Plant Leaves Classification

Mohamed Elhadi Rahmani, Abdelmalek Amine, Mohamed Reda Hamou

GeCoDe Laboratory
Department of computer Science
University of Dr. Tahar Moulay
Saida, Algeria
r_m_elhadi@yahoo.fr; amine_abd1@yahoo.fr; hamoureda@yahoo.fr

Abstract:- **A leaf is an organ of vascular plant and is the principal lateral appendage of the stem. Each leaf has a set of features that differentiate it from the other leaves, such as margin and shape. This paper proposes a comparison of supervised plant leaves classification using different approaches, based on different representations of these leaves, and the chosen algorithm. Beginning with the representation of leaves, we presented leaves by a fine-scale margin feature histogram, by a Centroid Contour Distance Curve shape signature, or by an interior texture feature histogram in 64 element vector for each one, after we tried different combination among these features to optimize results. We classified the obtained vectors. Then we evaluate the classification using cross validation. The obtained results are very interesting and show the importance of each feature.**

Keywords:-** **Plants leaves classificatin; supervised classification; KNN; Decision tree; Naïve Bayes.

I. INTRODUCTION

For all forms of life, plants form the basic food staples, and this is just one reason why plants are important. They are the major source of oxygen and food on earth since no animal is able to supply the components necessary without plants. The fish we eat consume algae and the cattle we eat as beef feed on grass, so even if you are not a fan of salads, your food source relies on plants. Plants also provide animals with shelter, produce clothing material, medicines, paper products, reduce noise levels and wind speed, reduce water runoff and soil erosion. Coal is also produced from plant materials that were once alive. All that gives plants its important role in life on earth. For example, as natural resource managers, they must understand what they manage, and plant identification is a key component of that understanding. The ability to know, or identify plants allows them to assess many important rangeland or pasture variables that are critical to proper management: range condition, proper stocking rates, forage production, wildlife habitat quality, and rangeland trend, either upward or downward. Natural resource managers, especially those interested in grazing and wildlife management must be able to evaluate the presence or absence of many plant species in order to assess these variables.

In nature, plant leaves are two dimensional containing important features that can be useful for classification of various plant species, such as shapes, colours, textures and structures of their leaf, bark, flower, seedling and morph. According to Bhardwaj et al. [8], if the plant classification is based on only two dimensional images, it is very difficult to study the shapes of flowers, seedling and morph of plants because of their complex three dimensional structures.

The present paper proposes a comparison of the classification of different representation of plant leaves based on its margin, shape and textures; we used for each representation different classical supervised data mining algorithms. The organization of this paper is given as follows: Section 2 provides an overview of the related works; Section 3 gives details about dataset used in our experiment, Section 4 presents used approaches, discussion of the results will show in Section 5, and finally Section 6 gives the overall conclusion and the scope for future research.

II. RELATED WORK

Recently, plant classification became one of major researches. Shanwen et al. [2] used a combination between semi-Supervised locally linear embedding (semi-SLLE) and KNN algorithms for plant classification based on leaf images and showed its performance. James Cope et al. [6] presented plant texture classification using Gabor co-occurrences; where joint distributions for the responses from applying different scales of the Gabor filter are calculated. The difference among leaf textures is calculated by the Jeffrey divergence measure of corresponding distributions. Also Kadir et al. in [3] incorporates shape and vein, colour, and texture features to classify leaves using probabilistic neural network and proves that it gives better result with average accuracy of 93.75%. Plant leaf images corresponding to three plant types, are analysed using two different shape modelling techniques in Chaki et al. [5], authors proposed an automated system for recognizing plant species based on leaf images. One of the last works released by Bhardwaj in [8], that presented a simple computational method in computer vision to recognize plant leaves and to classify it using K-nearest neighbours. Anang Hudaya also worked on plant classification in his paper [18], presenting a scalable approach for classifying plant leaves using the 2-dimensional shape feature, using distributed hierarchical graph neuron (DHGN) for pattern recognition and k-nearest neighbours (k-NN) for pattern classification.

