Classification of vegetation using CBIR technique

1. Abstract

As the propagation of video and image data in digital form has increased, content based image retrieval (CBIR) has become a prominent research to pick; therefore an important problem that needs to be addressed is fast retrieval of images from large databases, to find images that are perceptually similar to a query image. Image retrieval systems attempt to search through a database. CBIR can greatly enhance the accuracy of the information being returned and is an important alternative and complement to traditional text-based image searching.

1. Methodology

Content-based image retrieval (CBIR) is a technique for retrieving image on the basis of automatically-derived features such as color, texture and shape. The architecture of a CBIR system can be understood as a basic set of modules that interact within each other to retrieve the database images according to a given query. In typical content-based image retrieval system, the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. To retrieve images, users provide the retrieval system with query image or sketched figures. The system then changes the query image into its internal representation of feature vectors. The similarities/differences between the feature vectors of the query example and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme.

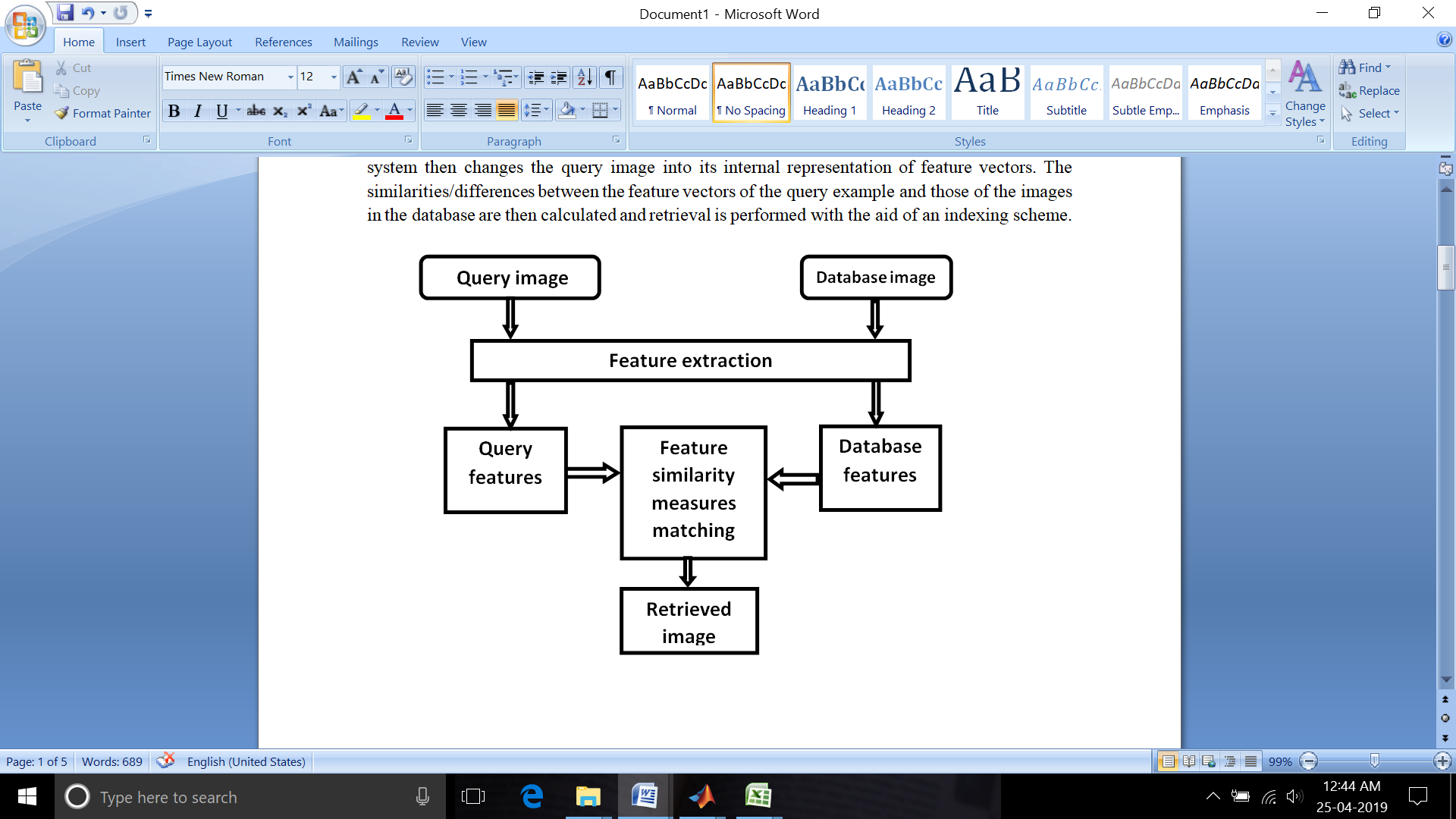


Fig 1. Block diagram of content-based image retrieval system

* Given a query image, the system uses the interactive segmentation step to segment the plant from the background.
* Segmentation steps ensures the extraction of relevant features, by focusing on the plant region rather than the background.
* Database is the collection images and their features.
* In typical content-based image retrieval system, the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors.
* The feature vectors of the images in the database form a feature database.
* To retrieve images, users provide the retrieval system with query image or sketched figures.

The system then changes the query image into its internal representation of feature vectors.

* The similarities/differences between the feature vectors of the query example and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme.

1. Implementation

Implementation of this project follows these steps:

1. Collect the image samples of leaves with different shape, size, and colour. And number them in a sequence as shown in the figures below.

