Classification of Indian Herbal Leaf with Random Forest Classifier

Shobnom Dasgupta[1], Trisa Das[2], DolaSaha[3],Chiranjib Dutta [4],Ananjan Maiti[5]

[1,2] Department of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal
[3]Assistant Professor, Dept. of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

[4]Assistant Professor, Dept. of Computer Applications
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

[5]Assistant Professor, Dept. of Computer Science and Engineering
Guru Nanak Institute of Technology
157/F Nilgunj Road, Panihati, Kolkata – 700114, West Bengal

Corresponding author's email: shobnomdasgupta56@gmail.com

Abstract – Herbal Leaf is normally used for preparing medicines. This includes images taken from the locality of West Bengal mainly Kolkata, with white background. Consideringthis problem of image classification, this research is trying to identify herbal leaves based on the images which are in white background. After that, in the research, an image processing algorithm and Otsu segmentation are used. Then, the Herbal leaf image features are classified, using their shape and texture properties. Shape and colour features of Herbal leaf assemble high accuracy only when at a time both are used. Morphological features were applied for shape feature uproot in the research. The contribution of this research is applying the herbal leaf image enhancement and segmentation algorithm so that the image features can be uprooted based on shape and colour attributes. The accuracy of the classification in the research has got 97.8% with Random Forest Classifier.

Keywords – Image Processing, Shape Texture, Texture Feature, Segmentation, Feature Extraction.

1.0 Introduction

India is not only one of the most biologically but also culturally diverse countries in the world. Also, its per capita income is low. On an enormous scale, mass poverty's existence depends on mainly this fact: the country, India exports natural products which can be found in our locality very easily and the leaves are very useful to cure many health issues, as well as Herbal plants that are mostly

used to expand industries quickly. The original database which is made by us was used in the study. The herbal leaf images were collected by capturing in our locality. Much previous research collected their images which are captured by high-resolution cameras with natural backgrounds for easily segmented. In the study, these images had been captured by a low camera resulting from the smartphone being picked from its trunk and then put on a white background. Collecting the images with the natural background is a complex segmentation process so the method might be got a bigger risk. Each type of leaf has a different shape and colour features. Shape and colour features of Herbal leaf assemble high accuracy only when at a time both are used for classification. If only one feature either shape or colour is used, the accuracy becomes low. Using both features together to identify the Random Forest Classifier an accuracy of 97.8%. In this Research, morphological features are used for both shape and colour features.

2.0 Previous Study

This study also demonstrated that combining leaf shape and colour features yields the highest categorization accuracy.Additionally, they contrasted how accurately shape characteristics, colour features, and the use of both features were performed.

A. Leaf Segmentation

The majority of researchers rather use databases that are already available on the internet such as Flavia, ICL, and the Swedish Leaf Dataset. They believe that simplifying the segmentation procedure is desirable. Few researchers employ photographs of leaves. Using max-flow min-cut (MFMC), Kebapci et al. performed leaf segmentation in their natural context (Kebapci,2011). Their categorization accuracy stood at 73%. There was another investigationthat attempted to identify a leaf by looking at it from several angles, recreating its three-dimensional (3D) position, and then classifying it.While performing segmentation, James Nesaratnam et al. converted the RGB image into an HSV image, applied thresholding, and then converted it back to RGB.With this complicated background, leaf segmentation using the Region Growing method is successful.Based on previous findings, we attempted to make use of our database by taking pictures of leaves in their natural settings rather than cutting them from the plants.

B. Feature Extraction

Some studies have achieved great accuracy when classifying leaves using shape and colour data (Chithra, P. L., & Janes, P. S. 2018).According to some prior research, leaf categorization findings based on the extraction of leaf shape and colour attributes had the highest accuracy.It made use of both a support vector machine and an intersection cortical model. They carried out a study in India using attributes extracted from texture, colour, and form to classify herbal plants.Plant classification using leaf shapes—A pattern counting approach Pattern Recognition.The accuracy reached 72.16% (Wang, Z, 2014). From another research, we found that they made the application

of herbal plant classification in India. A fresh approach to automatically classifying plant species that uses sparse representations of leaf tooth characteristics. This time, accuracy improved to 97.8% using solely the extraction of shape and texture data (Herdiyeni, 2012). We learned from another study that they integrated the Gray Co-occurrence Matrix and Gabor Filter to extract texture features.Compared to individual texture feature extractions, they had greater accuracy. They also take into account feature extraction based on leaf form. The accuracy percentage was 97.60% using a neural network classifier as the classification algorithm. With 97.5 per cent accuracy, this result was superior to their prior study's use of solely leaf form for feature extraction (Chaki, J, 2015).

