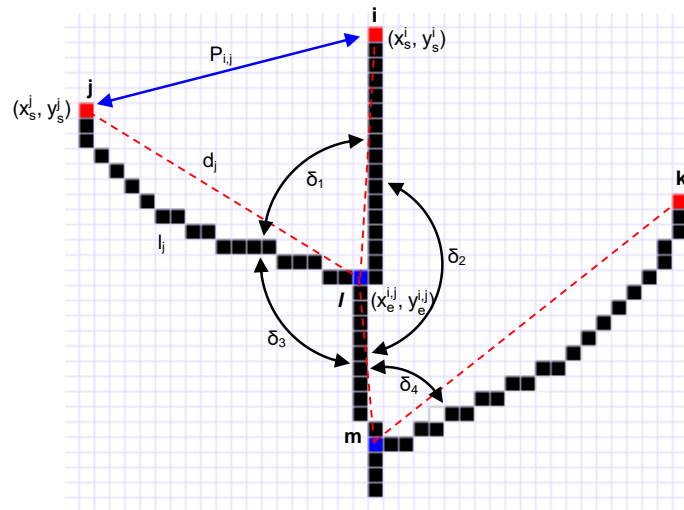

Figure 6. Illustration of the leaf venation feature extraction

2.5.1. Straightness

Straightness is the measurement of the straightness value of a segment. From the Figure 6, the straightness value can be calculated using equation 1.

$$\text{straightness} = \frac{l_j}{d_j} \tag{1}$$

where l_j is the length of the segment j represented by the number of pixels connected by the points j and point l, d_j is the distance between the pixel x_s^j, y_s^j coordinates with the pixel x_e^j, y_e^j coordinates. The coordinate values were determined from the segment extraction process. This value indicates the distance of straightness that has by the leaf venation. The value of d_j can be calculated using equation 2.

$$d_i = \sqrt{\left(x_e\text{-}x_s\right)^2 + \left(y_e\text{-}y_s\right)^2} \tag{2}$$

where x_s, y_s is the absys, ordinate from the initial pixel and x_e, y_e is the absys, ordinate from the final pixel.

2.5.2. Different Angle

The different angle is the measurement of the angular difference between segments that coincide. From Figure 6, different angle (δ_1) values can be calculated using equation 3.

$$\delta_1 = \left|\alpha_i\text{-} \alpha_j\right| \tag{3}$$

Each segment will calculate the angle (α) formed from the center of the segment with equation 4.

$$\alpha = \mathrm{acos}\left(\frac{x_e - x_s}{\sqrt{(x_e - x_s)^2 + (y_e - y_s)^2}}\right) \tag{4}$$

The resulting angle will have a range of "$0 \le \alpha \le \pi$." To produce αi, αj, and αk values, the angle should have the range "$0 < \alpha \le 2\pi$" . The angle values of that range are obtained by equation 5.

$$\alpha_i = \begin{cases} \alpha_i, \ (y_e - y_s) > 0 \\ \alpha, \ (y_e - y_s) < 0 \end{cases} \tag{5}$$

If the value $(y_e - y_s) < 0$, then the value of α_i can be determined using equation 6

$$\alpha_i = \begin{cases} \alpha - 2\pi, \ \alpha > 0 \\ \alpha + 2\pi, \ \alpha < 0 \end{cases} , 0 < \alpha \le 2\pi \tag{6}$$

After the angle value changed with the range $0 < \alpha \le 2\pi$, it can be determined the value of the different angle between the segments that coincide using equation 3.

2.5.3. Length Ratio

The length ratio value measured by comparing the length of each segment to the maximum length of the segment in a leaf venation image. From Figure 6, the value of the length ratio can be calculated using equation 7.

$$R_i = \frac{l_i}{\max(\vec{l})} \tag{7}$$

where R_i is the length ratio of the segment i, l_i is the length of the segment i and $\vec{l}$ is the segment vector in the leaf image.

2.5.4. Scale Projection

Scale projection is the measurement of projection length between segments that coincide. In Figure 6, to determine the length of the projection between segments i and j, can use equation 8 and derived to Equation 9.

$$P_{i,j} = \frac{\vec{i} \cdot \vec{j}}{(\max(|i|,|j|))^2} \tag{8}$$

$$P_{i,j} = \frac{(x_e^i - x_s^i)(x_e^j - x_s^j) + (y_e^i - y_s^i)(y_e^j - y_s^j)}{(\max(d_i,d_j))^2} \tag{9}$$

The value of skeleton length, straightness, different angle, length ratio, and scale projection on each leaf data has more than one value. So in this research, the five features were measured density using the statistical approach that is mean, variance, and standard deviation that refer the research of Plotze and Bruno [11]. In addition to these features, also obtained other features such as skeleton length, number of segments and features used to measure the leaf venation density.

