

# Identification of Vegetable Plant Species Using Support Vector Machine



**K. Deeba and B. Amutha**

**Abstract** This study proposes a method for the identification of vegetable plant species. Each plant leaf has its own features that can be used to identify the species it belongs to. Some of the features of a leaf that enable specific identification of a plant species are its shape, vein pattern, apical and basal features, and color patterns. Those salient features extracted from the leaf image are used along with a data mining algorithm, such as support vector machine, to identify of the species that the leaf belongs to. In this study two vegetable species, namely eggplant and ladies' fingers were considered. Support vector machine is suited to situations where the data need to be classified into two groups.

**Keywords** Shape information • Vein pattern • Apical and basal features  
Colour patterns • Support vector machine

## 1 Introduction

India is a county based on agriculture. It is the second largest country in the production of fruits and vegetables in the world, after China. It is essential to identify vegetable plant species automatically using images. Each plant leaf has its own features that allow it to be classified as different from other plants. Automatic identification of plant species is a type of pattern recognition problem. Images that are taken in real life always tend to show changes in the physical structure. Even in such cases there are few common features that can be used to identify the plant

---

K. Deeba (✉) · B. Amutha

Department of Computer Science and Engineering, SRM University, Chennai, India  
e-mail: deeba.k@ktr.srmuniv.ac.in

B. Amutha

e-mail: amutha.b@ktr.srmuniv.ac.in

species, such as the shape of the leaf, color pattern, and vein structure etc. Some of the salient features considered in this paper are leaf shape, vein pattern, apical and basal features, and color patterns.

Shape descriptors are of two types. One is a region-based descriptor and the other is a counter-based descriptor. A region-based descriptor is used to describe the shape of the object based on boundary and inner pixel information. A counter-based descriptor describes only the outer boundary information of the object and includes conventional representation and structural representation.

Nine different vein patterns are considered widely and for our example we considered two classes; ladies' fingers with palmate venation and eggplant with reticulate venation. After the extraction of the features a support vector machine classifier is built for classification.

This paper is organized as follows: Sect. 2 describes related works, section three discusses proposed work, and Sect. 4 contains the conclusion.

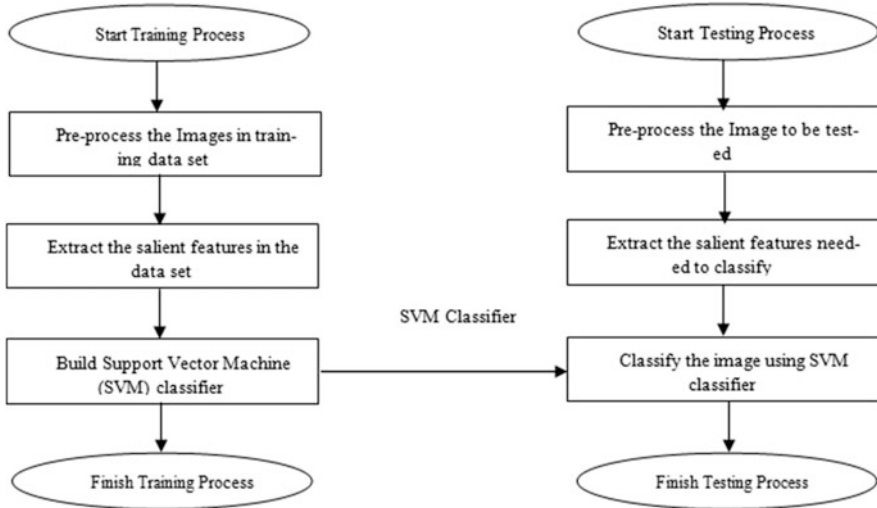
## 2 Related Work

Leaf recognition based on shape [1, 2] is an old method and when the leaf samples are incomplete this method cannot be used for correct recognition. Crowe and Delwiche [3], developed an algorithm for analyzing apple and peach defects. He obtained an accuracy of 70% by using near-infrared (NIR) images. Nakano [4], applied a neural network for color grading of apples, and achieved an accuracy of 75%.

Pydipati et al. [5], used statistical and NN classification for citrus disease detection using machine vision. He gained an accuracy of 90% using a texture analysis method, depends on lab results. Lei et al. [6], worked on plant species identification based on the neural network algorithm called a self-organizing map (SOM). Liu et al. [7], worked on plant leaf recognition based on a locally linear embedding and moving center hypersphere classifier and achieved an accuracy of 92%. Shilpa et al. [8], used different neural network algorithms for plant species identification and made a comparative analysis.

## 3 Proposed Approach

This proposed work aims to identify plant species based on the following leaf features: shape information, vein pattern, and apical and basal features. The images used were pre-processed in order to enhance feature extraction.



