AUTOMATED CLASSIFICATION OF HAIR CARE PLANTS USING GEOMETRICAL AND TEXTURAL FEATURES FROM LEAF IMAGES: A PATTERN RECOGNITION BASED APPROACH

A. Shaukat, S. Farhan, M. Tahir, M. A. Fahiem and H. Tauseef

Department of Computer Science, Lahore College for Women University, Lahore, Pakistan
Corresponding author's email: saifar79@hotmail.com

ABSTRACT: Automated classification plays a vital role in content based image retrieval systems in addition to many more. Inter-class similarity and intra-class dissimilarity is the main challenge posed by leaf classification. This research work proposed a plant classification system using textural and geometrical features from leaf images. Six classification models, among which three were ensemble methods, were considered to evaluate the accuracy of proposed technique. Train and test strategy was adopted to evaluate the performance of different classifiers. Experimental results showed that the proposed technique outperformed the state of the art. Moreover, it was observed that textural features outperformed geometrical features. The best accuracy achieved with textural features was 100%, whereas it was 98.8% when geometrical features were used. SVM, IBk and Random Tree remained the best classifiers in leaf identification using both types of features. Hence, textural and geometrical features could be effectively used for plant classification.

INTRODUCTION

Uses of plants are crucial in every field of life such as foodstuff, medicine, industry and protection of environment (Rates, 2001; Farnsworth, 1988; Secoy and Smith, 1983). A number of plants species are beneficial for human health including hair care. The uses of plants increase the need of automated techniques for plants classification e.g. to diagnose diseased plant from healthy plants (Martinelli *et al.*, 2015), to identify the same species of plants (Wu *et al.*, 2007), and to recognize/classify a plant. Shape (Du *et al.*, 2007), color (Kadir *et al.*, 2013), veins (Larese *et al.*, 2014), texture (Sathwik *et al.*, 2013; Cope *et al.*, 2010), morphology (Nesaratnam and BalaMurugan, 2015; Wu *et al.*, 2007) and geometry (Kalyoncu and Toygar, 2015) are most commonly used features for automated plant classification. Some researchers have worked on combination of these features as well for plant leaf classification (Arafat *et al.*, 2016; VijayaLakshmi and Mohan, 2016).

Mostly used parts of plants for classification are leaf, bark or flower. Supervised and unsupervised classifiers are used for this purpose. Several classifiers like Probabilistic Neural Network (PNN) (Kadir *et al.*, 2013), Support Vector Machine (SVM) (Tomar and Agarwal, 2016; Chang and Lin, 2011), Move Median Centers (MMC) (Du *et al.*, 2007), Probability Distribution Functions (PDFS) (Cope *et al.*, 2010), Linear Discriminant Classifier (LDC) (Kalyoncu and

Toygar, 2015), Penalized Discriminant Analysis (PDA) (Larese *et al.*, 2014), Neural Network (NN) (Fu *et al.*, 2004), Linear Discriminant Model (LDM) (Neto *et al.*, 2006), Random Forest (RF) (Larese *et al.*, 2012), Neuro-Fuzzy Controller (NFC) and Multi-layered Perceptron (MLP) (Chaki *et al.*, 2015) are available for use in plant classification. These classifiers work well with different sets of features. Mostly existing techniques are good enough and provide acceptable accuracy levels, but a large feature set is limitation. Consequently, in the present study, a recognition system for plant classification/recognition is proposed which is based on the geometrical and textural features of the leaf. Geometrical features are easy to calculate and robust. Moreover, geometrical features can capture the shape of a leaf easily. Textural features, on the other hand, capture the texture of leaves and are one of the main features for image analysis. The reason behind selection of leaves as an object for plant classification is that leaves remain on plant for most part of the year. The proposed recognition system is capable to classify hair care plants from other species of plants. A number of classifiers were used here in order to understand the appropriateness of use of textural and geometrical features for plant classification.

MATERIALS AND METHODS

The dataset selected belonged to vascular plants; a sub-category of plant kingdom. Images in the dataset were self-captured with 8MP digital camera. Total leaf

images (n=1250) were considered each belonging to one of 12 plant classes (Table 1). It consisted of botanical and common name of a plant and its family name. The total number of leaf images considered against each class were mentioned.