3.0 Methods

The following image tends to explain the method of image processing. It starts with image data retrieval, leaf segmentation, and feature extraction, then continued with classification.

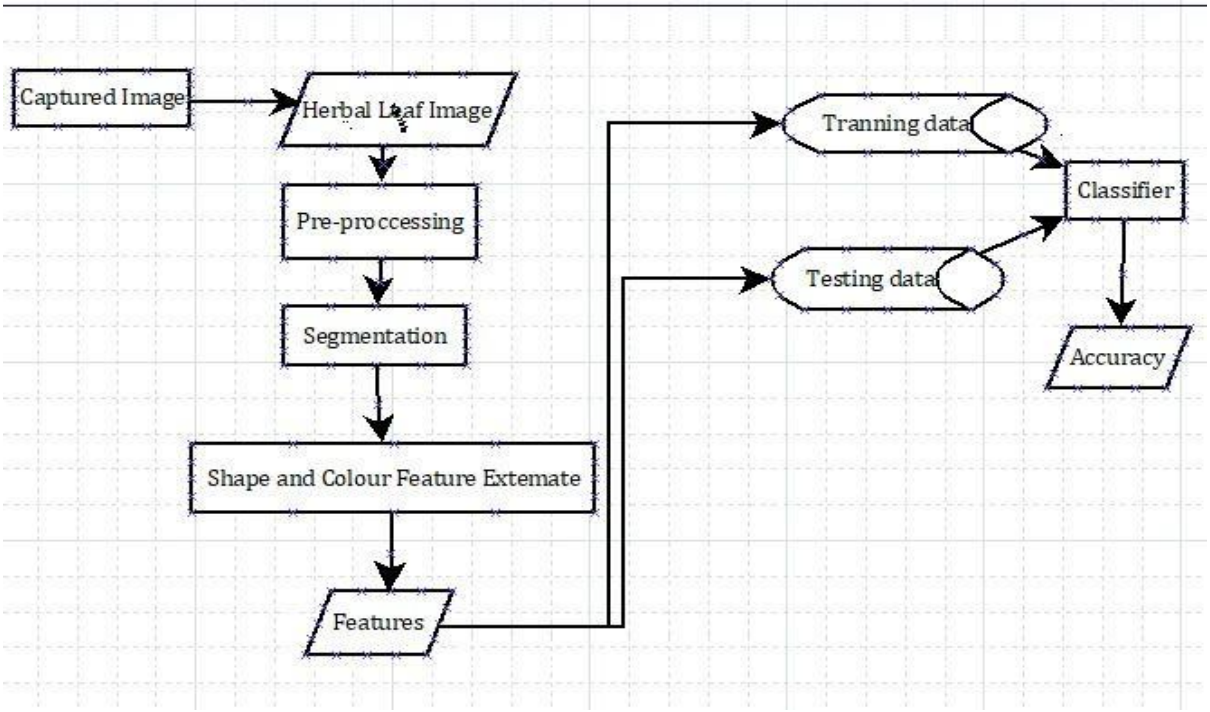

Fig 1. Image processing procedure

A. Preprocessing

The pre-processing stage is performed to standardize the image before the feature extraction is performed. In figure 2, the pre-processing stage begins by converting the RGB image to a LAB image. Then do the threshold process. After that, turn it back into an RGB image. Once it became an RGB image, it was converted into a grayscale image to be processed into a binary image using the Otsu algorithm. From the pre-processing stage performed (Chaki, J, 2015). The image used for colour feature extraction is a grayscale image and a binary image for shape feature extraction. Then rotate the major axis of the leaf. Lastly, shrunk the background using a bounding box.

