

Weighted Ensemble Classifier for Plant Leaf Identification

R. Putri Ayu Pramesti*, Yeni Herdiyeni, Anto Satriyo Nugroho

Department of Computer Science, Bogor Agricultural University,
Jl. Meranti Wing 20 Level V Kampus Dramaga IPB, Bogor, Indonesia

*Corresponding author, e-mail: putri_pramesti@apps.ipb.ac.id

Abstract

Plant leaf identification using image can be constructed by ensemble classifier. Ensemble classifier executes classification of various features independently. This experiment utilized texture feature and geometry feature of plant leaf to find out which features are more powerful. Each classifier trained by specific feature produced different accuracy rate. To integrate ensemble classifier the results of the classification were weighted, so as the score obtained from better features contributed greater to the final results. Weighted classification results were combined to get the final result. The proposed method was evaluated using dataset comprises of 156 variety of plants with 4559 images. Weighting and combining classifier used in this study were Weighted Majority Vote (WMV) and Naïve Bayes Combination. Both of those method result showed better accuracy than using single classifier. The average accuracy of single classifier was 61.2% for geometry classifier and 70.3% for texture classifier, while WMV method was 77.8% and Naïve Bayes Combination was 94.6%. The calculation of classifier's weight by using WMV method produces a weight value of 0.54 for texture feature classifier and 0.46 for geometry feature classifier.

Keywords: ensemble classifier, weighted classifier, naïve bayes combination, weighted majority vote

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Indonesia is one of the countries having a high plant diversification [1]. High plant diversity is frequently used for various purposes, one of which is for researches. Previous studies that utilized plant diversity were abundance which resulted in a large number of research data availability. Some of these data were presented in the form of images. Despite the availability of images, efficient usage is hindered due to the existing process for searching images' that use only the common features of images, such as texture and color. Therefore, we argue that there is a need for an image retrieval system which runs based on specific characteristics or features of the image.

From a leaf image, there are various features that can be extracted, such as leaf shapes, venation, and texture [2]. In the previous studies [3] using color and [4] using both color and texture to retrieve image, while [5] conducted a study using features such as leaf and geometry texture to retrieve images of leaf plants [6] use Local Binary Pattern to extract leaf texture then classify the image use Probabilistic Neural Network. To distinguish one type of leaf from another, one can observe the shape of the leaves. However, there are many types of leaves that have similar shapes. Thus, apart from the shapes of leaves, other characteristics or features can also be put into consideration when distinguishing leaves' types or species, such as its texture feature. Therefore, plant identification using leaf images requires combining several features in order to get better result.

Ensemble classifier method, also known as Multiple Classifier System (MCS), includes the use of several features for the classification process. MCS is used because each available feature has different information and characteristics, which resulted to the need of using different training process. MCS is one of classification methods which combines the results of several independent classification processes which is hoped to enhance single classifier's classification performance [7]. All features used are later classified separately by using certain classification techniques [8] using SVM for classification process on ensemble classifier. In that research, the data resulted from feature extraction were divided into several parts. Each part

then is treated separately in terms of its classification process using SVM. The last decision is taken after the best-performance classifier is known. This kind of decision making will ignore other classifiers with lower performance. Thus, a method that can be used to combine every classification result on ensemble classifier is needed.

There is method to combine ensemble classifier such as PDR (Product Decision Rule) [9]. PDR method uses prior and posterior probability value of each classifier in determining the final decision. Other than that, there are several ensemble classifier techniques used to combine the result of classifications that can be applied in ensemble classifier, one of which is simple vote as used in a research [10] and Majority Vote [11]. However, both of those techniques applied the same weight to all classifiers used, therefore it takes a method of combining that is influenced by the weight of each classifier, so it will show which features are more powerful depending on the weight of classifier. Naïve Bayes combination method also uses the value of each classifier prior to the calculation of the combined probability. Naïve Bayes Combination method performed by calculating combined probability input data by using two or many classifier are mutually independent. The calculation of these probability is obtained using prior probability and the amount of data in the class predictions of each classifier. Basically, each feature is different from each other in terms of its accuracy result. Hence, a certain classifier weighting technique is required to determine the influence of each classifier on its classification results. Combination method with the weighting on the so-called weighted ensemble classifier classifier. [12] used weighted ensemble classifier methods namely Weighted Majority Vote to determine the final outcome of the ensemble classifier. WMV uses the weight values obtained from the accuracy values during training phase. These weights will later be used in the calculation of the probability of data entered into each set of existing classes