Copyright (c) IARIA, 2015. ISBN: 978-1-61208-445-9

III. DATASET

The 'Leaves' dataset contains one-hundred species of leaves [7], each species represented by three 64 element vector for each of three distinct features extracted from images: a fine-scale margin feature histogram, a Centroid Contour Distance Curve shape signature, and an interior texture feature histogram. This dataset contains 1600 samples, whereas there are sixteen distinct specimens for each species, photographed as a colour image on a white background. Figure 1 shows the first 27 species from the dataset.

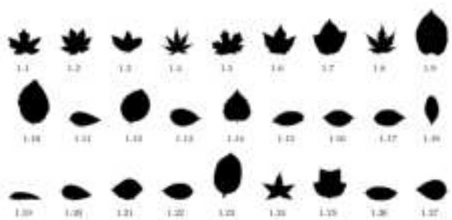

Figure 1. A silhouette image of one plant specimen each from the challenging one-hundred species leaves data set.

The data set inherently consists of having a wide set of classes with a low number of samples. Additionally, many sub species resemble the appearance of other major species, as well as many sub species with a major species can resemble a radically different appearance [7].

IV. PROPOSED APPROACHES

The present work shows a comparison of classification of seven different representations of plant leaves using three features extracted from the images; Figure 2 shows the architecture of proposed approaches:

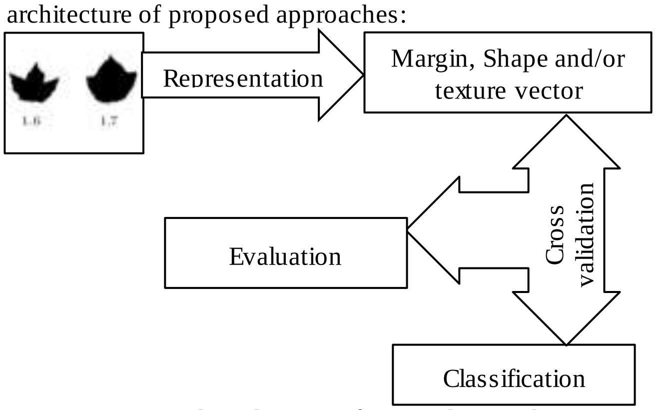

Figure 2. The architecture of proposed approaches

A. Representation of samples

Beginning with representation of species by three features extracted from images: a fine-scale margin feature histogram, then a Centroid Contour Distance Curve shape signature, and finally an interior texture feature histogram. We put values of each feature in 64 elements vector, then we tried to combine these vectors two by two in one 128 elements vector, and finally we presented species combining the three vectors together in one 192 elements vector.

B. Classification

In each case, we used three different approaches for classification: probabilistic approach using Naïve Bayes algorithm, hierarchical approach using Decision Tree C4.5 algorithm, and finally, approach based on distance calculation using K-nearest neighbours (K-NN) algorithm with k = 3, 4, 5, 6, or 7 and using Euclidian distance.

C. Evaluation

To evaluate classification, we used 10-folds cross validation. Training and testing sets are performed 10 times by partitioning randomly the dataset into 10 mutually in iteration exclusive subsets or "folds"; i, a subset Di is reserved as the test set, and the remaining partitions are collectively used to train the model.

D. Calculated measures for the evaluation

To calculate different metrics used for evaluation of classification, we have to introduce other measures:

1) True Positive (TP) present the average of the vectors that are correctly predicted relevant obtained in each iteration

2) True Negative (TN) present the average of the vectors that are correctly predicted as not relevant obtained in each iteration

3) False Positive (FP) present the average of the vectors that are predicted relevant but they are not relevant obtained in each iteration

4) False Negative (FN) present the average of the vectors that are correctly predicted not relevant but they are relevant obtained in each iteration

Using these four measures, we calculated the most famous measures that are used to evaluate classification algorithms:

1) *For classification, the accuracy estimate is the is the overall number of correct classifications from the 10 iterations, divided by the total number of tuples in the initial data.[16]*

- Accuracy = (TP + TN) / (TP + FP + TN + FN)

2) Precision and recall are the measures used in the information retrieval domain to measure how well an information retrieval system retrieves the relevant elements requested by a user. The measures are defined as follows[17]

- Precision = (TP) / (TP + FP)
- Recall = (TP) / (TP + FN).

