Automatic Identification and Classification of Nutritional Defeciencies and Diseases in Coffee Plants using Digital Image Processing

Shenthan Vardhi, Chandana K R, Dayananda S, Yogaraj J, Mrs. Nagarathna R

Student, Student, Student, Student, Assistant Professor
Department of Telecommunication
Dayananda Sagar College Of Engineering, Bangalore, India

Abstract : In India the hill tracks of south Indian states are dominating in coffee production while Karnataka process 70% followed by Kerala 21% and Tamilnadu. Finding nutritional deficiencies in coffee leaves is so problematic since some deficiencies have same visual characteristics such as iron, manganese. These deficiencies cause major effects in producing coffee. This paper presents automatic classification of nutritional deficiencies using machine learning approach based on decision tree using local and global features.

Keywords—Deficiencies, decision tree, local features, global features.

1. Introduction

Plants require sufficient nutrition's for its normal and healthy growth. However, when plant grows without enough nutrition's abnormal changes appears in the plant such as change in shape, size, color and others. The characteristics of leaf is important in finding the deficiencies of leaf in which two characteristics of leaf is considered: the first is local features which describes the properties of pixels relating to its neighbors. Second is global features which includes shape, height, tip of the leaf and color features. Some deficiencies can be identified visually by symptoms such as shape, chlorosis, and changes in normal leaf coloration. However, when deficiency is detected the leaf is cut and analyzed to take corrective actions but this process is expensive and takes time.

In the literature, there are few works on image processing. Image identification using shape-based descriptor is proposed in [1] it presents a novel boundary-based approach to identify and differentiate the leaf. Deficiency is also depending up on shape example calcium and boron so this method will be helpful for finding deficiency using this approach. In [2] the cloud-based approach automatic identification of leaf disease using mobile phone is proposed. They used AT Mega microcontroller for sensor readings, cloud servers for storing data and raspberry pi for android application is proposed. In [3] different models of bag of features are proposed. Bag of features outlining has 4 steps: first one is extracting features. Second one is learning visual vocabulary. Third one quantizing feature using vocabulary. Forth one representing images by frequencies or binary codes. In [4] in this the performance of Squared root velocity function (SRVF) of planar curves for identifying leaf shape is proposed. In [5] [6] the new approach to extract feature is proposed where 2-Dimensional Empirical wavelet transform is used for extracting entropy features. In [7] leaf image contrasting using SIFT descriptor is used. This descriptor is used to classify the local features of an image. A new method of identifying leaf using leaf vein extraction is proposed in [8] this extraction is based on mathematical morphology. A new machine learning technique is used for image identification [9] the SVM classifier is used for classification and identification of leaf images.

2. Methodology

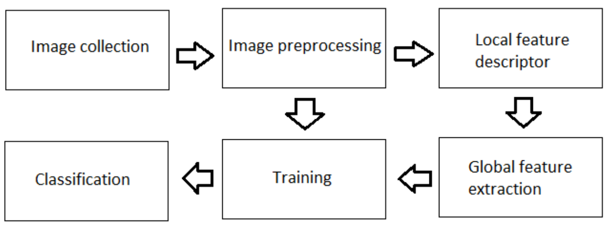


Figure 1: flow of proposed approach

2.1. Image preprocessing

Image preprocessing is significant for genuine data that are frequently noisy or uneven. During this phase the transformation is applied to convert image to another image to improve quality for better analyzing. Then the image is segmented. In the obtained segmented image, white pixels represents the background and black pixels represent the leaf area.

Finally, the image size is reduced and the background with- out information is removed to keep only the leaf area in order to reduce the computational cost of feature extraction.

2.2. Local feature extraction

A local feature refers to pattern or distinct structure found in an image. An image can be described by key points. The content of interest regions around each key point can be found by using a descriptor. The common algorithm used , in find local features, is the SIFT descriptor.

The Scale Invariant Feature Transform (SIFT) is feature extraction algorithm introduced by Lowe in 1999 . The SIFT algorithm has two steps: Key points detection and key points description. Each key point is described with a vector of length 128*128.

2.3. Global feature extraction

Global features include contour representations, shape descriptors, and texture features represents the texture in image patch. Global features include 5 shape and 8 color features.