B. Segmentation

In computer vision and image processing, Otsu's method, named after Nobuyuki Otsu is used to perform automatic images. In the simplest form, the algorithm returns a single intensity threshold that separates pixels into two classes, foreground and background. This threshold is determined by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance.

$$\sigma_{\omega}^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)....(1)$$

Weights ω_0 and ω_1 are the probabilities of the two classes separated by a threshold t, and σ_0^2 and σ_1^2 are variances of these two classes.

C. Morphological Features

This study uses nine morphological features based on research conducted (Chaki, J, 2015). Then searched for length (A), width (B), area (C), and perimeter of the leaf (X). Then formulate five additional parameters, namely P, Q, R, S, and T as below.

$$P = A/B \quad (2)$$
$$Q = 4\Lambda\ C/X2 \quad (3)$$
$$R = A*B/C \quad (4)$$
$$S = X/AC \quad (5)$$
$$T = X/(A+B) \quad (6)$$
$$\text{Feature Vector} = \{A,B,C,X,P,Q,R,S.T\} \quad (7)$$

D. Colour Features

In this study, Three colour features were used based on a previous study which is researched (Chaki, J, 2015). Then searched for the red channel, blue channel, green channel and perimeter of the leaf.

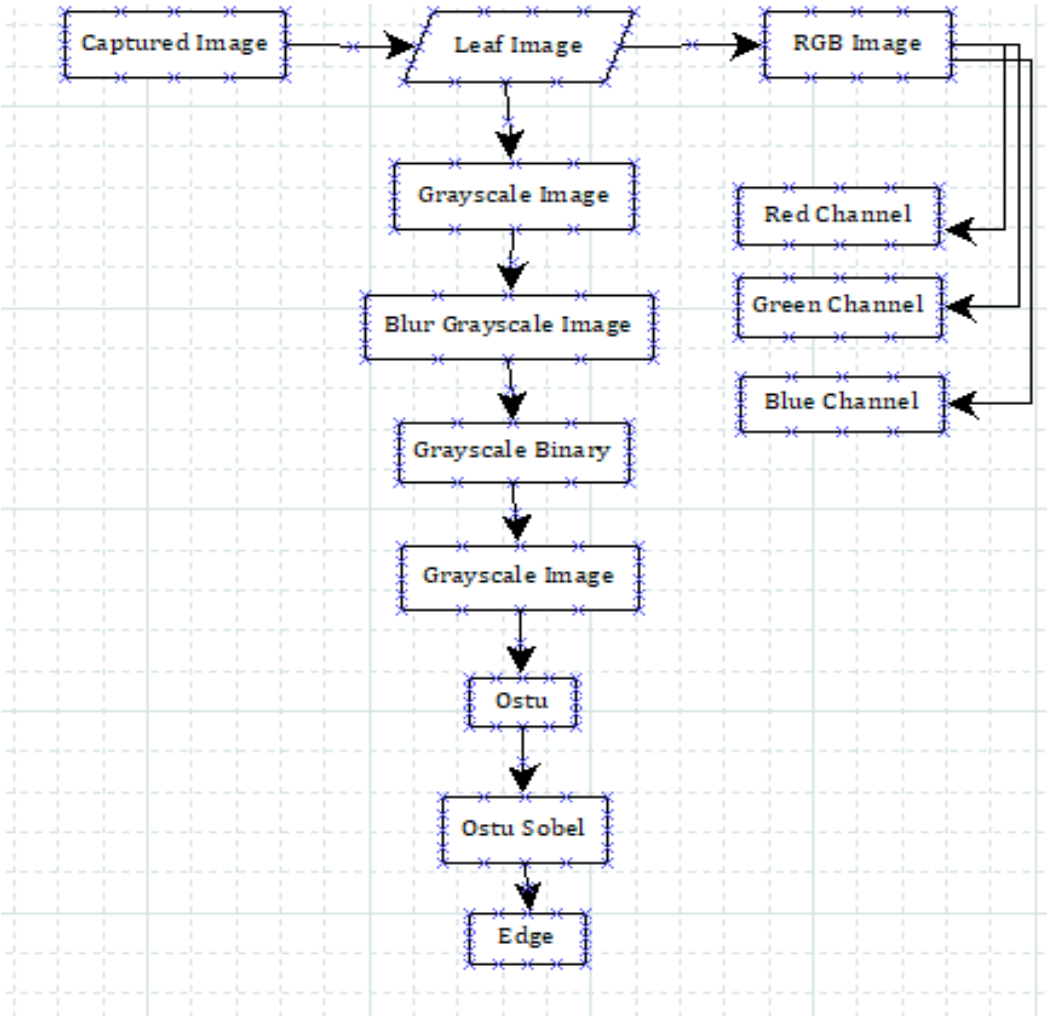

Fig 2. Preprocessing procedure

E. Classification

In this research supervised learning is also included. In this leaf Classification, the Ostu algorithm with Random Forest Classifier is used. The reason for using this algorithm is to motivate 97.8% accuracy.

4.0 Results