3) Instead of two measures, they are often combined to provide a single measure of retrieval performance called the F-measure as follows[17]

- F-measure = (2 * Recall * Precision) /(Recall + Precision)

Copyright (c) IARIA, 2015.

V. RESULTS AND DISCUSSION

The following section shows the different results obtained for each representation with each algorithm.

TABLE I. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE MARGIN EXTRACTED FROM IMAGES

Algorithms	Evaluation %			
	Accuracy	**Precision**	**Recall**	**Fmeasure**
Naïve Bayes	**85.125**	**85.9**	**85.1**	**85.5**
Decision Tree	47.437	48.2	47.4	47.7
K-NN k=3	75.5	77.1	75.5	76.2
K-NN k=4	76.5	77.9	76.5	77.2
K-NN k=5	77.062	78.3	77.1	77.7
K-NN k=6	75.75	77.3	75.8	76.5
K-NN k=7	77.312	78.4	77.3	77.8

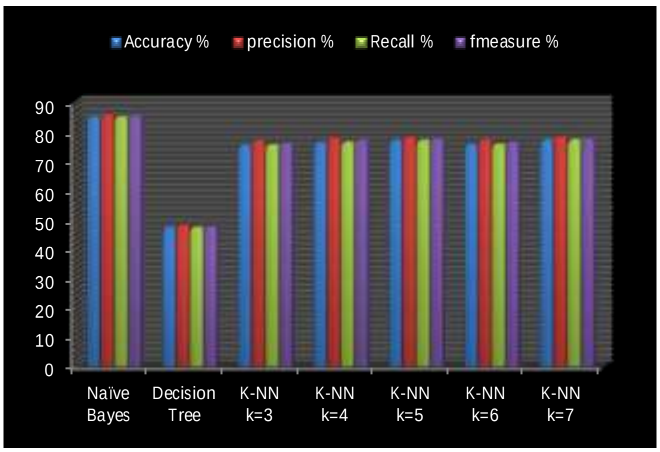

Figure 3. Results obtained by classification of species represented by the margin extracted from images

Table 1 and Figure 3 show the obtained results of the classification of species represented by the margins extracted from images where samples are 64 elements vectors. As we see, the probabilistic approach using Naïve Bayes gives best result compared with the approach based on distance calculation using K-Nearest Neighbours where the initial K (=3, 4, 5, 6, or 7) value can affect the result even a little, otherwise, hierarchical classification approach using Decision Tree gives the worst results.

TABLE II. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE SHAPE OF LEAVES

Algorithms	Evaluation %			
	Accuracy	**Precision**	**Recall**	**Fmeasure**
Naïve Bayes	52.625	53.8	52.6	53.2
Decision Tree	42.125	42	42.1	42
K-NN k=3	60.437	61.9	60.4	61.1
K-NN k=4	**61.187**	**62.5**	**61.2**	**61.8**
K-NN k=5	59	60.9	59	59.9
K-NN k=6	57.562	59.6	57.6	58.6
K-NN k=7	56.937	58.4	56.9	57.6