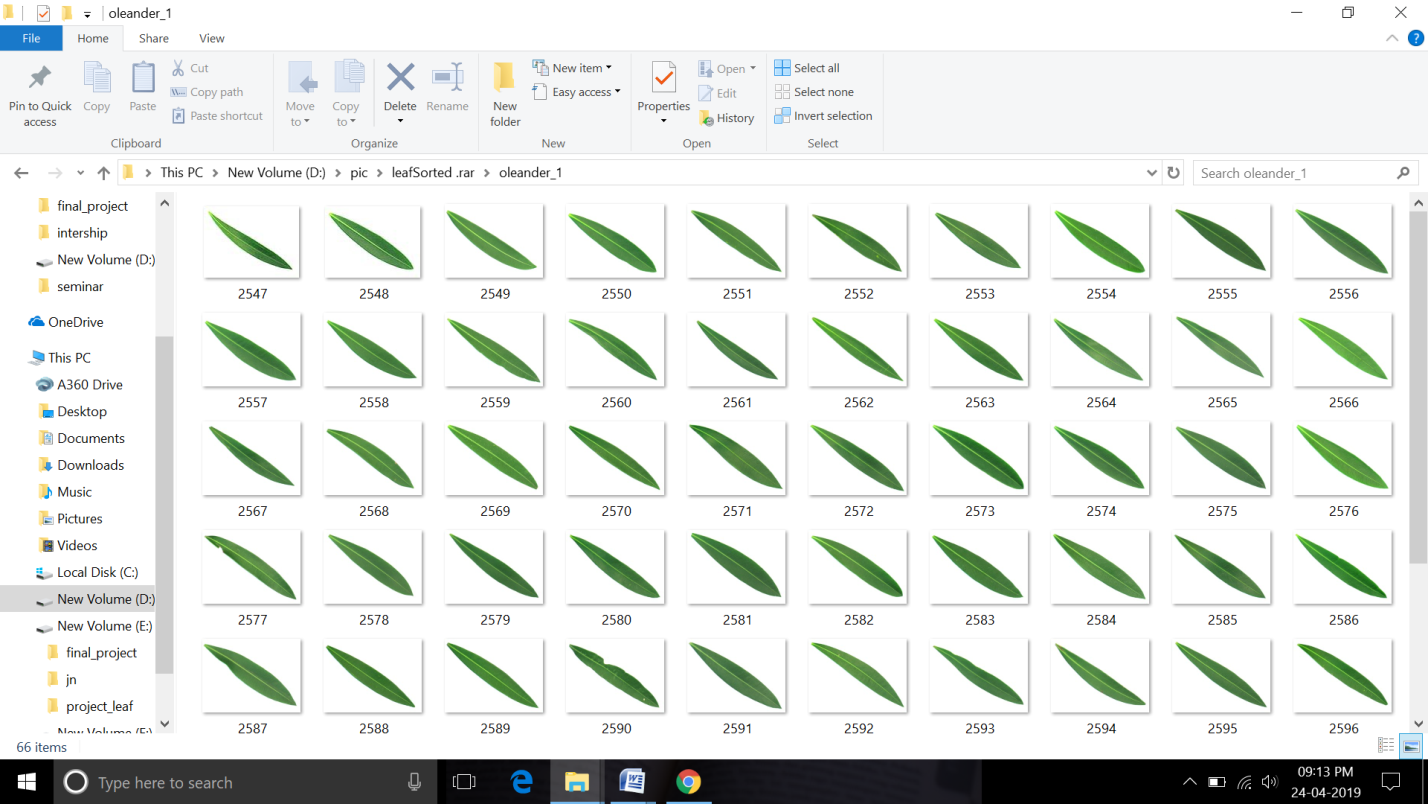


Fig 1. Collection of oleander leaves

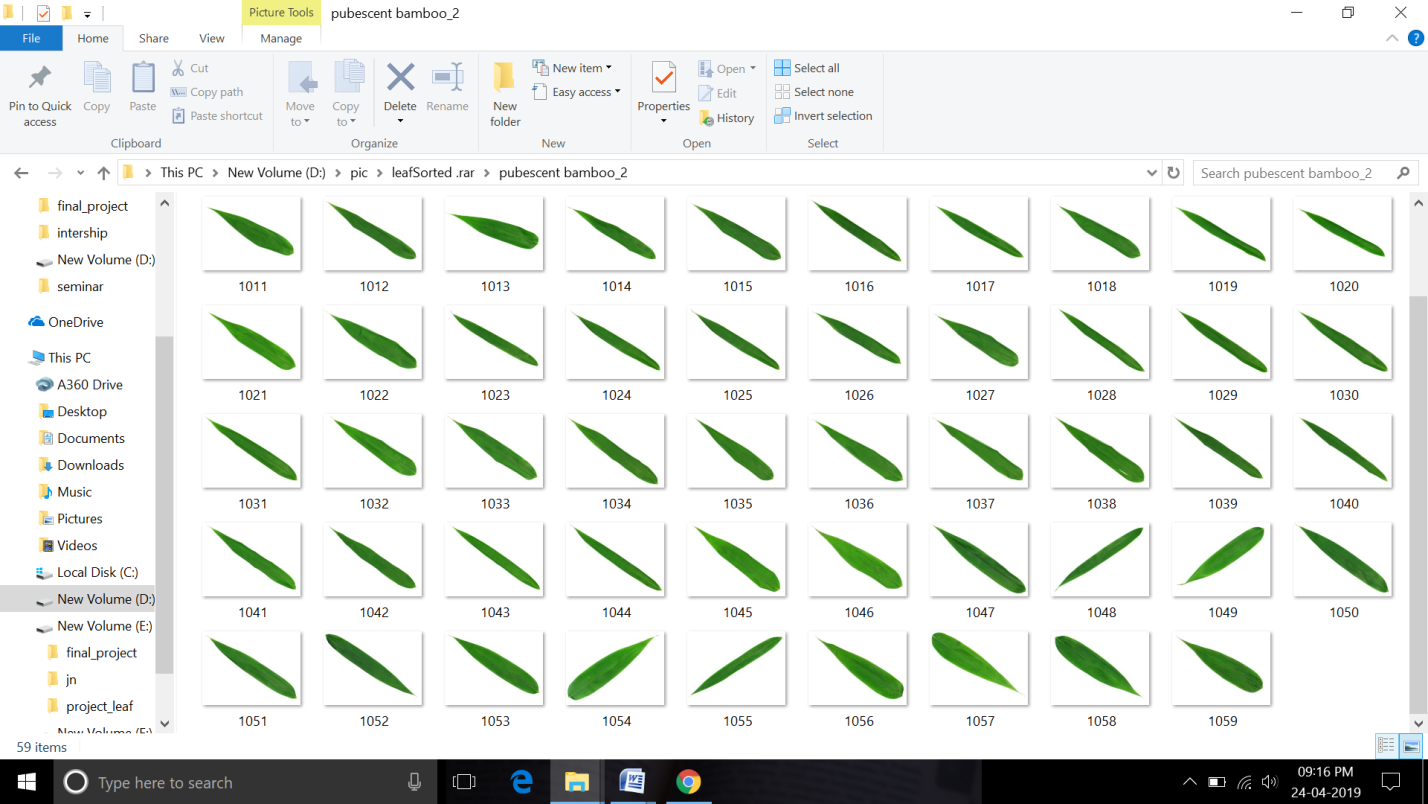


Fig 2. Collection of Bamboo leaves

1. Perform image threshold operation by adjusting