2.6. The Leaf Venation Density Measurement

The leaf venation density measurement consists of three, which is leaf venation density, branch point density, and end point density. Leaf venation density, density of the branching point and density of the branching end points can be calculated using the equation 10-12 [14].

$$D_V = \frac{L \ (total \ skeleton \ length)}{A \ (projected \ leaf \ area)} \tag{10}$$

$$D_{BP} = \frac{N_{BP} \ (no. \ of \ branching \ points)}{A \ (projected \ leaf \ area)} \tag{11}$$

$$D_{EP} = \frac{N_{EP} \text{ (no. of ending points)}}{A \text{ (projected leaf area)}} \qquad (12)$$

2.7. Feature Selection of the Leaf Venation

At the feature selection stage used Boruta Algorithm. Boruta is a feature selection algorithm that works as a wrapped algorithm around the Random Forest [15]. Boruta iteratively compares the original attribute with the shadow attribute (i.e., random data from a copy of all attributes). Attributes that have lower importance than the shadow attribute are marked as Rejected and removed from the system. On the other hand, attributes that have a higher importance than the shadow attribute are marked as Confirmed. The shadow attribute is recreated on each iteration. The algorithm stops when only the remaining Confirmed attributes are present, or when it reaches the specified iteration threshold.

2.8. SVM Classification

Model development was done by using SVM classifier with RBF Gaussian kernel function. In the RBF Gaussian kernel function required parameters C and γ. The value of C parameter tested were $[2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8, 2^9]$ and the value of γ parameter tested were $[2^{-3}, 2^{-2}, 2^{-1}, 2^0]$. To get the best parameter value, each value of C parameter is combined with the value of γ parameter, then in each combination was applied 4-Fold Cross Validation method. The best C and γ combination is which has the greatest accuracy value.

3. Results and Analysis

The collected leaf image data consisted of 5 types of venation, with the details of acrodromous as much 60 leaf images, actinodromous as much 32 leaf images, campylodromous as much 32 leaf images, parallelodromous as much 11 leaf images, and pinnate as much 136 leaf images. From this leaf images data is then done segmentation and vein detection. After that feature extraction of the leaf venation and obtained as many as 23 features, among others mean, variance, standard deviation of (straightness; different angle; length ratio; scale projection; skeleton length), number of segment, total skeleton length, projected leaf area, number of branching points, and number of ending points.

3.1. Analysis of Leaf Venation Density

The density analysis on leaf venation was performed to see the distribution of density and clustering based on leaf venation type. The leaf venation density measurement consists of three, which is leaf venation density, branch point density, and end point density. The result of measurement of leaf venation density level of 5 types of venation shown in Figure 7.

According to Figure 7, of the five types of venations having an enough good density are the paralellodromous and acrodromous types. The pinnate type is relatively good; this is because the variety of data is the most. The venation type of actinodromous and campylodromous have a density less good because the density spreads from the range of 0.0 to 1.0. Data in the range 1.0 means having a high density, while the data in the range of 0.0 means having a low density. The examples of vein detection images for each venation type with high and low venation density levels shown in Table 1.

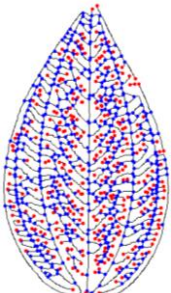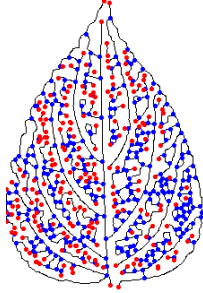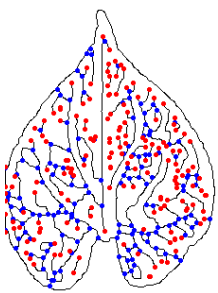Table 1. Comparison of Leaf Venation Density on Each Type of Venation
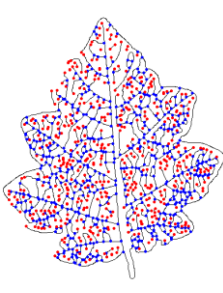
Acrodromous	Actinodromous	Campylodromous	Paralellodromous	Pinnate