The proposed research work comprised of three stages ie preprocessing, feature extraction and classification (Figure- 1). Preprocessing stage prepared the images for feature extraction and further analysis. It included grayscale conversion, noise removal and binarization. Textural features (n=314) and geometrical features (n=12) were extracted during feature extraction stage, which were used as feature sets for classification. Classifiers used in the proposed research work were SVM, IBk, J48, Random Forest, Random Tree and Bagging.

Table: 1. Description of Dataset used in this Research Work.

| Sr. No.	Botanical Name	Common Name	Family	No. of Images
1	*Aloe barbadensis*	Barbados aloe, Aloe vera, Lily of the desert, Curacao aloe,	Liliaceae	120
2	*Trigonella foenum-graecum*	Cooper's clover, Fenugreek, Fresh menthe, Bird's foot, Greek clover, Sicklefruit, Greek hay	Fabaceae	100
3	*Rosmarinus officinalis*	Compass plant, Polar plant, Compass-weed	Lamiaceae	130
4	*Ginkgo biloba*	Maidenhair tree	Ginkgoaceae	80
5	*Nepeta cataria*	Catnip, Catnep, Catmint, Catrup, Catswort	Lamiaceae	110
6	*Simmondsia chinensis*	Jojoba, Goat nut, Coffeeberry	Simmondsia	90
7	*Mentha xpiperita*	Peppermint	Lamiaceae	100
8	*Lavandula angustifolia mill*	English Lavender	Lamiaceae	110
9	*Arctium xmixtum Nyman*	Burdock	Asteraceae	100
10	*Moringa oleifera*	Ben oil tree/Benzoil tree, Drumstick tree, Moringa, Horseradish tree	Moringaceae	120
11	*Cocos nucifera*	Coconut palm	Arecaceae	100
12	*Centella asiatica*	Spadeleaf	Apiaceae	90
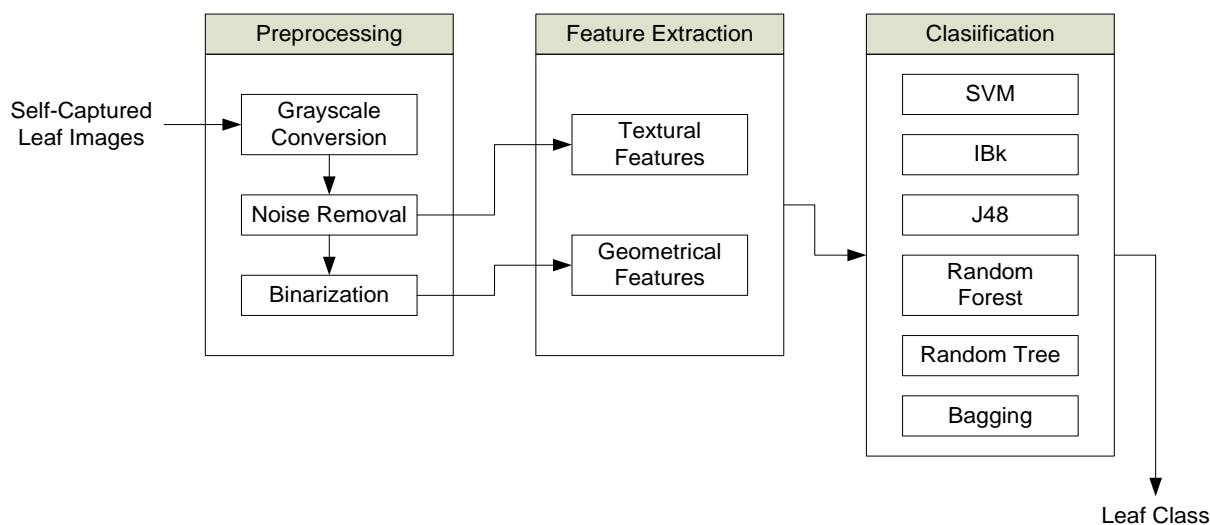
Fig-1: Block Diagram of the Proposed Work