the hue, saturation and intensity values. Using this, an object with a certain colour can be detected and to reduce the influence of light intensity from the outside

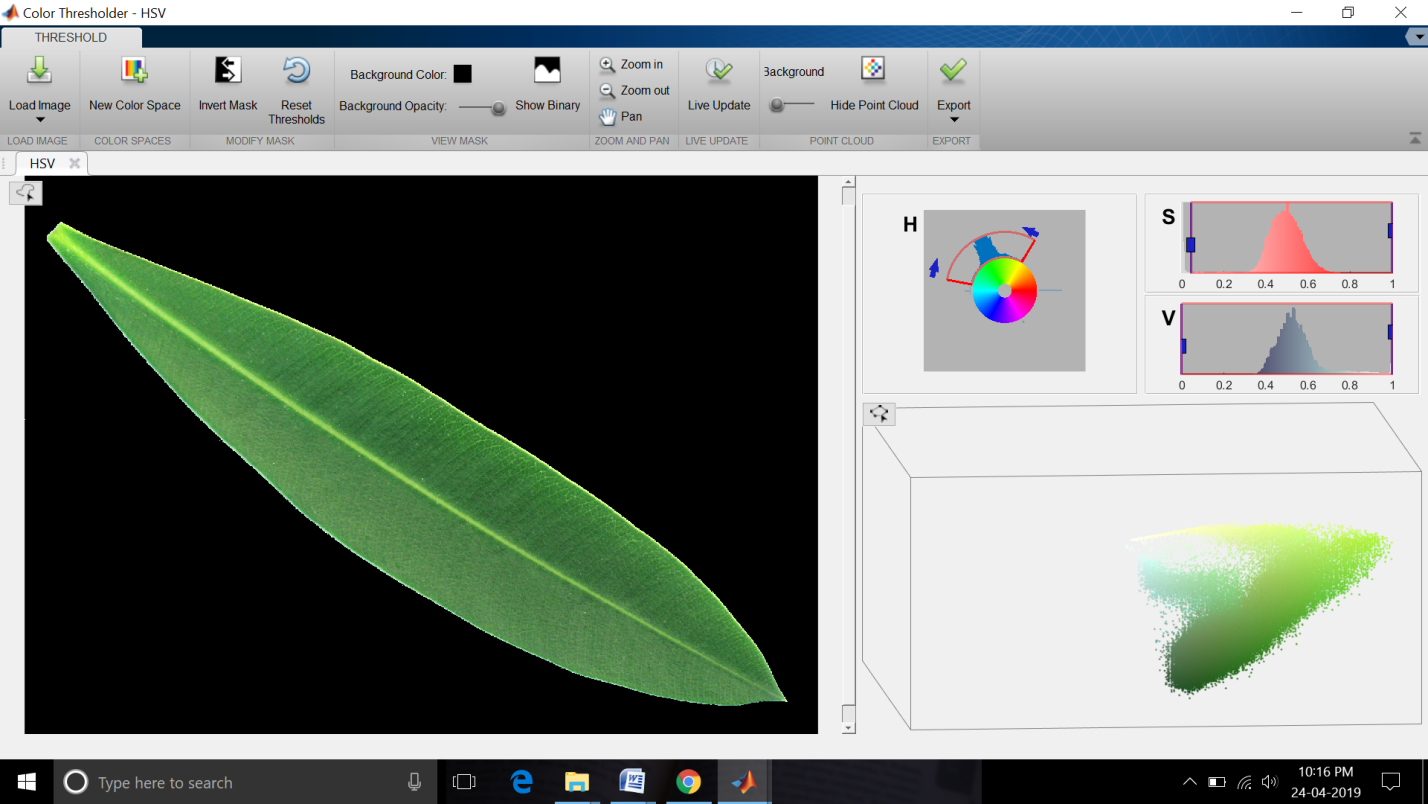


Fig 3. Image threshold

1. Perform image segmentation operation and morphology operations such as dilation, erosion, etc. using this, a foreground object can be distinguished from the background.