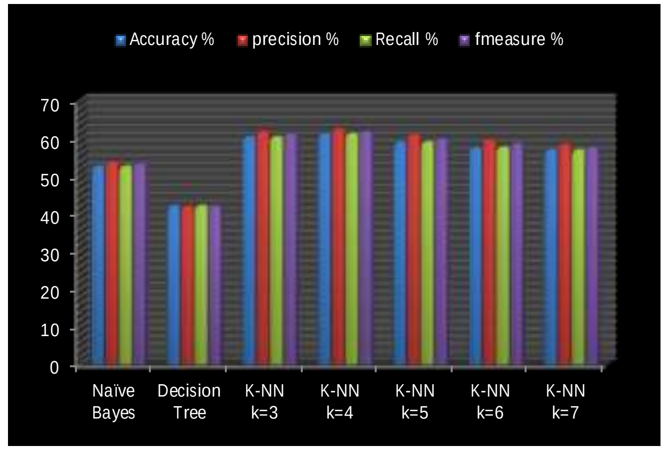

Figure 4. Results obtained by classification of species represented by the shape of leaves

In Table 2, unlike the margin representation, representation of leaves by its shape gives totally different results, in which approach based on distance calculation gives better result than probabilistic approach; even initial K value effect is more visible as we see in Figure 4.

TABLE III. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE TEXTURE EXTRACTED FROM IMAGES

Algorithms	Evaluation %			
	Accuracy	**Precision**	**Recall**	**Fmeasure**
Naïve Bayes	74.359	77.5	74.4	75.9
Decision Tree	51.97	52.9	52	52.4
K-NN k=3	76.923	78.2	76.9	77.5
K-NN k=4	76.673	78.1	76.7	77.3
K-NN k=5	**76.923**	**78.4**	**76.9**	**77.6**
K-NN k=6	76.548	78	76.5	77.2
K-NN k=7	76.86	78.4	76.9	77.6

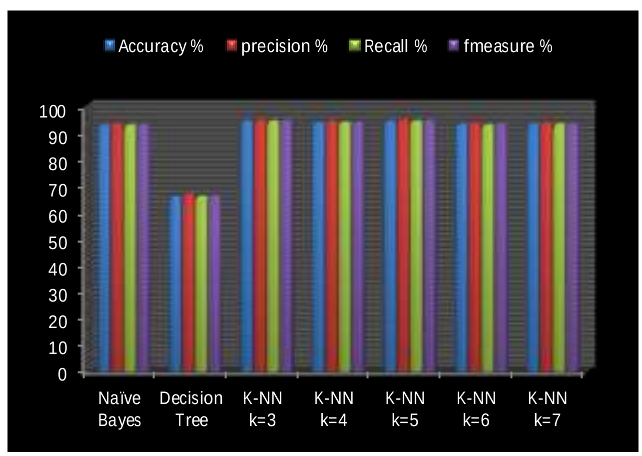

Figure 5. Results obtained by classification of species represented by the Texture extracted from images

In Figure 5 and Table 3, representation by texture vectors extracted from images gives convergent results, either between K-nearest neighbours with different initial K value, or between K-nearest neighbour and Naïve Bayes algorithm.

For hierarchical algorithm, it gives better results than the representation of leaves by margins or shape.

In order to optimize obtained results, we used to combine these features, where we get more efficiency in classification; the following Tables and Figures prove this idea. So, the supposed question here is, which combination can give the best result?

Beginning with combination between margin vector and shape vector in one 128 elements vector to represent each leaf, the obtained results are shown in Table 4.

TABLE IV. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE COMBINATION OF MARGIN AND SHAPE

Algorithms	Evaluation %			
	Accuracy	Precision	Recall	Fmeasure
Naïve Bayes	93.187	93.7	93.2	93.4
Decision Tree	66.187	67.1	66.2	66.6
K-NN k=3	**94.687**	**95.1**	**94.7**	**94.9**
K-NN k=4	94.187	94.7	94.2	94.4
K-NN k=5	94.687	95.2	94.7	95
K-NN k=6	93.5	94.2	93.2	93.8
K-NN k=7	93.562	94.2	93.6	93.8