Table 1. Comparison of Leaf Venation Density on Each Type of Venation
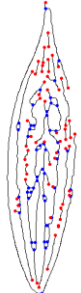
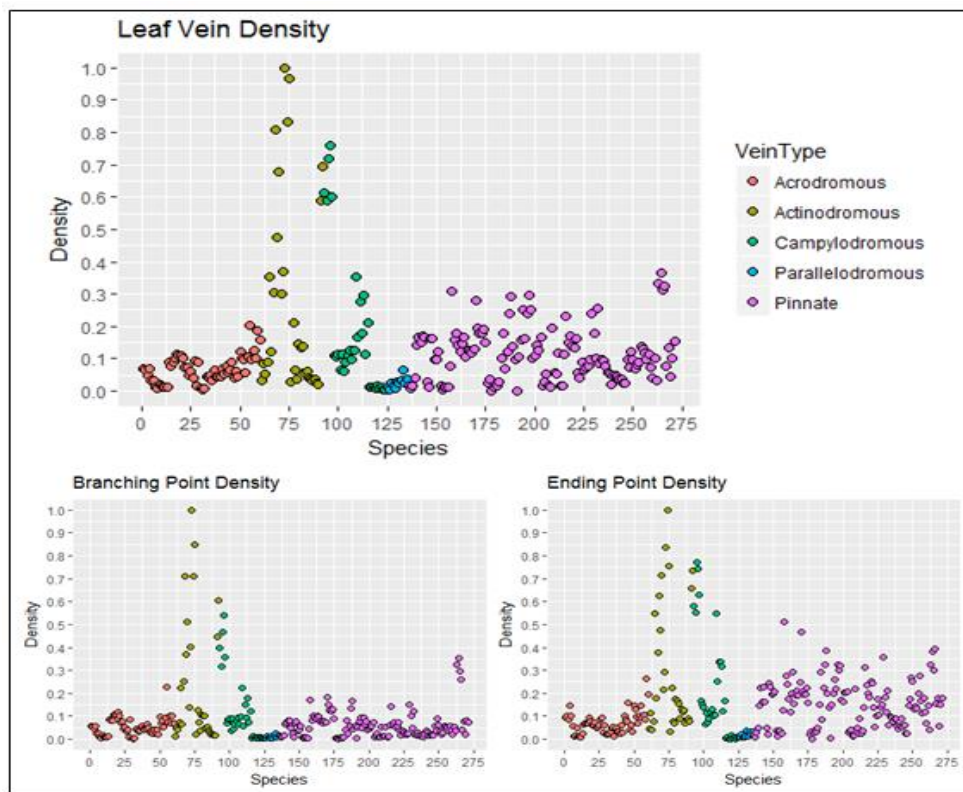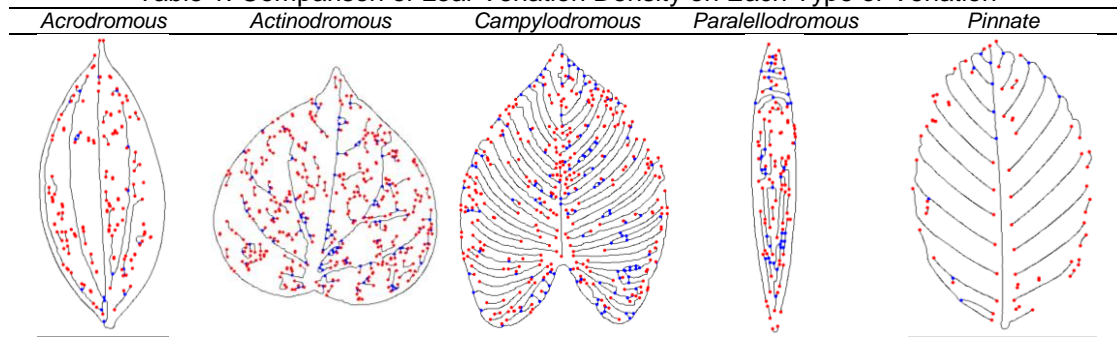| Acrodromous	Actinodromous	Campylodromous	Paralellodromous	Pinnate

Figure 7. Value of leaf venation density of 5 types of venation

3.2. The Result of Feature Selection Using Boruta Algorithm

Based on the features selection using Boruta Algorithm, from 23 leaf venation features obtained 19 features that are best or relevant to the venation type. The best features selection is base on the mean, median, min and max values of each leaf venation feature. The result of feature selection using Boruta Algorithm shown in Table 2.