2.3.1 Basic features

1. **Leaf area:** It is calculated by counting the number of pixels segmented as leaf in a binary image.
2. **Leaf perimeter:** It is calculated counting the number of pixels in the leaf margin.
3. **Leaf diameter:** It is the maximum distance between two points, (x_i, y_i) and (x_j, y_j), in the leaf margin:

$$D = max\{(x_i - x_j)^2 + (y_i - y_j)^2\}. \quad (1)$$

2.3.2 Shape features

Using this feature, we can find deficiencies like boron, calcium, iron.

Apex height: we can find boron deficiency. The top position (xa, ya) on the tip of leaf is found. Then, the distance from the position (xa − m, ya) which is h1 to the position (x1, y1) is calculated. similarly, the distance from the position (xa + m, ya) that is h2 to the position (x2, y2) is calculated. It is as shown in below figure.

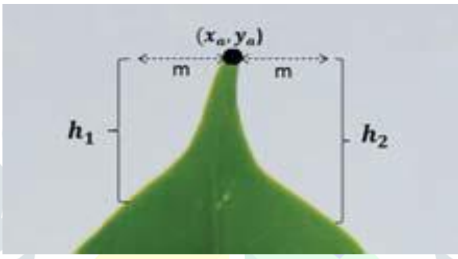


Figure 2: apex height of a leaf

$$\text{Apex height} = \frac{h1+h2}{2}. \quad (2)$$

Roundness: Roundness is defined as the ratio between the leaf area and the area of the perfect circle that content the leaf:

$$Roundness = \frac{100A}{\pi r2} \quad (3)$$

colour features are calculated to find the deficiency of nitrogen. Potassium. For example, we can calculate percentage of red colour by deriving hue, value, and saturation component from an image. The percentage of red colour can be defined as:

$$\text{Red percentage} = \frac{\sum_i \sum_j R(x_i, y_j)}{A} \quad (4)$$

2.4. Training

Machine learning can be appear in many guises. We now discuss a number of applications, the types of data they deal with, and nally, we formalize the problems in a somewhat more stylized fashion. The latter is key if we want to avoid reinventing the weel for every new application. Instead, much of the art of machine learning is ti reduce a range of fairly disparate problems to a set of fairly narrow prototypes. Much of the science of machine learning is then to solve problems and provide good guarantees for the solutions.

2.5. *Classification*

There are 8 different deficiency to classify in coffee leaf.

1. **Iron**: leaves are chlorotic between veins; veins remain green. Necrotic spots usually absent.
2. **Phosphate**: foliage is dark green. Growth is retarded. Lower leaves sometimes yellow between vein, leaves often dropping early.
3. **Nitrogen**: foliage is light green, growth is stunted, stalks slender. Leaves small, lower one's lighter yellow than upper. Yellowing followed by drying to a light brown color.
4. **Boron**: Shape deformity, lack of special shape in leaf margin and total loss of apex (tip of leaf).

5. **Calcium:** shape deformation in leaf and undulations in leaf margin.

6. **Magnesium**: the visual effects of this is similar to that of iron deficiency but yellow coloration appears between secondary veins.

7. **Manganese:** chlorosis in leaf area. Veins remain with green coloration and margin appear on both sides of the veins with color green.

8. **Potassium:** beginning of necrosis on the leaf tip, with a yellow halo limiting it.

3. PROPOSED APPROACH

The proposed approach for automatically classifying nutritional deficiencies is plain. The input image is basic RGB image, this image is preprocessed, then visual features are extracted using local and global features. Finally, a machine learning technique called SVM (Support Vector Machine) is used to classify deficiencies in image. SNM based on decision tree technique.

3.2. SIFT Algorithm

SIFT algorithm transformers an image into a collection values of local feature vectors. Each of this vector are supposed to be unique and fixed to any scaling, rotation or translation of image.

In order to detect the local maxima and minima of an image each sample point is compared with to its different points in same image. It is elected

+ only if it is larger than all of these neighbours or smaller tan all of them. The cost of this check is reasonably low due to the fact that most sample points will be eliminated following the first few checks.

