Microscopic Image Analysis Method for Identification of Indian Herbal Plants

[1]Bhupendra Fataniya
Electronics and Communication Department
Nirma University
Ahmedabad, India
e-mail: bdfataniya@nirmauni.ac.in

[2]Prachi Manishkumar Patel
Electronics and Communication Department
Nirma University
Ahmedabad, India
e-mail: pmpatel4593@gmail.com

[3]Tanish Zaveri
Electronics and Communication Department
Nirma University
Ahmedabad, India
e-mail: ztanish@nirmauni.ac.in

[4]Sanjeev Acharya
Pharmacy Department
Nirma University
Ahmedabad, India
e-mail: sanjeev.acharya@nirmauni.ac.in

***Abstract*—Identification of medicinal plant from its powder is a common task performed in pharmaceutical labs and industries. In this paper, we are proposed a identification using machine intelligence. The algorithm is designed for the microscopic images of plant powder and segment the specific object from these images. From the extracted image unique features are identified and using this we train our algorithm and classify the test images. We have prepared the image database at our pharmacy lab for two plants namely Licorice and Kurchi plant. The proposed algorithm is applied on the variety of fifty images and results are described using confusion matrix and ROC curves.**

Keywords— K-Means, Microscopic Images, Segmentation

I. INTRODUCTION

Many pharmaceutical companies and industry have demand for identification of plant from its powder form. These plants have many medicinal purposes described later in this section. In today's scenario, this task being done manually and hence is extremely time-consuming, complex and monetarily inefficient. This paper focuses on development of automatic intelligent system for microscopic image analysis and identification system for Indian herbal plants. In this paper various microscopic images of two herbal plants namely licorice and kurchi are considered for the experiment and the analysis of unique characteristic features present in respective plant images described as follows.

Medicinal uses of each plant and characteristic features of microscopic images of the same are explained.

A. Licorice Powder

1. Characteristic feature: The main feature used for detection is blunt ended phloem fibers with calcium oxalate crystal [1]. The said fiber hidden under the cell array and are clearly visible when the image is converted to a binary image. The main properties used for detection are major axis length, minor axis length and area. The feature is effectively shown using circle in microscopic image of licorice in Fig. 1 (a).
2. Medicinal use: The compound glycyrrhizin, found in licorice, has been proposed as being useful for liver protection in tuberculosis. Glycyrrhizin has also demonstrated antiviral, antimicrobial, anti-inflammatory, hepatoprotective and blood-pressure increasing effects in vitro and in vivo, glycyrrhizin also slows the progression of viral and autoimmune hepatitis [2]. Liquorice has also demonstrated efficacy in treating hyperlipidaemia, inflammation-induced skin hyperpigmentation and in preventing neurodegenerative disorders and cavities [3].

(a) (b)

Figure 1 Microscopic image of (a) crystal phloem fiber of licorice (b) rhombic crystal of kurchi

B. Kurchi Powder

1. Characteristic feature: The main feature used for detection is its crystal shape which is smaller in size and has rhombic shape. The feature is effectively shown using circle in microscopic image of kurchi in Fig. 1 (b).

2. Medicinal use: The plant's numerous local medicinal uses include as a treatment for dysentery, diarrhoea, fever, snakebite, infertility, venereal disease, diabetes and malaria. H. floribunda has been used as arrow poison [4].

We face some problems like high number of unnecessary blobs in image as shown in Fig. 1(b). Along with this, there are several marks with medium saturation as shown in Fig. 1 (b) using rectangles. Different images have different orientation of the same object which adds to the difficulty in segmentation and feature extraction.

To over-come this we proposed algorithm which extracts unique features from large feature database and provides computationally faster solution using k-means classifier.

Similar methods are used to detect tumour cells from images retrieved from patients and even to classify detected tumor as benign or malignant [5].

We have used microscopic images of licorice and kurchi herbal plant. Our algorithm initially enhances the images by RGB to gray conversion and application of between class variance threshold (BCV). After this, it segments the main object from each image and measures various features of the same. It then gives some of these feature values are given as input to learning algorithm and rest for testing. Using K-Means data mining algorithm, it finally classifies these test images.

Initially we had used area as a feature for classification. Since it is scaling variant, accurate results were not obtained when scaled images were given for testing. To overcome this limitation, we used the ratio of major axis length to minor axis length for the segmented object as a classification feature. We were able to obtain higher true positive rate for this feature.