We took pictures with the poor smartphone camera. The amount of light present is not taken into consideration while retrieving images, and the best-looking and most colourful leaves are not chosen. Included are even a few juveniles.

A. Dataset

The photograph of a leaf below was captured with a smartphone's low-resolution camera. For each type of plant, 40 leaf photos were taken. For training data, we used 368 photos, and for testing data, 32 images. 462 RGB pictures altogether made up the leaf images. Each photograph had an average resolution of 750 by 600 pixels.

Fig 3a. Neem Leaf images

Fig 3b. Pudina Leaf images

Fig 3c. Marigold Leaf images

B. Pre-processing

As seen in figure 4, the RGB image was first converted into a grayscale image. We converted the grayscale image into a blurred grayscale image from the original. After that, we used the Otsu technique to assign the threshold value. Because we convert the grayscale image to a binary image, the threshold was required. We multiplied the binary image by the RGB image to isolate the leaf portion of the image. After that, we bound the box and shrank the image to make it translation invariant.

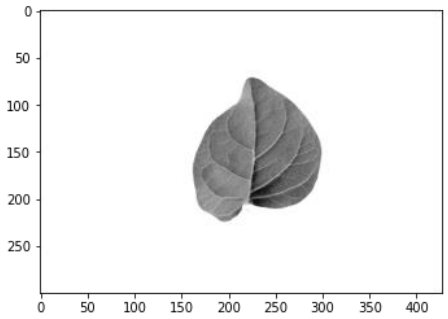

Fig 4 Ashwagandha. Grayscale image

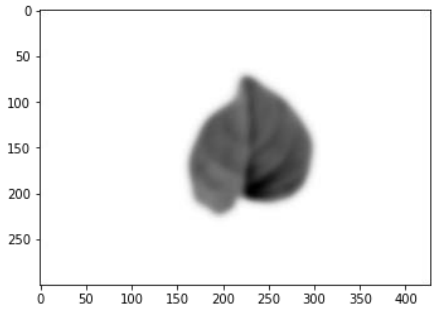

Fig 5. AshwagandhaBlur Grayscale image

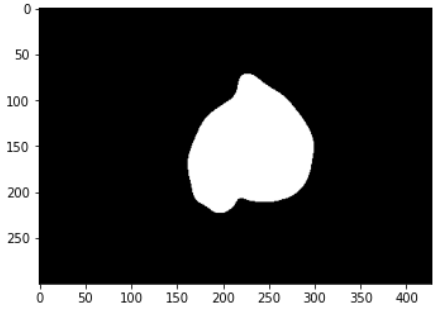

Fig 6. Ashwagandha Otsu algorithm image

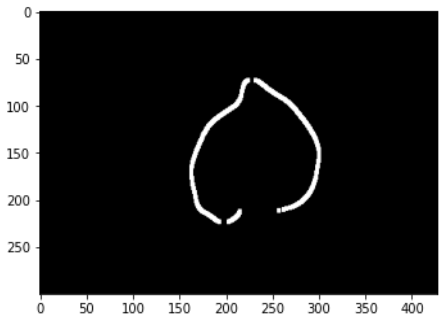

Fig 7. Ashwagandha Edge Detection image

C. Shape-Based Modelling

The extraction of shape features from binary pictures was done utilizing morphological characteristics. Hu moments and Zernike moments have been compared to the outcomes.