The most common color of plant leaves was green but their shades kept on varying with time due to changes in atmosphere, water and nutrients. For automated identification of leaf class, the RGB leaf images were converted to grayscale. Another artifact that degraded classification accuracy was the presence of noise in images. Median filtering technique by (Chan *et al.*, 2005) was adopted in this research work for noise removal. The preprocessing algorithms were necessary to be applied for later extraction of features from leaf images. Grayscaled and noise removed images were used to extract textural features from images. For geometrical features extraction binarization was also performed in addition to previous preprocessing steps. Noise Removed

images were converted to binary to find the exact leaf shape and to extract other geometrical features. Fixed threshold value (Otsu, 1975) equal to 0.2 was used for binarization. A sample preprocessed image is shown in Figure- 2.
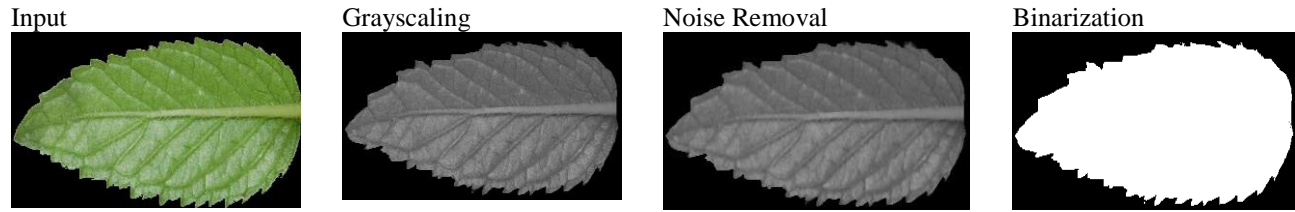
Input   Grayscaling   Noise Removal   Binarization

Fig-2: Sample Preprocessed Leaf Image

Textural features were extracted from grayscaled, noise removed images whereas geometrical features were extracted from binarized images. These features were helpful in the analysis and identification of leaves.

Textural features (Fernandez-Lozano *et al.*, 2015) consisted of numerical values that represented the texture of leaf surface. MaZda, a freely available software tool by (Szczypiński *et al.*, 2009), was used for this purpose. The features were divided into six feature sets according to the type of features. The feature sets constructed were Autoregressive features (ARF), Co-occurrence Matrix Features (COMF), Gradient Features (GF), Histogram Features (HF), Run Length Matrix Features (RLMF) and Wavelet features (WF). In addition to these feature sets, a sum of all textural features (ALLF) was also considered.

ARF calculated the sum of weight pixel intensities of neighboring pixels (Materka and Strzelecki, 1998). COMF examined the relationship between pixels over a selected image area (Pharsook *et al.*, 2011). GF in an image was the force at every point, giving the direction of the biggest possible increment from light to dim and the rate of change in that direction (Raju *et al.*, 2014 and Drzewiecki *et al.*, 2013). For HF, individual pixel intensities were counted over an image area and were used to calculate the gray level histogram (Materka and Strzelecki, 1998). RLMF counted the progressive running of white pixels from rows and columns (Raju *et al.*, 2014; Selvarajah and Kodituwakku, 2011and Galloway, 1975). WF was used to retrieve information from the image about different regions. It divided the image into different regions that implied high or low grey level variation (Selvarajah and Kodituwakku, 2011).

The binarized leaf shape images were used to extract geometrical features (n=12). The features included Perimeter, Area, Major axis length, Minor axis length, Filled area, Orientation, Equiv diameter, Euler number, Solidity, Extent, Eccentricity and Convex Area (Russ, 1999). Mathematical models of the features are presented in Table 2.

Table 2: Geometrical Features used in Proposed Research Work.

S. No	Feature Name	Mathematical Model
1	Perimeter (P)	$P = 2 \times (width + length)$
2	Area (A)	$A = (Actual\ Number\ of\ Pixels\ in\ an\ image)$
3	Major Axis Length (Maj_AL)	$Maj\_AL = d_1 + d_2$
4	Minor Axis Length (Min_AL)	$Min\_AL = \sqrt{(d_1 + d_2)^2 - f^2}$
5	Filled Area (FA)	$FA = \sum\limits_{m,n}^{N-1} f(m,n)$
6	Orientation (O)	$O = \tan^{-1}(\dfrac{f - d_1 + \sqrt{(f - d_1)^2 + (d_2)^2}}{d_2}$
7	Equiv Diameter (ED)	$ED = \sqrt{\dfrac{4 \times A}{\pi}}$

S. No	Feature Name	Mathematical Model
8	Euler Number (EN)	$EN = Number\ of\ objects - Number\ of\ holes\ in\ object$
9	Solodity (s)	$S = \dfrac{A}{CA}$
10	Extent (Ex)	$Ex = \dfrac{A}{Area\ of\ Bounding\ Box}$
11	Eccentricity (Ec)	$Ec = \dfrac{f}{Maj\_AL}$
12	Convex Area (CA)	$CA = smallest\ portion\ of\ polygon\ that\ contains\ region$

For the classification of data, a number of machine learning algorithms were available. The classifiers used in proposed work were SVM, IBk, J48 and 3 ensemble classifiers Random Forest, Random Tree and Bagging.