This paper is organized as follows. In the section II, the entire basic algorithm used is presented. Explanations of each step of the algorithm along with results of each step have been given in section III. All the problems faced in the algorithm and solutions used are also elaborated in this section. The identification results so obtained are discussed in section IV.

II. Proposed Algorithm

In this paper our aim is to provide new framework for microscopic image analysis and intelligent identification systems based on its optimal extracted feature set. As described above we have considered two herbal plants to apply our proposed algorithm.

The characteristic feature of Licorice is the shape of its crystal phloem fiber and that of Kurchi is its rhombic crystal structure as shown in Fig. 1(a) and (b) respectively. The results after each step of the algorithm are shown below. We had 50 microscopic images of each herbal plants. The proposed algorithm is shown in Fig. 2. The algorithm is described in following six steps.

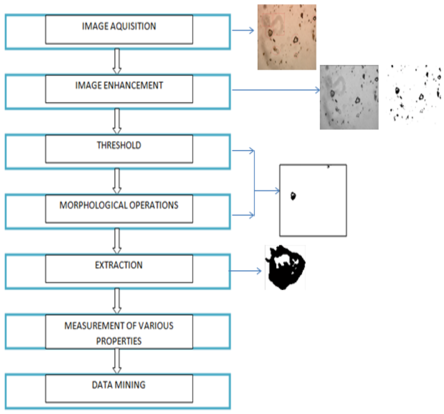


Figure 2 Proposed algorithm.

1. Image Acquisition: 50 licorice images and 50 kurchi images are obtained using the microscope and by application of operations like scaling, rotation and both to the images.
2. Image Enhancement: These images are enhanced in spatial domain by conversion to gray scale followed by application of BCV threshold.
3. Threshold and Morphological Operations: By setting a constant threshold, morphological operations are applied on the image to remove pixels or pixel areas smaller than a predefined value.
4. Extraction: Objects in the image so obtained are labelled and separate images are formed for each object.
5. Measurement of various properties: Various physical characteristics of the feature extracted and measured.
6. Data-Mining: The data so obtained is fed to algorithm of K-Means thus classifying the test image into either kurchi or licorice.

In this section we have described basic steps of our algorithm. In the next section, the detailed analysis of our simulation results is described.

III. Explanation of Algorithm with Simulation Results

1. Image Acquisition and Enhancement:

In the original image, there are several blobs apart from the main kurchi cell. This image has low contrast between background and objects. Also, there are several marks with medium saturation as shown as rectangles in Fig. 1 (b) which can neither be categorized as dark object nor as light background. So initially the image is converted into a gray scale image. Then to separate light background from dark image objects and also to remove these medium saturated blobs, a suitable threshold is found out using the method of between class variance (BCV). This method segments the image into nearly uniform regions. Fig. 3 (a) and (f) show original images, Fig. 3 (b) and (g) show the respective gray scale images, Fig. 3 (c) and (h) show the images after application of BCV threshold for licorice and kurchi respectively.

2. Threshold and Morphological Operations:

As it is apparent from Fig. 1(b), there are many other blobs found apart from kurchi cell. To remove these extra blobs we apply the morphological operator bwareaopen. It eliminates the blobs smaller than specified area in the field. Keeping the constant threshold as 4750 for all Licorice images and 1750 for all Kurchi images, we can eliminate the unnecessary blobs as shown in Fig. 3 (d) and (i).