The SIFT descriptor is calculated for each image in the data-set. These descriptors are quantized using K-means algorithm and result is stored. We can't use same or similar window to detect key points with different scale. It is ok with detecting small corners. But to detect large corners large windows are needed. For this purpose, scale-space filtering is used. In this, Gaussian values are found for the image with different lambda values. LoG acts as a blob detector which detects blobs in various shapes and sizes. But using LoG will be costly, so SIFT algorithm uses difference of Gaussians which is an approximation of LoG

RGB values are extracted for calculating global features of an image.

3.3. K-means clustering

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means due to the name. Applying the 1-nearest neighbor classifier to the cluster centers obtained by k-means classifies new data into the existing clusters.

3.4. *SVM(Support Vector Machine*)

It is a machine learning technique used to classify and detect the type of leaf disease. Classification deals with associating a given input pattern with one of the distinct class. In the given system a linear support vector machine is used for classification of lead disease.

SVM is a binary classifier which uses a hyper plane called the decision boundary between two classes. The hyper plane tries to divide, one class containing the target training vector which is labeled as +1, and the other class containing the training vectors which is -1. Using this labeled training vectors. SVM optimizer finds a hyper plane that will then maximizes the margin of separation among the two classes. Example is as shown below.

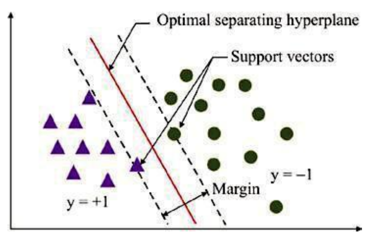


Figure 3: SVM in linearly separable condition

4. RESULTS AND DISCUSIONS

There are two stages in the process of classifications, training and testing using classifier. In training phase, classifier is trained using feature values and its respective target values. This trained classifier is used to classify test image.
First step is to preprocess the input image below figure 4 shows input image vs preprocessed image.

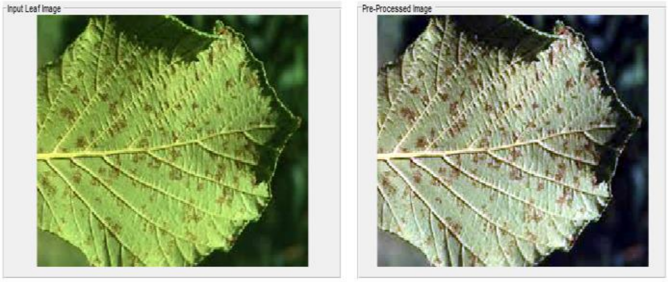


Figure 4: input mage and preprocessed image

The preprocessed image is segmented and clustered using k-means clustering. The HSV values are obtained.

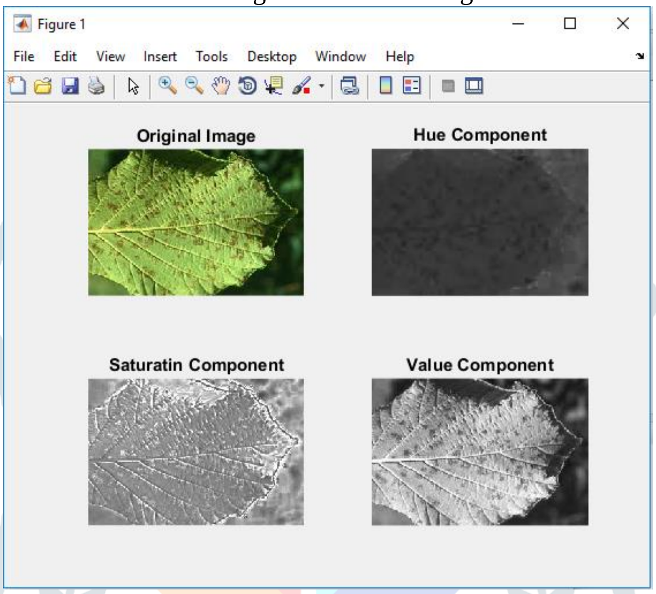


Figure 5: HSV components

Figure 6: segmented image

In this work 8 type of nutritional deficiencies and disease of coffee pants are detected automatically. Nine texture features and 4 color features are calculated for all segmented parts of single leaf image. This feature values, collectively called as feature vector, is given to trained SVM classifier which classifies the input leaf into two classes.