SVM, a type of linear classifier, tried to find a hyperplane that separated two classes of data (Burges, 1998). K-Nearest Neighbor (IBk) estimated the class probabilities with a Laplace assessment of the extent of the neighbors in every class. Here value of K=1 was used for all images (Aha *et al.*, 1991). Decision Tree (J48) was a predictive machine learning model which created a binary tree for classification (Witten and Frank, 2005). Random Forests worked by merging several weak tree classifiers. The final output was the majority votes received from weak classifiers (Cutler *et al.*, 2012; Breiman, 2001). Random Tree was basically a decision tree that was built on random subset of attributes at each node. Bagging was used for improving unstable estimation or classification for solving regression problems to get an aggregated predictor (Bauer and Kohavi, 1999). On the basis of datasets, different algorithms resulted in varying performance. Classification was performed using training and test set method (James, 1985).

RESULTS AND DISCUSSION

Results obtained using textural and geometrical features with different classifiers were reported and subsequently compared with existing research work. The performance parameters used included accuracy, Kappa statistics and Root Mean Square Error (RMSE).

Textural features were grouped into six feature sets (ARF, COMF, GF, HF and WF) as well as a sum of all textural features (ALLF). Each of six classification models were used with individual textural feature set and accuracies were calculated. After that sum of all textural features (ALLF) were used with all classifiers to calculate the accuracy. Highest accuracy of 100% was achieved using SVM, IBk and Random Tree, when using sum of all textural features (n=314). Whereas J48, Random Forest and bagging produced 91.20%, 98.00% and 88.00% accuracies respectively over sum of all textural features (ALLF). Table 3 should be referred for results (% accuracies) using individual and combine feature sets with six classifiers. Similarly Table 4 and Table 5 reported Kappa statistics and RMSE respectively.

It was evident that SVM, IBk and Random Tree outperformed J48, Random Forest and Bagging. Best performance was achieved using Random Tree over all feature sets whereas worst performance could be seen using Bagging.

For geometrical features, Accuracy, Kappa statistics and RMSE were calculated using all classifiers (Table 6). It was observed that SVM, IBk and Random Tree presented highest accuracy of 98.80%, whereas J48, Random Forest and Bagging produced 86.00%, 96.00% and 73.60% accuracies respectively.

Table 3: Classification Results using Textural Features Expressed as % Accuracy.

Classifier	Textural Feature Sets						ALLF
	ARF	COMF	GF	HF	RLMF	WF	
SVM	82.00	**100.00**	99.60	**100.00**	**100.00**	**100.00**	**100.00**
IBK	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.60	**100.00**
J48	85.20	91.20	90.00	85.60	87.60	85.60	91.20
Random Forest	99.20	97.60	97.60	97.20	97.20	99.20	98.00
Random Tree	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Bagging	84.00	86.40	80.80	84.80	78.40	82.00	88.00

Table 4: Kappa Statistics using Textural Features.

| Classifier | Textural Feature Sets | | | | | | ALLF |
	ARF	COMF	GF	HF	RLMF	WF	
SVM	0.80	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
IBK	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
J48	0.84	0.90	0.89	0.84	0.86	0.84	0.90
Random Forest	0.99	0.97	0.97	0.97	0.97	0.99	0.98
Random Tree	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Bagging	0.82	0.85	0.79	0.83	0.76	0.80	0.87

Table 5: RMSE using Textural Features.

| Classifier | Textural Feature Sets | | | | | | ALLF |
	ARF	COMF	GF	HF	RLMF	WF	
SVM	0.17	**0.00**	0.03	**0.00**	**0.00**	**0.00**	**0.00**
IBK	0.01	0.01	0.01	0.01	0.01	0.03	0.13
J48	0.13	0.10	0.10	0.13	0.12	0.13	0.10
Random Forest	0.11	0.10	0.11	0.11	0.11	0.11	0.10
Random Tree	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Bagging	0.19	0.17	0.18	0.18	0.19	0.20	0.18

Table 6: Classification Results using Geometrical Features Expressed as % Accuracy.