Table 2. The Result of Feature Selection Using Boruta Algorithm

The leaf venation feature	meanImp	medianImp	minImp	maxImp	normHits	decision
Mean of diff. angle	6.31	6.33	4.99	7.76	1.00	Confirmed
StdDev. of diff. angle	5.44	5.28	4.53	7.07	1.00	Confirmed
Var. of diff. angle	5.10	4.98	3.93	6.53	1.00	Confirmed
Mean of length ratio	10.92	10.80	9.98	12.34	1.00	Confirmed
StdDev. of length ratio	11.30	11.46	9.95	12.29	1.00	Confirmed
Var. of length ratio	11.24	11.13	10.14	12.78	1.00	Confirmed

Table 2. The Result of Feature Selection Using Boruta Algorithm

The leaf venation feature	meanImp	medianImp	minImp	maxImp	normHits	decision
Mean of scale projection	11.99	11.96	10.92	13.58	1.00	Confirmed
StdDev. of scale projection	1.28	1.26	-0.86	3.01	0.06	Rejected
Var. of scale projection	1.27	1.16	0.34	2.87	0.00	Rejected
Mean of straightness	7.86	7.81	6.66	9.12	1.00	Confirmed
StdDev. of straightness	0.23	0.36	-1.66	1.66	0.00	Rejected
Var. of straightness	0.66	0.59	-0.93	2.39	0.06	Rejected
Mean of skeleton length	12.43	12.39	11.47	13.13	1.00	Confirmed
StdDev. of skeleton length	6.97	7.14	5.38	8.33	1.00	Confirmed
Var. of skeleton length	6.77	6.90	4.64	7.88	1.00	Confirmed
Projected leaf area	18.54	18.56	16.71	20.14	1.00	Confirmed
Total skeleton length	18.55	18.82	16.96	19.38	1.00	Confirmed
No. of ending points	15.89	15.69	14.54	17.85	1.00	Confirmed
No. of branching points	15.17	15.05	14.07	17.17	1.00	Confirmed
No. of segment	14.25	14.24	13.60	15.24	1.00	Confirmed
Leaf vein density	13.84	13.81	13.01	15.34	1.00	Confirmed
Branch point density	11.69	11.53	10.25	12.96	1.00	Confirmed
End point density	17.41	17.31	15.95	19.12	1.00	Confirmed

3.3. Classification Results using SVM

The initial stage of the SVM classification is the parameter selection in the RBF kernel. Then the dataset was divided into training data and testing data, with portion 75% of training data and 25% of testing data. So that obtained as much as 203 of training data and 68 of testing data. In this research, we tested two types of datasets, which are the dataset with all features (23 features) and the dataset the result of feature selection (19 features). From each of these datasets, the best combination of C and γ parameters that selected for classification modeling shown in Table 3.

Table 1. The Best C and γ Parameter Values That Selected for Each Dataset

No	Dataset	C parameter	γ parameter	The average of model accuracy
1	Selected feature (19 features)	23	20	69.32%
2	All features (23 features)	23	20	68.59%

After the best parameters obtained, the next step is the selection of the classification model. The selection of the classification model of the four models built using the best C and γ value pairs is to look at the accuracy of each model produced. In this research, we selected the best classification model with the best accuracy, i.e., in the 3rd fold with 85.29% accuracy for the dataset with Boruta feature selection and 83.82% for the dataset with all feature.

Table 4. Details of Accuracy (in %) for Each Model That Formed from the Best C and γ Pair

Fold	Accuracy	
	Dataset with Boruta feature selection	Dataset with all features
1	66.18	63.24
2	70.59	72.06
3	85.29	83.82
4	55.22	55.22
Average	69.32	68.59

From the selected classification model is then evaluated. Evaluation of classification model was done by testing the accuracy of each class. The total of test class is 68 data for the whole class, with the portion of acrodromous as much 15, actinodromous as much 8, campylodromous as much 8, parallelodromous as much 3, and pinnate as much 34.

Based on Table 5, the average of accuracy obtained from the test results for each type of leaf venation on the dataset with Boruta feature selection is 77.57%, and the dataset with all features is 76.82%. This shows that the dataset with feature selection using Boruta Algorithm has the highest level of accuracy. Although by reducing the original features to 19 features, but the information contained in the data is still maintained.