The objective of this paper are and to show wich feature is stronger to classify plant leaf using image and to combine the ensemble classifier using WMV and compare it result with Naive Bayes Combination in plant image retrieval system which are based on texture and geometry features found in the leaf image. This paper consists of: Section 2, about the feature extraction; Chapter 3, introduction of ensemble classifier and combination of ensemble classifier; Section 4, where the result of the experiment is shown, and Section 5, conclusion and summary.

2. Proposed Method

2.1 Feature extraction

Feature extraction is performed to obtain significant information that can be used to identify and distinguish an object with other objects. In this research, we identify plants species using texture and leaves' geometry features. The texture feature extraction was done by using FLBP (Fuzzy Local Binary Pattern), while geometry feature extraction was retrieved from the characteristics of the area, circularity, eccentricity, and centroid-radii.

2.1.1 Fuzzy local binary pattern (FLBP)

FLBP is used to get a better representation of the texture contained in the image. FLBP is a method that applies the concept of fuzzy on local binary pattern method. Fuzzification of LBP .approach include the transformation of input variables into fuzzy variables according to a set of fuzzy rules. Two fuzzy rules were choosen to obtain a binary value and fuzzy values based on the relationship between the value of circular sampling (p_i) and center pixel (p_{center}) similar with [13] and [14] research.

After we use two fuzzy rule, then two membership function will be formed. In FLBP, each LBP value will have different level of contribution, it depends on the value of the membership function of m_0 and m_1 . Then after obtaining LBP code, LBP code will be represented in the histogram.

2.1.2 Geometrical feature

There are many geometrical feature from leaf image that can be used to classify [15] using perimeter, area, roundness, and others to classify the object. Geometrical feature used in this study were the area, circularity, eccentricity and centroid radii. Geometry feature extraction process performed on the image of the leaves that had been converted into a binary image. Area feature is a feature that describes value of the leaf surface area. Circularity is a feature

that measures how an object is closely resembled as a perfect circle. Eccentricity explain how the regions points are scattered around the centre of the region. While the centroid-radii is a measurement of the distance between the center point of the leaf with the edges of the leaves.

2.2. Ensemble classifier combination

Ensemble classifier is a decision making method towards incoming data taken from some separate classification processes to enhance the stability of accuracy from a single classifier [16]. Ensemble is often called Multiple Classifier System (MCS). MCS is used because every existing feature has different information and characteristics, leading to the conducting of training processes with different classifiers. MSC has four approaches [17], which are MCS with different combination schemes, MCS with different models, MCS with different feature subsets, and MCS with different training sets. These four MSC approaches can be seen in Figure 1.