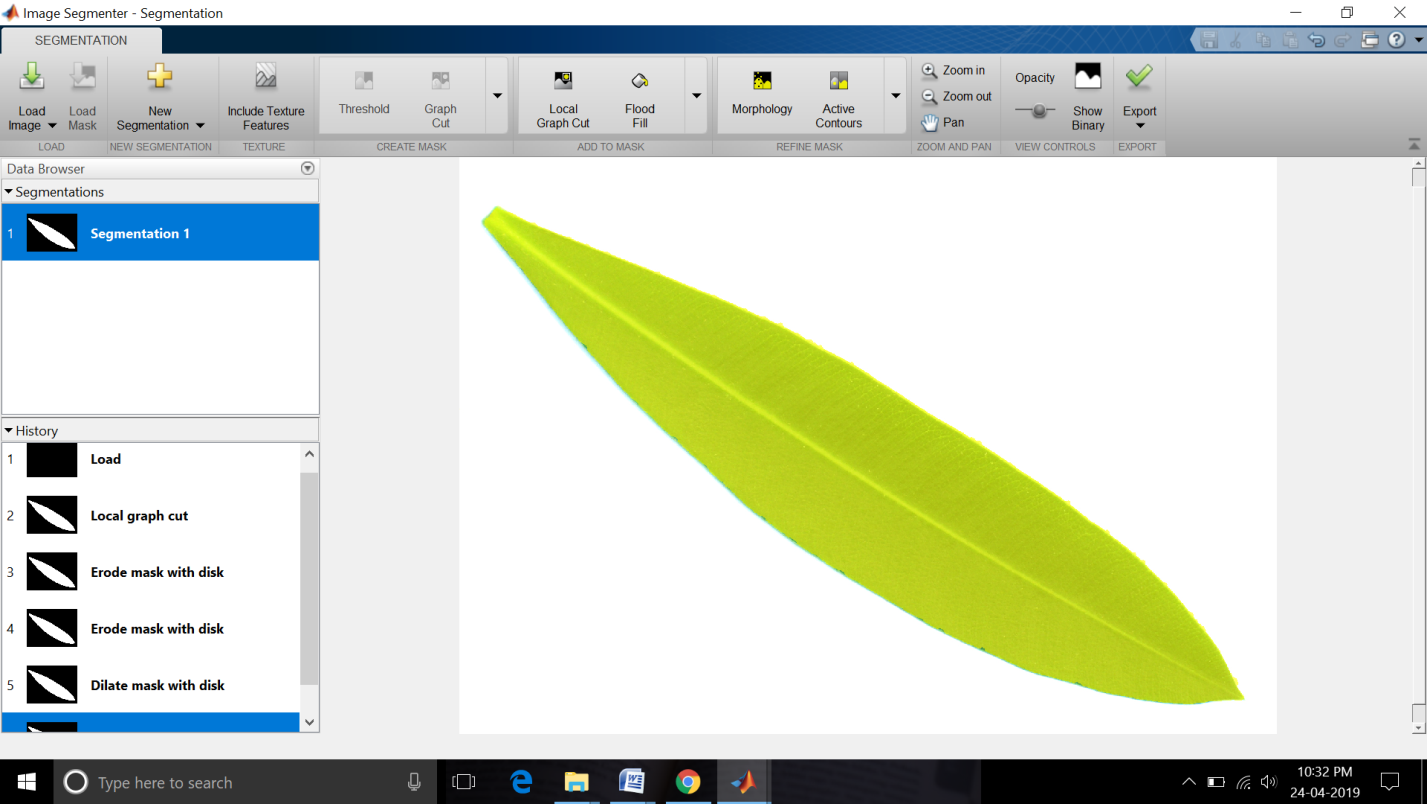