TABLE 1:Results of classification

Classifier	Classification	
	Accuracy	Log Loss
Random Forest Classifier	97.8495%	0.2424877488174202
Decision Tree Classifier	95.6989%	1.4855387696735864
Gradient Boosting Classifier	93.5484%	0.20097482888172627
Quadratic Discriminate Analysis	68.8172%	2.0099513206211355
Linear Discriminate Analysis	47.3118%	1.6168957194088576
neighbours	38.7097%	2.188134180159424
Gaussian NB	32.2581%	3.2014282467007016
Logistic Regression	31.1828%	2.0177543831393483
NVC	26.8817%	2.0318509971249266
AdaBoost Classifier	25.8065%	2.1096214917560885
SVC	23.6559%	2.117167312464825
Bernoulli NB	13.9785%	2.278998282207057

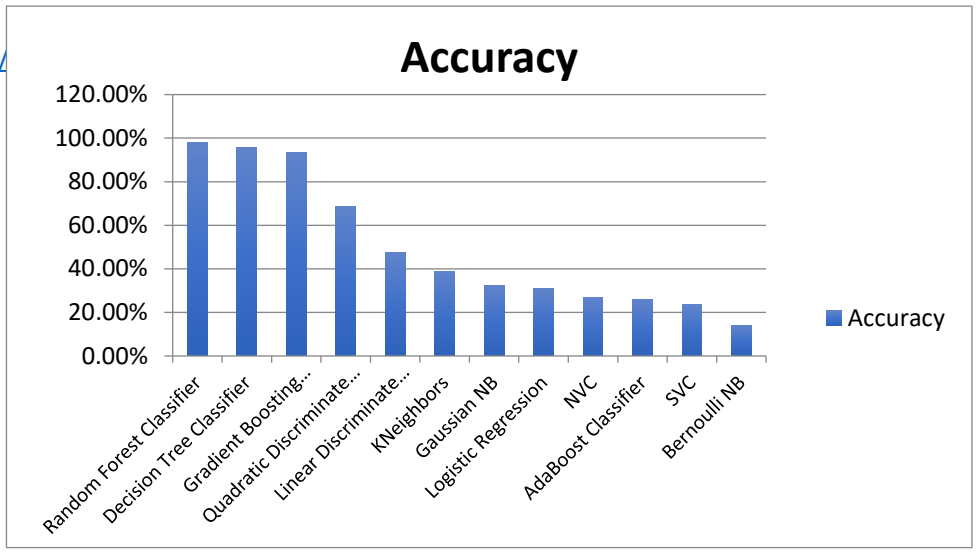

Fig 8. Classifier Accuracy

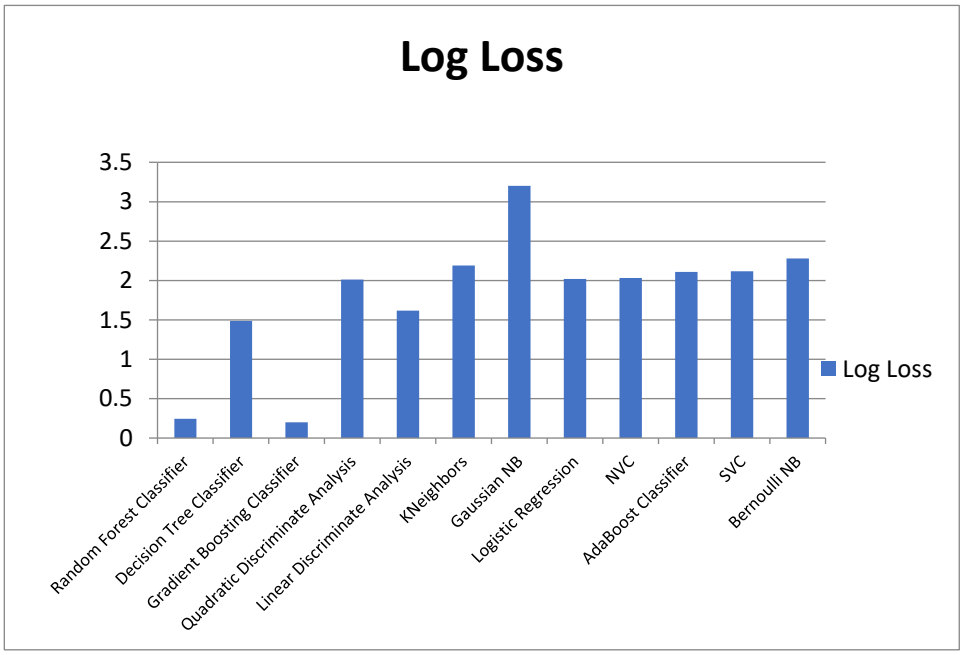

Fig 9. Classifier Log Loss

5.0 Discussion

It seems that sometimes, after segmenting the leaf image, the segmentation outcomes were not what was anticipated. As seen in figure 10, it occurred because the proportion of pixels that represented the leaf in the total image was less than 60%.

Fig 10. Leaf part below 60% of the overall image

It will require some segmentation algorithms if the findings of this research are applied at the feature matching stage. As a result, more than one algorithm will be employed to segment leaves in their natural context. The best outcomes from some of these segmentation techniques will be applied to feature matching. To determine which of these algorithms has features resembling leaf features, the output of some of these algorithms will be filtered. As a result, there will be three matching results if three segmentation algorithms are applied. The user then selects the method with the highest degree of accuracy. The characteristics of the leaves will help us identify plant illnesses.

6.0 Conclusion

Herbal leaf and its effect are vital for daily life for medical purposes. Herbal leaves produce various critical medicines. There are different sources of herbal leaves available in the locality of Kolkata. The study includes local Herbal leaves available in the locality of Kolkata, and it has been reproduced in the current image set. The image set has been processed with image processing techniques and