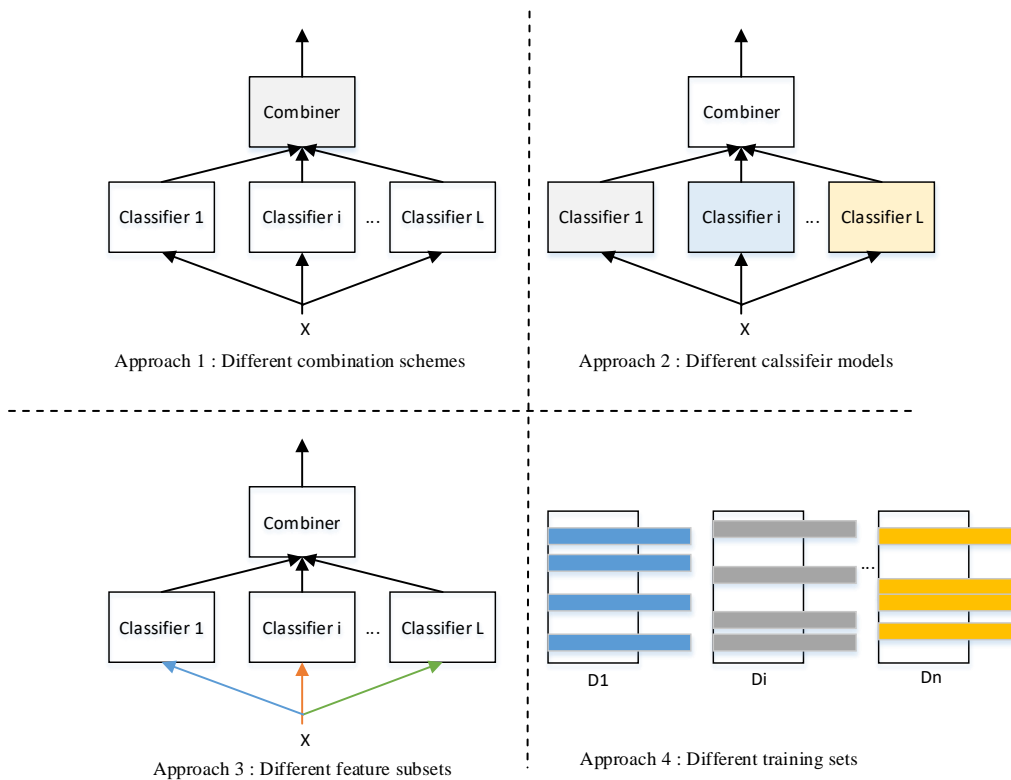


Figure 1. Ensemble classifier approaches

This research employs the third approach, which is MCS or ensemble classifier with different feature subsets. This approach is chosen since the two different features from texture and geometry characteristics are extracted from the leaf to be used in this research. Later, both features will be adopted in conducting classification processes, which will be done separately by using SVM.

In order to get the final results from several classifiers, the process of combining classifiers with *ensemble classifier* needs to be done. There are three ensemble classifier combination categories that can be used. Those are *simple vote strategy*, *weighted classifier ensemble*, and *selective or pruning classifier*. *Ensemble classifier* combination method using simple vote strategy combines the results of several classification processes with same coefficient or weight classifiers for entire classifiers that are used. Second combination method is weighted classifier ensemble. Weighted classifier ensemble method combines several classification result by weighting each classifier first. The last combination method is selective or pruning method. This method combines the result of several classifications with a weight vector for every classifier, including classifier with zero weight value, to indicate classifier not having a

significant influence or even negative influence on the classification. This research will use the ensemble classifier combination with weighted ensemble classifier methods *weighted majority vote* and compare the result obtained by *naive bayes combining* method.

Weighted ensemble classifier combines ensemble classifiers by weighting each classifier first. This research will use ensemble classifier combination with weighted classifier methods, namely *weighted majority vote* and compare the result with that obtained by *naive bayes combination* method.

2.2.1. Weighted majority vote

Weighted Majority Vote (WMV) is a decision-making method retrieved from several classifiers that are separated and independent by giving each classifier more abilities [18]. The scheme from this classifier incorporation can be seen in Figure 2.

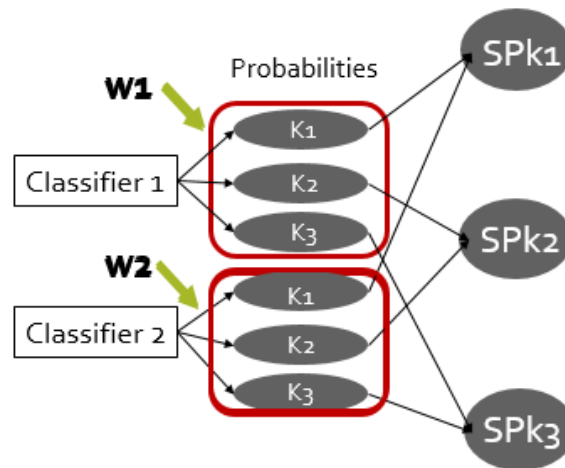


Figure 2. The scheme of classifier and weighting combination

To determine the final decision, each classifier will be given weight that is suitable with its accuracy value, retrieved during training phase. These weights are a measure of each classifier's influence on the result of the final decision. Classifier's weight is a value that will later be multiplied by the probabilities of a certain object entering each class. The class with the highest multiplying result will eventually become the class from that new object, as seen in equation (1).