Classifier	Accuracy	Kappa Statistics	RMSE
SVM	98.80	0.99	0.04
IBk	98.80	0.99	0.03
J48	86.00	0.85	0.13
Random Forest	96.00	0.96	0.12
Random Tree	98.80	0.99	0.03
Bagging	73.60	0.71	0.19

Results achieved using the proposed research was compared with a number of well-known existing research works. Three important aspects were considered when comparisons were made, i.e. types of features, accuracy and classification model. It was found that the proposed approach was capable of producing better results with smaller features sets as compared to existing approaches.

When only textural features were taken into account the accuracy achieved was 100% for SVM, IBk and Random Tree. In a study carried out by Cope *et al.* (2010) and Sathwik *et al.* (2013) also used textural features for same purpose but accuracy remained 79.69 % using PDFS and 94% using texture analysis respectively. For better accuracy, Chaki *et al.* (2015) combined textural features and shape features with neural classifiers and got an accuracy of 97.6% using NFC and 85.6% using MLP. Here the types of features were increased but still accuracy remained below the accuracy achieved by the proposed approach using only textural features (100%) while using only geometrical features (98.8%). Similarly Kadir *et al.* (2013) added color features to textural features and shape features and achieved an accuracy of 93.75% when PNN was used, still below proposed approach accuracy.

In contrast to textural features, Du *et al.* (2007), Wu *et al.* (2007) and Nesaratnam and BalaMurugan (2015) used morphological features for plant species classification using MMC, PNN and SVM respectively. Their accuracies were 91%, 90% and 86.66% respectively, which were lower than the accuracy achieved by using proposed approach. It became obvious that the use of textural features and SVM, IBk or Random Tree as classifiers produced better results than existing works. It was observed that even selection of a subset of textural features (COMF, HF, RLMF) produced 100% accuracy with SVM, IBk and Random Tree. Use of multiple classifiers as well as ensemble classifier confirmed the accuracy achieved by the proposed work.

Conclusion: It was concluded that for leaf image classification using textural and geometrical features SVM, IBk and Random Tree performed better as compared to J48, Random Forest and Bagging. Moreover textural features resulted in better classification rates (100%) compared to geometrical features.

REFERENCES

Aha, D.W., D. Kibler and M.K. Albert (1991). Instance-based learning algorithms. Mach Learn, 6(1): 37-66.

Arafat, S.Y., M.I. Saghir, M. Ishtiaq and U. Bashir (2016). Comparison of techniques for leaf classification. 6th Int. Conf. IEEE DICTAP, 136-141.

Bauer, E. and R. Kohavi (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. Mach Learn, 36(1-2): 105-139.

Breiman, L. (2001). Random forests. Mach Learn, 45(1): 5-32.

Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov, 2(2): 121-167.

Chaki, J., R. Parekh and S. Bhattacharya (2015). Plant leaf recognition using texture and shape features with neural classifiers. Pattern Recog Lett, 58: 61-68.

Chan, R.H., H. Chung-Wa and M. Nikolova (2005). Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. IEEE Trans Image Process, 14(10): 1479-1485.

Chang, C.C. and C.J. Lin (2011). LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol, 2(3): 27.

Cope, J.S., P. Remagnino, S. Barman and P. Wilkin (2010). Plant texture classification using gabor co-occurrences. Int. Symposium VC, 669-677.

Cutler, A., D.R. Cutler and J.R. Stevens (2012). Random forests. Ensemble Machine Learning. 157-175.

Drzewiecki, W., A. Wawrzaszek, M. Krupinski, S. Aleksandrowicz and K. Bernat (2013). Comparison of selected textural features as global content-based descriptors of VHR satellite image - the EROS-a study. Int. Conf. FedCSIS, 43-49.

Du, J.X., X.F. Wang and G.J. Zhang (2007). Leaf shape based plant species recognition. Appl Math Comput, 185(2): 883-893.

Farnsworth, N.R. (1988). Screening plants for new medicines. Biodiversity, 3: 83-97.