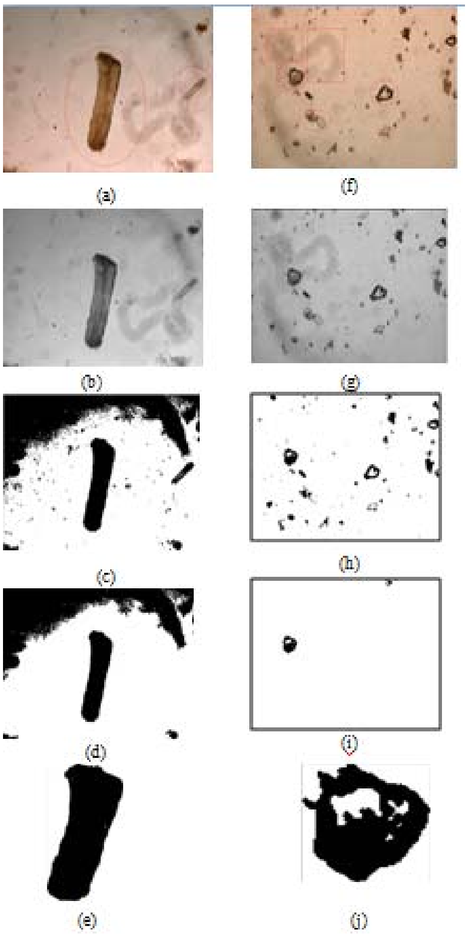


Figure 3 (a) and (f) Original images of licorice and kurchi respectively, (b) and (g) gray scale images, (c) and (h) obtained after BCV thresholding, (d) and (i) obtained after morphological operation, (e) and (j) Images of segmented object

3. Extraction and measurement of various properties:

We now label the object so obtained in the clear image and form a new image carrying only that object. Fig. 3 (e) shows the Licorice component so obtained and Fig. 3 (f) shows the Kurchi counterpart.

We then extract various properties of this object by using regionprops operator. We are considering three properties namely area, major axis length and minor axis length for classification since they have fairly constant range of values for various images of same plant. These ranges are also far apart for kurchi and licorice and hence aid in their separate detection.

We had initially obtained several other properties along with those mentioned above. Since, these other properties either vary largely for various images of the same plant or have similar range for both plants, we considered only the three mentioned above.

Different values of various properties for a number of images are shown in table 1. For scaling invariant properties, we have calculate the new features using equation 1.

$$\text{Ratio} = \text{Major Axis Length} / \text{Minor Axis Length} \quad (1)$$

TABLE 1

Various features of Licorice and Kurchi

	Area	**Major Axis Length**	**Minor Axis Length**	**Ratio**
Licorice	22854	384.40	82.8	4.64
	16119	458.24	67.82	6.75
	36454	407.03	125.28	3.24
	14661	417	108.12	4.78
	29862	473.79	98.99	6.75
Kurchi	2427	71.98	48.83	1.47
	2184	64.43	55.03	1.17
	4272	120.20	54.85	2.19
	7561	143.65	105.47	1.36
	2130	70.53	53.01	1.33

4. Data Mining (K Means Algorithm):

Finally, after the features were extracted, we applied K-Means data mining algorithm for classification of these images. Steps of the algorithm are:

a. Plot all points corresponding to dataset and plot k points for k number of clusters as centroids such that they are as far away from each other as possible.
b. Assign each data point to a centroid such that distance from data point to centroid assigned is minimum. After above step is done, we get k groups of points.
c. Now, calculate new centroid positions and repeat step b to assign each data point to closest new centroid point.
d. Repeat step c till new centroids become fixed points.

The algorithm basically works to minimize the objective function shown in equation (1). x is the data point and c is the cluster center.

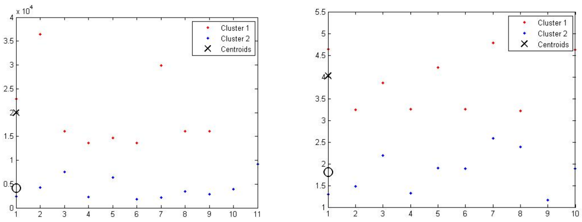
$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left| \left| x_i^{(j)} - c_j \right| \right|^2 \qquad (2)$$

IV. CLASSIFIERS RESULTS

K-Means algorithm is used to classify the test images into the class of either licorice or kurchi. In the algorithm, 2 K-Means clusters are formed using 10 licorice and 10 kurchi images that are used for learning. Respective centroids of the two clusters are calculated using equation 2. We have given 50 test images of each licorice and kurchi and calculate the confusion matrix. Then, from the formula of distance between 2 points we have computed the distance of each image feature value to both the centroid values. The centroid for which the output is a lower value, is the home cluster for the given test image.

Initially we had used area as a parameter to classify these images. Since this is a scaling variant feature, we get lower rate of true detection. We have then applied the algorithm using the feature of ratio of major axis length to minor axis length which is a scaling invariant feature. The K-Means cluster diagrams are shown in Fig. 4.

Fig. 4 shows the points corresponding to two clusters in red for licorice and blue for kurchi and the two respective centroids.



(a) (b)

Figure 4 K-Means clusters (a) Area as feature (b) Ratio as feature

The confusion matrices so obtained for area and ratio of major axis length to minor axis length are shown in Table 2 for licorice and Table 3 for kurchi.