$$p_j^{wmv}(x_r) = \sum_{i=1}^L w_i d_{ij}, \quad j \in \{1, \dots, C\} \quad (1)$$

Where, x_r is incoming data whose probability towards entire class will be calculated. $p_j^{wmv}(x_r)$ is data x_r 's probability that is included in class ω_j . L conveys the number of classifiers. C states the total number of classes, while w_i is coefficient weight from classifier to i , and d_{ij} is x_r 's probability to enter class ω_j by classifier to i . Before combining the result of classifiers from both features, each probability is multiplied by the weight value of each classifier retrieved from equation (2) [15].

$$w_i = \log \frac{Acc_i}{1 - Acc_i} \quad (2)$$

Where, Acc_i represent classifier's accuracy value to i during training phase. This number is used as a multiplying number for the probability of data's entering each existing class. The higher the classifier's accuracy during the training phase, the higher the weight will be, and vice versa; the lower the accuracy, the lower the weight will be. This is conducted in order to distinguish the contribution of each classifier to the final result.

2.2.2. Naïve Bayes Combination

In *naïve bayes combination*, it is assumed that each classifier does not depend on each other when it comes to determining class' lable. In determining the lable prediction of an object, the calculation is done using confusion matrix. Confusion matrix is a matrix which was created based on real class' lable from training data and on prediction lable produced by classifier. According to confusion matrix, prediction lable $H(x_r)$ from the sample is determined by using equation (3) [16].

$$H(x_r) = \arg \max_{j=1..C} \left(\frac{1}{N_j^{N-1}} \prod_{i=1}^L cm_{j,k_i}^i \right) \tag{3}$$

Where, N_j is number of actual samples with class label ω_j in the training data set, L is number of *classifier*, and cm_{j,k_i}^i is number of samples that belong to ω_j but goes to ω_k by *classifier-i*.

3. Results and Discussion

3.1. Image classification

To test the proposed ensemble classifier combination method, an experiment using IPBiotics' leaf image data which consists of 4559 images of 156 species was conducted. Separate classification was used on the data retrieved from the feature extraction results of the leaf images. There are two feature extractions that were conducted. Those are texture feature extraction and geometry feature extraction. The method used to extract these features was similar to the one used in [17]. After each feature was submitted to separated classification process by using SVM classification method, the classification results from both classifiers were combined using two proposed methods, which are *wighted majority vote* and *naïve bayes combination*. To see the accuracy of each ensemble classifier combination methods, an evaluation using k-fold cross validation method with k=5 was conducted. The accuracy of both combination methods are can be seen in Figure 3.