Fernandez-Lozano, C., J. Seoane, M. Gestal, T. Gaunt, J. Dorado and C. Campbell (2015). Texture classification using feature selection and kernel-based techniques. Soft Comput, 19(9): 2469-2480.

Fu, H., Z. Chi, D. Feng and J. Song (2004). Machine learning techniques for ontology-based leaf classification. 8th Int. Conf. IEEE ICARCV, 681-686.

Galloway, M.M. (1975). Texture analysis using gray level run lengths. Comput Vision Graph, 4(2): 172-179.

James, M. (1985). Classification algorithms, Wiley-Interscience.

Kadir, A., L.E. Nugroho, A. Susanto and P.I. Santosa (2013). Leaf Classification Using Shape, Color, and Texture. arXiv preprint arXiv:1401.4447.

Kalyoncu, C. and O. Toygar (2015). Geometric leaf classification. Comput Vis Image Und, 133: 102-109.

Larese, M.G., R.M. Craviotto, M.R. Arango, C. Gallo and P.M. Granitto (2012). Legume identification by leaf vein images classification. 17th Int. Conf. CIARP, 447-454.

Larese, M.G., R. Namias, R.M. Craviotto, M.R. Arango, C. Gallo and P.M. Granitto (2014). Automatic classification of legumes using leaf vein image features. Pattern Recogn, 47(1): 158-168.

Martinelli, F., R. Scalenghe, S. Davino, S. Panno, G. Scuderi, P. Ruisi, P. Villa, D. Stroppiana, M. Boschetti, L.R. Goulart and C.E. Davis (2015). Advanced methods of plant disease detection. A review. Agron Sustain Dev, 35(1): 1-25.

Materka, A. and M. Strzelecki (1998). Texture analysis methods–a review. Technical university of lodz, institute of electronics, COST B11 report, Brussels: 9-11.

Nesaratnam, J. and C. BalaMurugan (2015). Identifying leaf in a natural image using morphological characters. Int. Conf. IEEE ICIIECS, 1-5.

Neto, J.C., G.E. Meyer, D.D. Jones and A.K. Samal (2006). Plant species identification using Elliptic Fourier leaf shape analysis. Comput Electron Agric, 50(2): 121-134.

Otsu, N. (1975). A threshold selection method from gray-level histograms. Automatica, 11(285-296): 23-27.

Pharsook, S., T. Kasetkasem, P. Larmsrichan, S. Siddhichai, T. Chanwimaluang and T. Isshiki (2011). The texture classification using the fusion of decisions from different texture classifiers. 8th Int. Conf. ECTI-CON, 1003-1006.

Raju, G., B.S. Moni and M.S. Nair (2014). A novel handwritten character recognition system using gradient based features and run length count. Sadhana, 39(6): 1333-1355.

Rates, S.M.K. (2001). Plants as source of drugs. Toxicon, 39(5): 603-613.

Russ, J.C. (1999). The image processing handbook (3rd ed.), CRC Press, Inc.

Sathwik, T., R. Yasaswini, R. Venkatesh and A. Gopal (2013). Classification of selected medicinal plant leaves using texture analysis. 4th Int. Conf. IEEE ICCCNT, 1-6.

Secoy, D.M. and A.E. Smith (1983). Use of plants in control of agricultural and domestic pests. Econ Bot, 37(1): 28-57.

Selvarajah, S. and S. Kodituwakku (2011). Analysis and comparison of texture features for content based image retrieval. IJLTC, 2(1).

Szczypiński, P.M., M. Strzelecki, A. Materka and A. Klepaczko (2009). MaZda—A software package for image texture analysis. Comput Methods Programs Biomed, 94(1): 66-76.

Tomar, D. and S. Agarwal (2016). Leaf Recognition for Plant Classification Using Direct Acyclic Graph Based Multi-Class Least Squares Twin Support Vector Machine. Int J Image Graph, 16(03): 1650012.

VijayaLakshmi, B. and V. Mohan (2016). Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture, shape, and color features. Comput Electron Agric, 125: 99-112.

Witten, I.H. and E. Frank (2005). Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann.

Wu, S.G., F.S. Bao, E.Y. Xu, Y.-X. Wang, Y.-F. Chang and Q.-L. Xiang (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. Int. Symposium IEEE ISSPIT, 11-16.