TABLE II

CONFUSION MATRIX FOR LICORICE USING TWO PARAMETERS

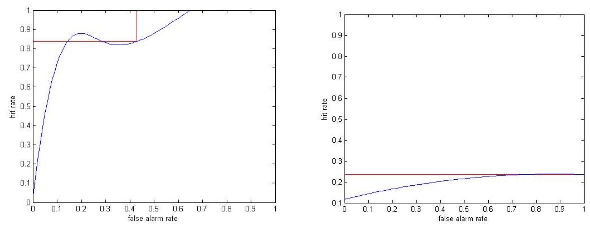
Licorice	Positive (Area)	Positive (Ratio)	Negative (Area)	Negative (Ratio)
True	34	43	46	48
False	4	2	16	7

TABLE III

CONFUSION MATRIX FOR KURCHI USING TWO PARAMETERS

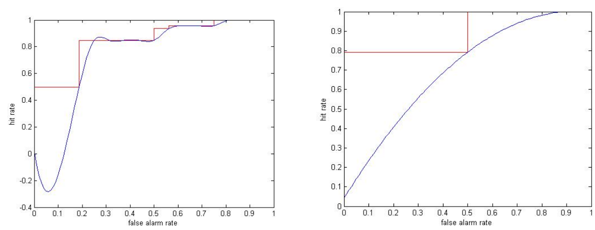
Licorice	Positive (Area)	Positive (Ratio)	Negative (Area)	Negative (Ratio)
True	46	48	34	43
False	16	7	4	2

The Receiver Output Characteristics (ROC) curve so obtained is shown in Fig. 5 (a) and (b) for licorice and Fig. 6 (a) and (b) for kurchi.



(a) (b)

Figure 5 ROC Curve for licorice (a) Area as parameter (b) Ratio as parameter



(a) (b)

Figure 6 ROC Curve for kurchi for (a) Area as parameter (b) Ratio as parameter

V. CONCLUSION

Features based image identification is an effective way of detecting and analyzing microscopic images of herbal plant. It can save a lot of time and effort. We have overcome the problem of misdetection due to scaling by using scaling invariant features. From fig. 4, we can conclude that if we take the area as a features then misclassification increases because of overlapping features of two class. But if we combine the area and ratio as a features for the classification then misclassification reduce.

The advantages offered by our algorithm are:

- Elimination of sample treatment, a highly cumbersome process. Thus, contributing to cost and energy saving.
- Drastic increase in speed and efficiency compared to Manual Identification of plants.
- It helps in automation of pharmaceutical labs and companies.

At the same instant, apart from being economical, this method is environmental friendly, as no pollution or waste is produced. Thus, this technique is highly beneficial for pharmaceutical industries and would soon be employed for various other fields, at large.

VI. FUTURE SCOPE

In this algorithm threshold used for segmentation of main object is manually set such that it satisfies most of our images. For algorithm's application to images with drastically different magnification levels, a new threshold value has to be calculated. Thus an algorithm could be made to automatically calculate threshold values.

Secondly, the number of images taken in our data set is limited. So a large data set could be applied to generalize the findings.

VII. REFERANCES

[1]. Medicinal Plants: Phytochemistry Pharmalogy and Therapeutics Vol 2, 2011, ISBN13: 9788170357681

[2]. Sato, H; Goto, W; Yamamura, J; Kurokawa, M; Kageyama, S; Takahara, T; Watanabe, A; Shiraki, K (May 1996). "Therapeutic basis of glycyrrhizin on chronic hepatitis B.". Antiviral Research 30 (2-3): 171–7. doi:10.1016/0166-3542(96)00942-4. PMID 8783808

[3]. Kannappan, R; Gupta, SC; Kim, JH; Reuter, S; Aggarwal, BB (October 2011). "Neuroprotection by spice-derived nutraceuticals: you are what you eat!" (PDF). Molecular Neurobiology 44 (2): 142–59. doi:10.1007/s12035-011-8168-2. PMC 3183139. PMID 21360003.

[4]. Medicinal Plants. PROTA. 2008. pp. 329–332. ISBN 978-9-05782-204-9.

[5]. Dijana Tralic, Jelena Bozek, Sonja Grgic, "Shape Analysis and Classification of Masses in Mammographic Images using Neural Networks", Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on 16-18 June 2011, ISSN:2157-8672, ISBN:978-1-4577-0074-3.