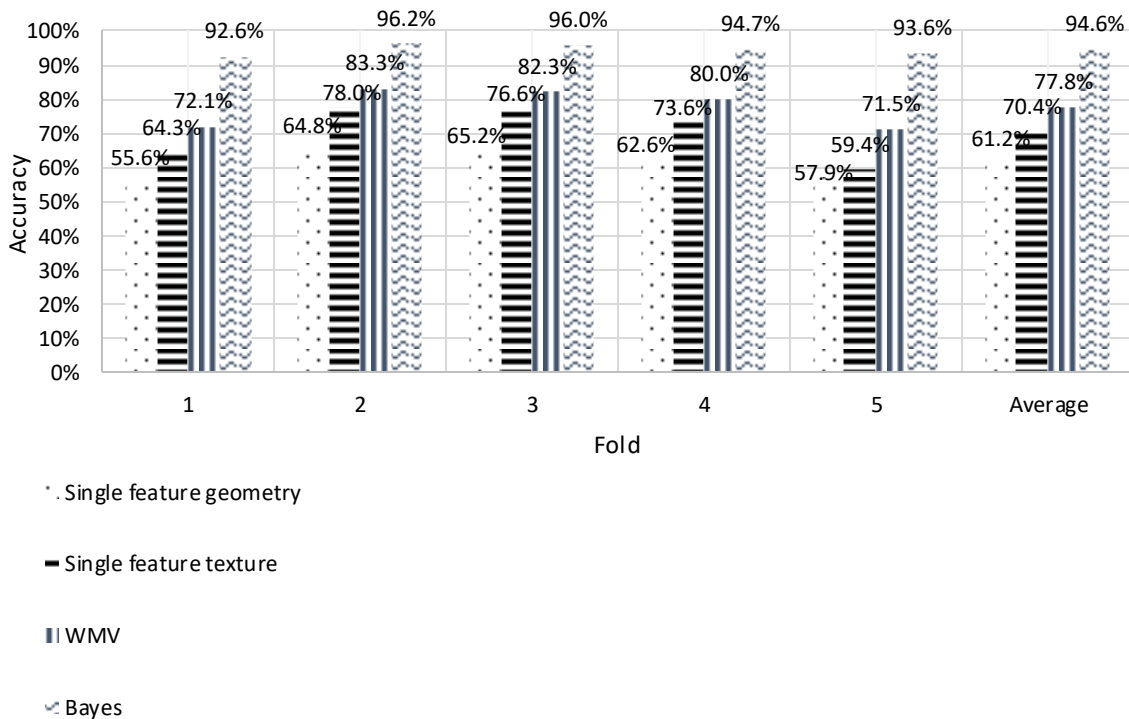


Figure 3. The correct classification rate using WMV's and Naïve Bayes combination methods

After measuring the accuracy of each method, the result showed that ensemble classifier method resulted in higher accuracy point compared to single feature and classifier. The average accuracy of Naïve Bayes Combination method was 94.6%, while WMV was 77.8%. It was stated in a research [13] that Naive Bayes Combination method is actually reliable when using small number of classifiers and large classes, but its accuracy decreases when many classifiers are used. In this research, the number of classes used is in accordance to the number of species from the data, which is 156. Therefore, the accuracy of Naive Bayes Combination method's result is higher than Weighted Majority Vote method. In addition, dissimilar to WMV and Naïve Bayes Combination methods that use the same classifier weight values for the combined probability of data going into each species, the calculations using the Naïve Bayes Combination method use different weights for each species. This weight is taken from the amount of data that is in the predicted class of both classifier, so each species will be given different treatments according to which feature is better for classifying the species. Therefore, by using a combination of Naïve Bayes, species that can not be classified well by the geometry classifier can still be classified with texture classifier, and vice versa.

The calculation of classifier's weight by using WMV method produces a weight value of 0.54 for texture feature classifier and 0.46 for geometry feature classifier. Based on those weight values, it can be concluded that texture feature has greater contribution in determining the classification's final result.

3.2 Result analysis

After obtaining accuracy rate of the two methods of ensemble classifier combination, it showed that classifier combination using Naïve Bayes Combination method scored higher accuracy. From the results of these experiments there were 131 species that successfully predicted correctly with an 100% of accuracy. It means that none of the images leaves incorrectly predicted in these species. Figure 4 shows a few leaves images of the species that have been successfully predicted correctly by 100%, the species include *Codiaeum variegatum*, *Centella asiatica*, *Melastoma malabatrikum* and *Coleus scutellarioides*.



Figure 4. *Codiaeum variegatum*, *Centella asiatica*, *Melastoma malabatrikum*, *Coleus scutellarioides*

From the experimental results there were also some species of the leaves are not successfully predicted correctly and got low accuracy. Such species include *Acanthus ilicifolius*, *Alyxia reindwardtu*, *Coleus scutellarioides*, *Kalanchoe pinnata*, *Melastoma Malabatrikum*. Examples of failed species-class images and examples of images in predicted species can be seen in Figure 5.

In Figure 5 shows that the leaf species have in common texture and geometrically with leaf species predicted results. From the data that were tested, several species of leaf that have 0% accuracy have a few of data training, while the other species have a much larger data training. Lack of image data in some species was one cause of low accuracy.