segmentation. The color and the shape features of different types of Herbal leaf-like Tulsi contribute to the model. The research on this Herbal leaf classification gave an accuracy of 97.8% with the Random Forest Classifier. In the future, various methods and techniques will be applied to recognize more Herbal leaves. This model will be applied in the embedded system for further investigation and creation of herbal medicine.

6.0 Reference

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

Chithra, P. L., & Janes, P. S. (2018). A Survey on Various Leaf Classification Techniques for Medicinal Plants. In Proceedings of the International Conference on Advancements in Computing Technologies (pp. 38-42).

Herdiyeni, Y., &Santoni, M. M. (2012, December). Combination of morphological, local binary pattern variance and colour moments features for Indonesian medicinal plants identification. In 2012International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 255-259). IEEE.

Jin, T., Hou, X., Li, P., & Zhou, F. (2015). A novel method of automatic plant species classification using sparse representation of leaf tooth features. PloS one, 10(10), e0139482.

Kebapci, H., Yanikoglu, B., &Unal, G. (2011). Plant image retrieval using colour, shape and texture features. The Computer Journal, 54(9), 1475-1490.

Mareta, A., Soesanti, I., &Wahyunggoro, O. (2018, March). Herbal leaf classification using images in natural background. In 2018 International Conference on Information and Communications Technology (ICOIACT) (pp. 612-616). IEEE.

Teng, C. H., Kuo, Y. T., & Chen, Y. S. (2009, July). Leaf segmentation, its 3d position estimation and leaf classification from a few images with very close viewpoints. In International Conference Image Analysis and Recognition (pp. 937-946). Springer, Berlin, Heidelberg.

Wang, Z., Sun, X., Ma, Y., Zhang, H., Ma, Y., Xie, W., & Zhang, Y. (2014, July). Plant recognition based on intersecting cortical model. In 2014 International joint conference on neural networks(IJCNN) (pp. 975-980). IEEE.

Zhao, C., Chan, S. S., Cham, W. K., & Chu, L. M. (2015). Plant classification using leaf shapes—A pattern counting approach. Pattern Recognition, 48(10), 3203-3215.