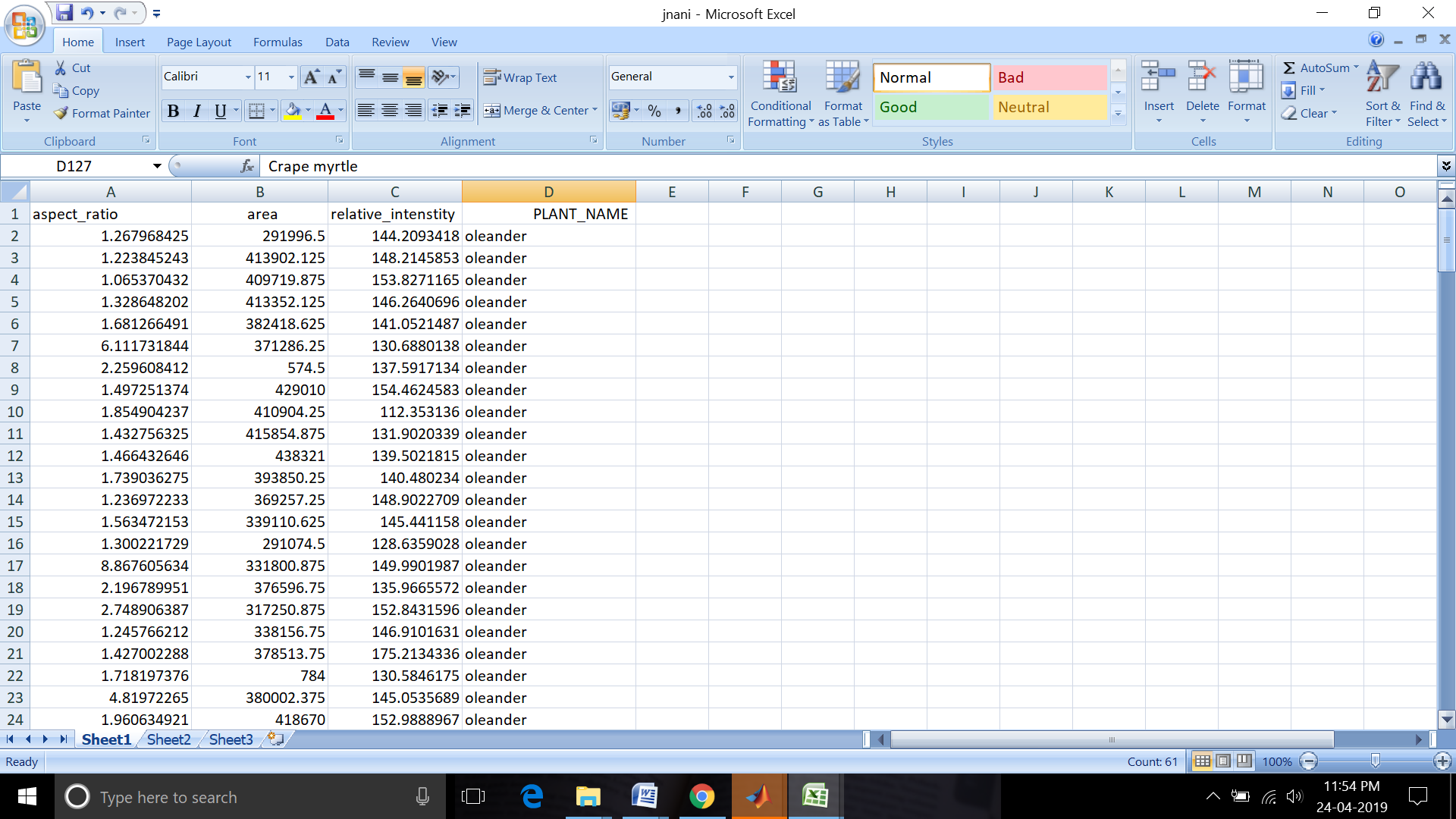