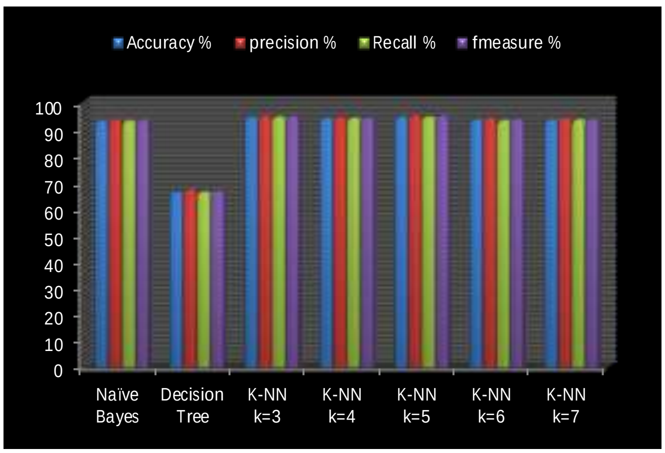

Figure 6. Results obtained by classification of species represented by the combination of margin and shape

In Figure 6 and Table 4, results given by Naïve Bayes algorithm and K-Nearest Neighbour are converged, and better than previous results (+90% of accuracy); even Decision Tree algorithm gives better result (almost 67% of accuracy).

TABLE V. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE COMBINATION OF MARGIN AND TEXTURE

Algorithms	Evaluation %			
	Accuracy	Precision	Recall	Fmeasure
Naïve Bayes	86.687	88.4	86.7	87.5
Decision Tree	59.375	59.9	59.1	59.5
K-NN k=3	**92**	**92.5**	**92**	**92.2**
K-NN k=4	91.562	92.1	91.6	91.8
K-NN k=5	91.312	92	91.3	91.6
K-NN k=6	91.25	91.9	91.3	91.6
K-NN k=7	91	91.8	91	91.4

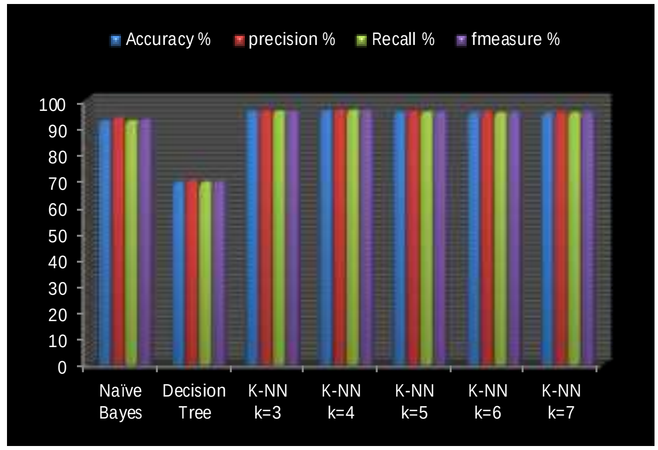

Figure 7. Results obtained by classification of species represented by the combination of margin and texture

Table 5 and Figure 7 show that obtained accuracy decreases compared with the combination between margin and shape, especially Naïve Bayes (from 93% to 87% of accuracy) and Decision Tree (from 66% to 59%).

TABLE VI. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE COMBINATION OF SHAPE AND TEXTURE

Algorithms	Evaluation %			
	Accuracy	Precision	Recall	Fmeasure
Naïve Bayes	84.25	86.5	84.3	85.3
Decision Tree	61.5	62.2	61.5	61.8
K-NN k=3	**87.687**	**88.6**	**87.7**	**88.1**
K-NN k=4	87.187	88	87.2	87.6
K-NN k=5	87	87.7	87	87.3
K-NN k=6	86.25	87.4	86.3	86.8
K-NN k=7	85.875	87	85.9	86.4