Table 5. The Percentage Values of Precision, Recall, and Accuracy from Each Dataset

Venation type	Dataset with Boruta feature selection			Dataset with all features		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Acrodromous	0.8000	0.8000	0.8000	0.8462	0.7333	0.7333
Actinodromous	1.0000	0.5000	0.5000	1.0000	0.5000	0.5000
Campylodromous	0.8000	1.0000	1.0000	0.8000	1.0000	1.0000
Parallelodromous	1.0000	0.6667	0.6667	1.0000	0.6667	0.6667
Pinnate	0.8378	0.9118	0.9118	0.8205	0.9412	0.9412
Average	0.8876	0.7757	0.7757	0.8933	0.7682	0.7682

In addition to the calculation of accuracy, to measure the performance of a classification model also calculated precision and recall. Comparison of the average value of precision and recall from each dataset among others on the dataset with all features, the precision of 89.33% and recall of 76.82%. While in the dataset with Boruta feature selection, the precision of 88.76% and recall of 77.57%. The difference in precision and recall values between the dataset with all features and the dataset with Boruta feature selection is not too much. This suggests that the effectiveness or performance of the classification model built using these two datasets is equally good. So it can be said that the use of feature selection with Boruta Algorithm is effective, because of the better accuracy. Besides, the use of feature selection on SVM classification can speed up computing time.

4. Conclusion

Of the five types of leaf venation, acrodromous and paralellodromous types have the density of the venation is good, whereas the pinnate type of density is relatively good compared with the actinodromous and campylodromous whose venation density is most diverse. The leaf venation features of the extraction result, which has the most important information on the venation type are as much as 19 features. The result of classification using Support Vector Machine (SVM) for leaf identification based on type of venation, obtained the highest accuracy rate of 77.57%, that is when using the selected feature with Boruta Algorithm.

References
[1]   Evert RF, Eichhorn SE. Raven Biology of Plants 8th Edition. New York: Freeman and Company. 2013.
[2]   Tjitrosoepomo G. Morfologi Tumbuhan. Yogyakarta: Gadjah Mada University Press. 2016.
[3]   Douaihy B, Sobierajska K, Jasinska AK, Boratynska K, Ok T, Romo A, Machon N, Didukh Y, Dagher-Kharrat MB, Boratynski A. Morphological versus molecular markers to describe variability in juniperus excelsa subsp. excelsa (cupressaceae). AoB Plants. 2012; (0): 1-14.
[4]   Le TL, Tran DT, Hoang VN. *Fully automatic leaf-based plant identification, application for Vietnamese medicinal plant search.* Proceedings of the Fifth Symposium on Information and Communication Technology - SoICT '14. New York. 2014: 146-154.
[5]   Rahmadhani M, Herdiyeni Y. *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. Bogor. 2010: 306-310.
[6]   Cope JS, Remagnino P, Barman S, Wilkin P. *Plant texture classification using Gabor co-occurrences. Advances in Visual Computing*: 6th International Symposium. Las Vegas. 2010: 669-667.
[7]   Wahyumiyanto A, Purnama IKE, Christyowidiasmoro. Identifikasi tumbuhan berdasarkan minutiae tulang daun menggunakan SOM kohonen. Thesis. Surabaya: Institut Teknologi Sepuluh Nopember; 2011.
[8]   Blonder B, Enquist BJ. Inferring climate from angiosperm leaf venation networks. New Phytol. 2014; 204(1):116-126.

[9]   Zalenski WV. Ueber die ausbildung der nervation bei verschiedenen pflanzen. Berichte Deutsche Botanicshe Gesellschaft. 1902; 20:433-440.

[10]   Zhu Y, Kang H, Xie Q, Wang Z, Yin S, Liu C. Pattern of leaf vein density and climate relationship of quercus variabilis populations remains unchanged with environmental changes. Trees. 2011; 26(2):597-607.

[11]   Plotze RDO, Bruno OM. Automatic leaf structure biometry: computer vision techniques and their applications in plant taxonomy. *International Journal of Pattern Recognition and Artificial Intelligence.* 2009; 23(02): 247-262.

[12]   Burges CJC. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 1998; 2:121-167.

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

[14]   Bühler J, Rishmawi L, Pflugfelder D, Huber G, Scharr H, Hülskamp M, Koornneef M, Schurr U, Jahnke S. Phenovein -A tool for leaf vein segmentation and analysis. Plant Physiology. 2015; 169:2359-2370.

[15]   Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. *Journal of Statistical Software.* 2010; 36(11):1-13.