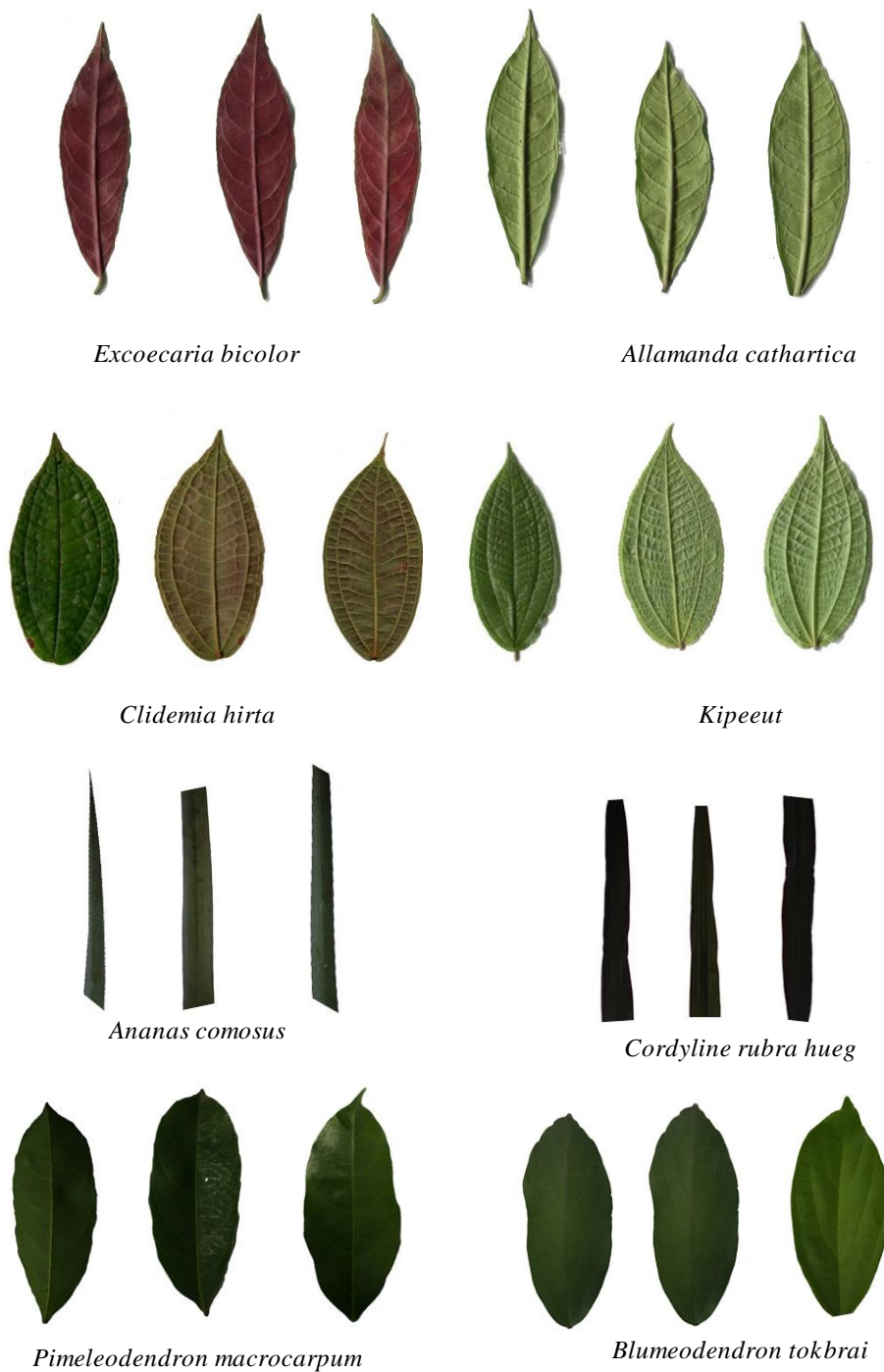


Figure 5. Picture of species leaves compared with leaf predictions results image

4. Conclusion

In this study, the combination of the ensemble classifier for texture classifier and geometry classifier were built using one of the methods weighted ensemble classifier, it is weighted majority vote and Naïve Bayes Combination. The calculation of classifier's weight by using WMV method produces a weight value of 0.54 for texture feature classifier and 0.46 for geometry feature classifier, in other words the classifier texture has a greater weight in the determination of the final classification result than the geometry classifier. The average accuracy point of WMV method was 77.8%, while naïve bayes combination's was 94.6%. This

result showed that ensemble classifier method performs better in this study. It showed that Naive Bayes Combination method is actually reliable when using small number of classifiers and large classes.