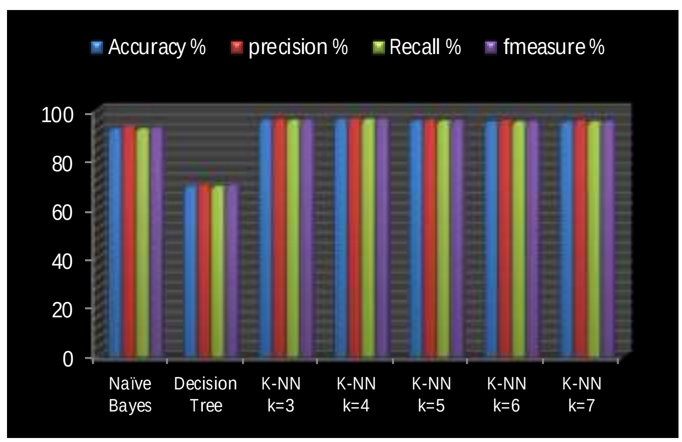

Figure 8. Results obtained by classification of species represented by the combination of margin and shape

In Table 6 and Figure 8, we see clearly that all algorithms had less performance (-90% of accuracy), where K-Nearest Neighbour gives the best result in this case.

We tried to improve results shown in Table 4 by combining the three features in one 192 elements vector in

Copyright (c) IARIA, 2015.

order to represent each sample, and we got better performance as demonstrated in Table 7 and Figure 9:

TABLE VII. RESULTS OBTAINED BY CLASSIFICATION OF SPECIES REPRESENTED BY THE COMBINATION OF THE THREE FEATURES

Algorithms	Evaluation %			
	Accuracy	Precision	Recall	Fmeasure
Naïve Bayes	92.437	93.5	92.4	92.9
Decision Tree	69.125	69.8	69.1	69.4
K-NN k=3	95.937	96.2	95.9	96
K-NN k=4	**96.25**	**96.5**	**96.3**	**96.4**
K-NN k=5	95.625	96	95.6	95.8
K-NN k=6	95.312	95.8	95.3	95.5
K-NN k=7	95.25	95.7	95.3	95.4

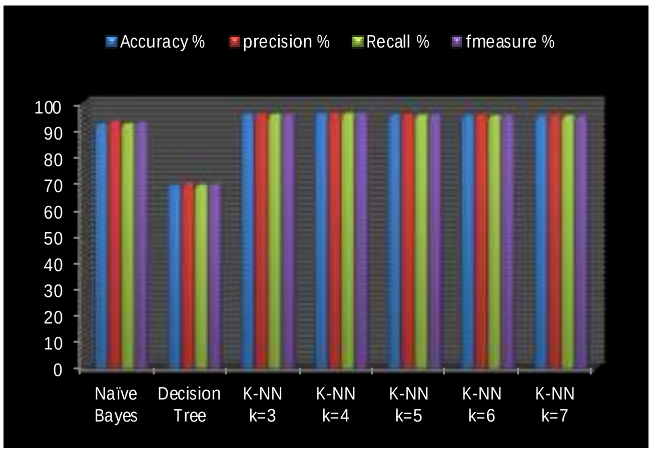

Figure 9. Results obtained by classification of species represented by the combination of the three features

Table 7 and Figure 9 prove that the combination of the three features extracted from images gives the best result in all tested representations; except Naïve Bayes that gives lower accuracy than the classification of vectors containing margin and shape values.

VI. CONCLUSION

Plants play an important role in our lives, without plants there will not be the existence of the ecology of the earth. The large amount of leaf types now makes the human being in a front of some problems in the specification of the use of plants, the first need to know the use of a plant is the identification of the plant leaf.

This work proposed a comparison of supervised classification of plant leaves, where we used to represent species in seven different representations, using three features extracted from binary masks of these leaves: a fine-scale margin feature histogram, by a Centroid Contour Distance Curve shape signature, and by an interior texture feature histogram. Results were very interesting in a way that gives as clear ideas:

- In term of representation: we can differentiate leaves by its margin better than shape or texture, but, experiments shown in this study prove our idea: the more we combine these features, the more precise the difference between samples is and that is what gives better results in classification.
- In term of classification: distance based algorithms give the best result for plant leaves classification. So, we can conclude that these algorithms are the most suitable for that task. On the other hand, the approach based on decision tree gives the worst results because of the overfitting problem. In general, a learning algorithm is said to overfit relative to a simpler one if it is more accurate in fitting known data but less accurate in predicting new data.