### 3.1 Pre-processing and Extraction of Shape Information

The major aim of pre-processing is to remove all noise and to discard unwanted descriptions. This step includes converting the image form RDG to a gray-scale image, this gray-scale image will be then converted into a binary image and finally that binary image will be converted into a contour image which only has the external boundary of the image. The following 11 features will be extracted:

1. Area: The area enclosed by the leaf
2. Perimeter: The number of pixels along the margin of the leaf
3. Diameter: The longest distance form a point to another point on the leaf.
4. Length (Major Axis): The distance between base and apex.
5. Width (Minor Axis): The longest line perpendicular major axis connecting two points in leaf.
6. Aspect ratio: The ratio between the major axis (x) and the minor axis (y).

$$\text{Aspect Ratio} = x/y$$

7. Rectangularity: Finding the similarity of the leaf shape to a rectangle.

$$\text{Rect} = (\text{Area}/\text{The area of the smallest rectangle that enclose leaf}) * 100$$

8. Circularity: Finding the similarity of the leaf shape to a circle.

$$\text{Circ} = (4 * \pi * \text{Area})/\text{Perimeter}^2$$

9. Eccentricity: Finding the similarity of the leaf shape to a cone.

$$\text{Cone} = ((x^2 - y^2)/x^2)^{1/2}$$

10. Convexity: Number of concaves on the contour.

#### **Apical and basal features:**

The two leaf classes that we considered in this study have different apical and basal features. To obtain these features an origin point O is considered for both apex A and base B. Two points were chosen from the origin; one on the left side (L), and one on the right side (R) of the contour. Then the angle is formed from that points to that of the main line connecting apex A and base B point. These angles were used as 4 features.

#### **Vein features extraction:**

These features are obtained by performing morphological operations on the gray-scale image. The purpose of applying morphology processing to a gray-scale image is to get rid of the gray overlap between the leaf vein and the background. For the proposed plant classification model, the RGB leaf image is converted to a gray-scale image. Then, opening operations are performed on the gray-scale image with a flat disk-shaped structuring element of varying radius (1, 2, 3, 4). The resultant image is then subtracted from the margin of the leaf. Finally, the converted image is binaried. Applying these stages leads to the obtaining of the two leaf vein features.

#### **Dataset:**

The images for this work were collected from the fields around Salem district in Tamil Nadu, India, using a Sony Cybershop W810 20.1MP digital camera with 20.1 mega pixels. The images were taken against a white background with 100 images of each species for training and 50 images for testing. Hence, there is a total of 300 images in the dataset.

#### **Leaf classification using support vector machine:**

Support vector machine (SVM) is a type of supervised classification algorithm. Here the training samples are associated with the class label and the classifier is

built based on the training samples which will be used for classification. With the extracted features of training set with instance label pair  $(a_i, b_i)$ ,  $i = 1, 2, \dots, l$ , where  $a_i \in \mathbb{R}_n$  and  $b_i \in \{-1, 1\}$ , the support vector machine requires the solution of the following optimization problem,  $\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i$  showing that  $b_i((w^T \cdot \phi(a_i) + b) \geq 1 - \xi_i$ ,  $i = 1, 2, \dots, l$  and  $\xi_i \geq 0$ ,  $i = 1, 2, \dots, l$ .

Here the training sample vectors  $a_i$  are mapped into a higher dimensional space by the function  $\phi$ . SVM finds a hyper plane which separates the region in a linear way in the given higher dimensional space. The penalty parameter of the error term  $C > 0$  is used and  $K(a_i, a_j) = \phi(a_i)^T \cdot \phi(a_j)$  is called a kernel function. With these functions the classifier is built so that the classification of the test samples can be verified. SVM works well for binary classification (classification of samples in two classes).

## 4 Conclusion and Future Work

Automatic identification of plant species is a type of pattern recognition problem. The images are taken in the real world and hence always tend to have changes present in the physical structure. This study proposes a method for the identification of vegetable plant species. The salient features were extracted from the leaf image and support vector machine (SVM) was used for the identification of the leaf species. Two different varieties of plant species were considered in this study and SVM is well suited to cases where the data need to be classified into two groups. In future work this can be extended to multi-class classification.

## References

1. Javed Amin H, Ashraf M (2010) Leaf shape identification based plant biometrics. In: 13th International conference on computer and information technology (ICIT), pp 458–463
2. Du J-X, Wang XF, Zhang G-J (2007) Leaf shape based plant species recognition. *Appl Math Comput* 185
3. Crowe TG, Delwiche MJ (1996) Real-time defect detection in fruit: part II. An algorithm and performance of a prototype system. *Trans ASAE* 39(6):2309–2317
4. Nakano K (1997) Application of neural networks to the color grading of apples. *Comput Electron Agr* 18(2–3):105–116
5. Pydipati, R, Burks TF, Lee WS (2005) Statistical and neural network classifiers for citrus disease detection using machine vision, by information & electrical technologies division of ASABE in August 2005, Florida Agricultural Experiment Station Journal Series R-10626
6. Zhang L, Kong J, Zeng X, Ren J (2008) Plant species identification based on neural network. In: Fourth international conference on natural computation, 2008. ICNC '08. vol 5, pp 90–94
7. Liu J, Zhang S, Liu J (2009) A method of plant leaf recognition based on locally linear embedding and moving center hypersphere classifier. In: International conference on intelligent computing ICIC 2009, LNAI 5755, pp 645–651. Springer

8. Ankalaki S, Majumdar J (2015) Leaf identification based on back propagation neural network and support vector machine. In: 2015 International conference on cognitive computing and information processing (CCIP), pp 1–7
9. Purusothaman G, Krishnakumari P (2015) A survey of data mining techniques on risk prediction: heart disease. *Indian J Sci Technol* 8(12)
10. Vedanayaki M (2014) A study of data mining and social network analysis. *Indian J Sci Technol* 7(S7):185–187