References

- [1] Badan Perencanaan Pembangunan Nasional. Indonesian Biodiversity and Action Plan 2015-2020. Jakarta (ID): Bappenas. 2016:65.
- [2] Cope JS, Remagnino P, Barman S, Wilkin P. *Plant texture classification using gabor co-occurrences. Advances in Visual Computing: 6th International Symposium (ISVC)*. Las Vegas. 2010;669-667.
- [3] Patil JK, Kumar R. Plant Leaf Disease Image Retrieval Using Color Moments. *IAES International Journal of Artificial Intelligence (IJAI)*. 2013; 2(1); 36-42.
- [4] Yue J et al. Content-based image retrieval using color and texture fused features. *Mathematical and Computer Modelling*. 2011; 54(3-4):1121-1127.
- [5] Herdiyeni Y, Ginanjar AR, Anggoro MRL, Douady S, Zuhud EAM. *Medleaf: mobile biodiversity informatics tool for mapping and identifying Indonesian medicinal plants*. 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR). Fukuoka, Japan. 2015; 54-59.
- [6] Prasvita DS, Herdiyeni Y. MedLeaf: mobile application for medicinal plant identification based on leaf image. *International Journal on Advanced Science, Engineering and Information Technology*. 2013; 3(2): 5-8.
- [7] Mao et al. Weighted classifier ensemble based on quadratic form. *Pattern Recognition*. 2015; 48(5): 1688-1706.
- [8] Hu Zhongzhui, Cai Yunze, Li Ye, Xu Xiaoming. *Support Vector Machine based Ensemble Classifier*. Proceedings of the 2005 American Control Conference. Oregon, Portland. 2005; 745-749.
- [9] Herdiyeni Y, Nurfadhilah E, Zuhud EAM, Damayanti EK, Arai K, Okumura H. A Computer Aided System for Tropical Leaf Medicinal Plant Identification. *International Journal on Advanced Science Engineering Information Technology*. 2013; 3(1):54 - 59.
- [10] Kuncheva LI, Riodriguez JJ. A weighted voting framework for classifier ensembles. *Knowledge and Information Systems*. 2012; 38(2):259-275.
- [11] Brown G, Kuncheva LI. *Good and Bad Diversity in Majority Vote Ensembles 9th International Workshop on Multiple Classifier Systems, MCS 2010*. 2010;124-133
- [12] Doan S, Collier N, Xu H, Duy PH, Phuong TM. Recognition of medication information from discharge summaries using ensembles of classifiers. *BMC Med Inform Dec Mak*. 2012; 12(1):36-45.
- [13] Iakovidis DK, Keramidas EG, Maroulis D. 2008. *Fuzzy local binary patterns for ultrasound texture characterization*. Campilho A, Kamel M, editor. *Image Analysis and Recognition: 5th International Conference, ICIAR 2008*. Povia De Varzim. 2008; 5(750-759).
- [14] Hansen LK, Salamon P. Neural network ensemble. *Trans Pattern Anal*. 1990; 12:992-1001.
- [15] Bangun MB, Herdiyeni Y, Herliyana EN. Morphological feature extraction of Jabonâ€™s leaf seedling pathogen using microscopic image. *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*. 2016; 14(1):254-261.
- [16] Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, (NJ): John Wiley & Sons, Inc. 2004:126.
- [17] Guo G, Neagu D. *Similarity-based classifier combination for decision making*. 2005 IEEE International Conference on Systems, Man and Cybernetics. Waikoloa. 2005; 176-181.
- [18] Herdiyeni Y, Ginanjar AR, Anggoro MRL, Douady S, Zuhud EAM. *Medleaf: mobile biodiversity informatics tool for mapping and identifying Indonesian medicinal plants*. 2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR). Fukuoka. 2015; 54-59.