Use of the three features proved that there is some information more important than other. We discovered that margin representation can affect results more than the shape of the leaf. However, the combination of the three features gives the best result. To solve this problem, we plan, as future work, use of feature extraction algorithms, like PSO, to clean dataset and keep the important information in order to optimize the obtained results and avoid overfitting problem posed by decision tree algorithm. We plan also to use bio-inspired algorithms. They are part of a new research domain that is becoming more important due to its results in different areas.

REFERENCES

[1] C. Mallah, J. Cope, and J. Orwell ,"Plant leaf classification using probabilistic integration of shape, texture and margin features," Signal Processing, Pattern Recognition and Applications, 2013

[2] S. Zhang, and K. W. Chau, "Dimension reduction using semi-supervised locally linear embedding for plant leaf classification," Emerging Intelligent Computing Technology and Applications, Springer Berlin Heidelberg, 2009, pp. 948-955

[3] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, "Leaf classification using shape, color, and texture features", arXiv preprint arXiv:1401.4447, 2013

[4] A. H. M. Amin, and A. I. Khan, "One-shot Classification of 2-D Leaf Shapes Using Distributed Hierarchical Graph Neuron (DHGN) Scheme with k-NN Classifier," Procedia Computer Science, 24, 2013, pp.84-96

[5] J. Chaki, and R. Parekh, "Plant leaf recognition using shape based features and neural network classifiers," International Journal of Advanced Computer Science and Applications (IJACSA), 2011, pp. 26-29

[6] J. S. Cope, P. Remagnino, S. Barman, and P. Wilkin, "Plant texture classification using gabor co-occurrences," In Advances in Visual Computing, Springer Berlin Heidelberg, 2010, pp.669-677

[7] C. Mallah, "Probabilistic Classification from a K-Nearest-Neighbour Classifier," Computational Research, 1(1), 2013, pp.1-9

[8] A. Bahrdwaj, M. Kaur, and A. Kumar, "Recognition of plants by Leaf Image using Moment Invariant and Texture Analysis," International Journal of Innovation and Applied Studies, 3(1), 2013, pp.237-248

[9] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," In Advances in neural information processing systems, 2005, pp.1473-1480

[10] A. R. Backes, W. N. Gonçalves, A. S. Martinez, and O. M. Bruno, "Texture analysis and classification using deterministic tourist walk," Pattern Recognition, 43(3), 2010, pp.685-694

[11] Q. P. Wang, J. X. Du, and C. M. Zhai, "Recognition of leaf image based on ring projection wavelet fractal feature," In Advanced Intelligent Computing Theories and Applications, with Aspects of Artificial Intelligence, Springer Berlin Heidelberg, 2010, pp.240-246

[12] T. Beghin, J. S. Cope, P. Remagnino, and S. Barman, "Shape and texture based plant leaf classification," In Advanced Concepts for Intelligent Vision Systems, Springer Berlin Heidelberg, January 2010, pp.345-353
[13] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, "A comparative experiment of several shape methods in recognizing plants," arXiv preprint arXiv:1110.1509, 2011
[14] N. Vallimmal, and S. N. Geethalakshmi, "Hybrid image segmentation algorithm for leaf recognition and characterization," In Process Automation, Control and Computing (PACC), 2011 International Conference on IEEE, 2011, pp.1-6
[15] K. Singh, I. Gupta, S. Gupta, "Svm-bdt pnn and fourier moment technique for classification of leaf shape," International Journal of Signal Processing, Image Processing and Pattern Recognition, 3(4), 2010, pp.67-78
[16] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and techniques, Morgan Kaufmann Publishers is an imprint of Elsevier, 2011
[17] C. Sammut, and G. I. Webb, "Encyclopedia of machine learning, " Springer Science & Business Media, 2011