Fig 4. Image segmentation

1. Extract the values of axial ratio, relative intensity and area of individual leaf. And create a excel file of these values including the category of the leaves.



1. Now train the data using the classification learner. From the below information we can say that the accuracy of traning is 87.6%.

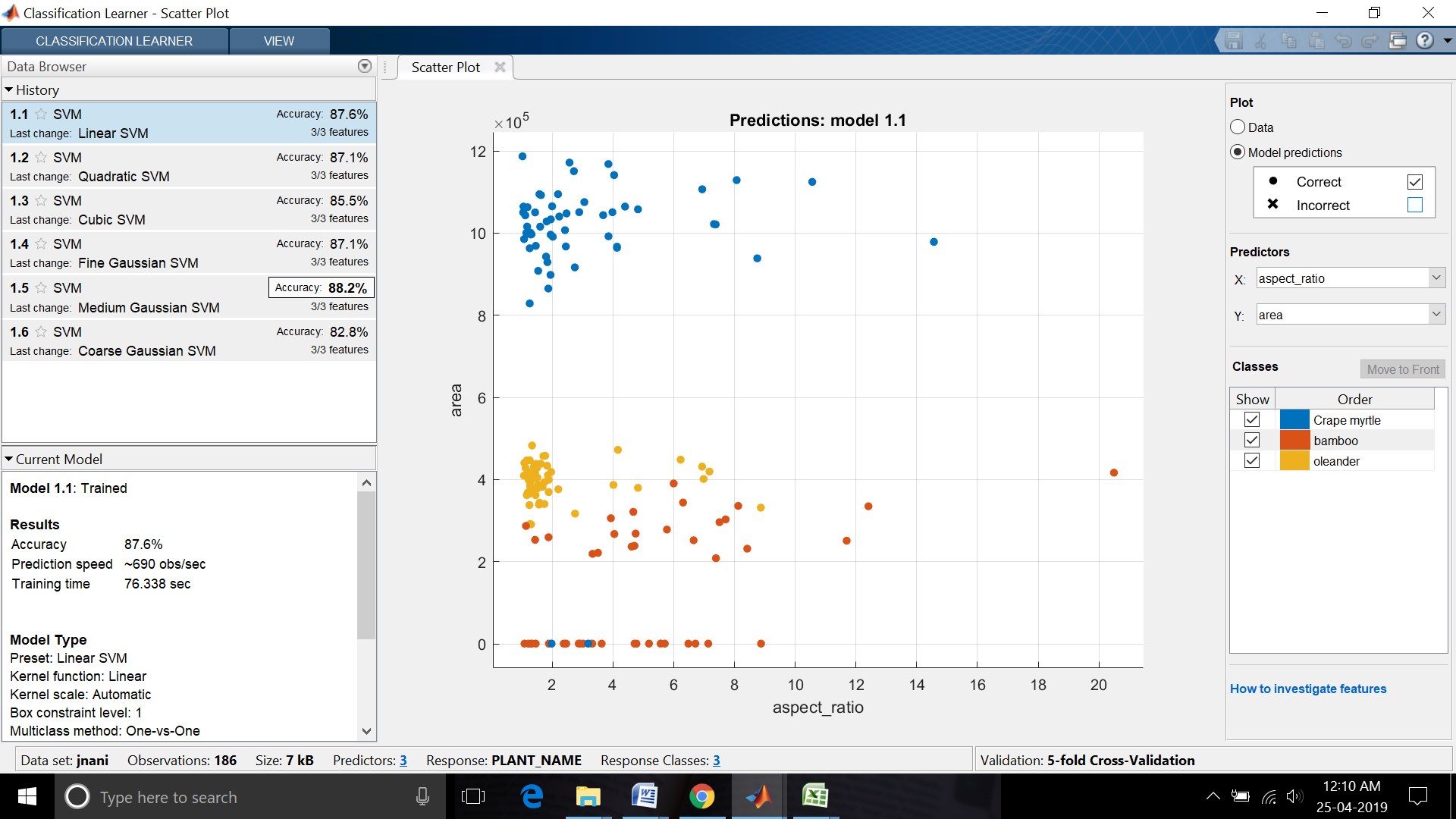


Fig 5. Scatter plot

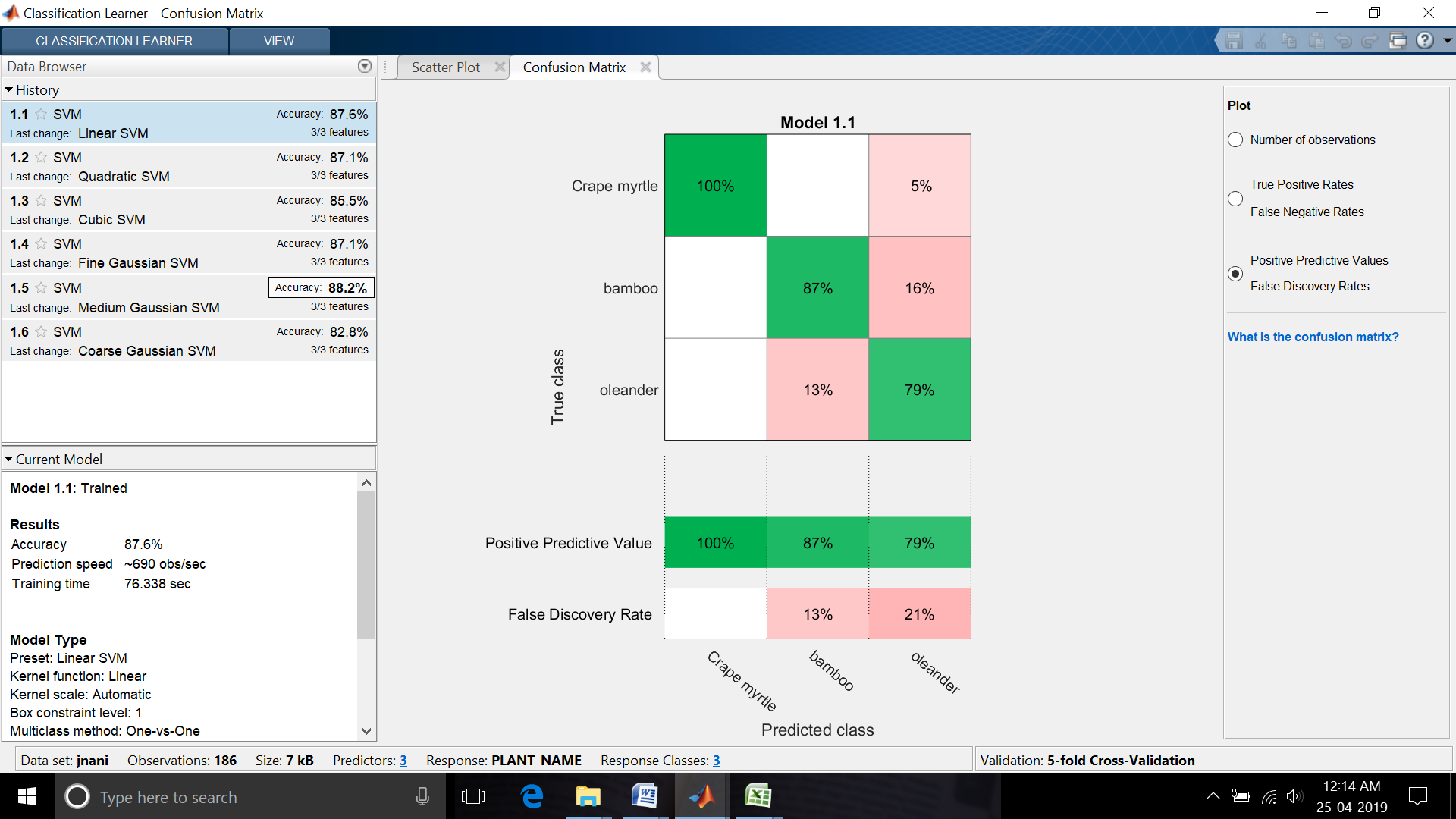


Fig 6. Confusion matrix

1. Give the user image information to the classifier function.
2. Finally, the classifier will predicit the name of the plant. And whether the plant exist or not in the pre-defined database.
3. ProjectStatus

In our project we have collected the different leaf samples of different shape, size and colour. And we have performed the thresholding and segmentation operations. Then we have coded the program to find axial ratio, relative intensity and area values of the leaves and stored these values in the excel file. And also we have trained the data using the classification learner, got the training accuracy of 87.6%.