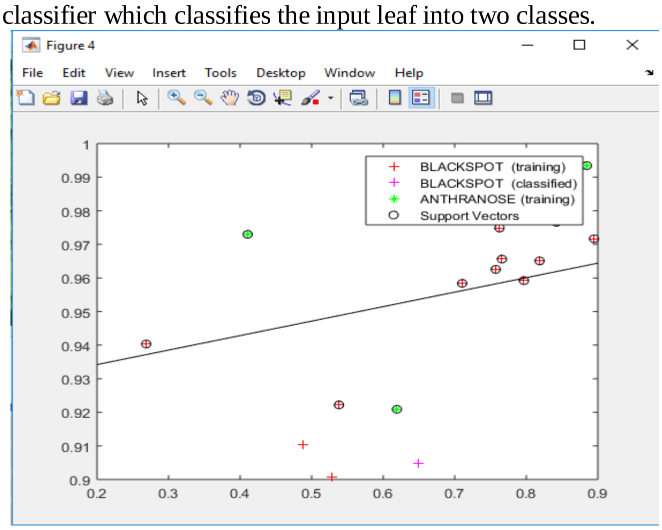
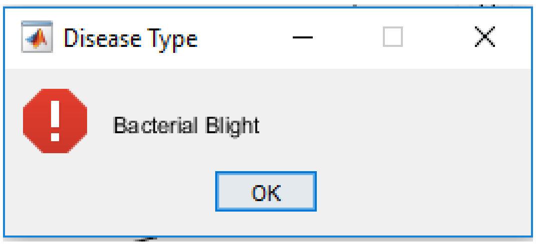


Figure 7: Machine learning output

The image shown in figure 8 is coffee leaf based on boron deficiency so the SVM classifies into boron class and result is presented as shown below



5. CONCLUSION AND FUTURE WORK

The given solution uses K-means clustering technique foe segmentation of image then feature extraction is done using both texture as well as color features. Then SVM classification technique is used to detect the type of leaf disease.

In future work more algorithms can be used to detect multiple deficiencies in a single image and on a single leaf. And also, can used more algorithms can be used to improve detection rate of the classification process.

ACKNOWLEDGMENT

We would like to thank our respective guide Mrs. NAGARATHNA and HOD Dr. A R ASWATHA for their support. we gratefully acknowledge both of them for imparting us valuable knowledge of Digital image processing and Embedded system.

REFERENCES

[1] Boikobo Tlhobogang and Muhammad Wannous, “On Desigh of plant disease detection system,” Department of information system, Graduation school of information Technology, IEEE ICASI 2018-Meen, Prior and Lam.

[2] Jagadesh D. Pujari, Rajesh Yakkundimath and Abdulmunaf S.Byadgi, “Identification and classification of fungal disease Affected on Agriculture,” 978—4799-3975-6/14©2014 IEEE.

[3] Qing Liu*, Xiao-long Zhao, Xiao-ping Yang, “Feature extraction and Recognition of chinese herbal Medicine Based on pilse coupled neural networks,” School of Electronic Information and Electrical Engineering, 978,1,5386,4991,6/18 ©2018 IEEE

[4] Diego Mondalve*,Maria Trukillo, Deisy Chaves, “Classification of Nutritional dsficiencies in doffee plants,” unpublished.

[5] Bhupendra Singh Kirar and Dheeraj K. Agrawal, “Maulana Azad Institute of Ttechnology,” IEEE ISBN 978-1-5090-4760-4 ©2017.

[6] Hamid Laga*, Sebastian Kurtek, Anuj Srivastava,”A Riemannian Elastic Matric for shape-based plant Leaf Classification,” Department of statistics, The Ohio state University, IEEE 978-1-4673-2181-5/12 ©2012

[7] Ranjith, Saheer Anas, Ibrahim Badhusla, Zaheema OT,”Cloud Based Automated Irrigation And Plant Leaf Detection System,” Eranad Knoeledge City Technical Campus, IEEE ICECA 978-1-5090-5686 ©2017.

[8] Ned Horning,”Random Forest: An Algorithm for Image classification and generation of continous fields data seta,” Center for Biodiversity and Conservation SDEAS 2010.

[9] Pranjali B. Padol and Prof. anjali A. Yadav,”SVM Classifier Based Grap leaf disease detection,” smt Kashibai Navle college of Engg, IEEE CASP, 978-1-5090-0849-0/16 ©2016.