NAME : SHARA RHAGHA WARDHAN B
ROLL NO: ME11B121

CLUSTERING

MATLAB CODE USED

```
clc;
clear all;
close all;
%storing all the images in one variable
A(:,:,1) = rgb2gray(imread('ashoka1.tiff'));
A(:,:,2) = rgb2gray(imread('ashoka2.tiff'));
A(:,:,3) = rgb2gray(imread('ashoka3.tiff'));
A(:,:,4) = rgb2gray(imread('ashoka4.tiff'));
A(:,:,5) = rgb2gray(imread('ashoka5.tiff'));
A(:,:,6) = rgb2gray(imread('banyan1.tiff'));
A(:,:,7) = rgb2gray(imread('banyan2.tiff'));
A(:,:,8) = rgb2gray(imread('banyan3.tiff'));
A(:,:,9) = rgb2gray(imread('banyan4.tiff'));
A(:,:,10) = rgb2gray(imread('banyan5.tiff'));
A(:,:,11) = rgb2gray(imread('banyan6.tiff'));
A(:,:,12) = rgb2gray(imread('mango1.tiff'));
A(:,:,13) = rgb2gray(imread('mango2.tiff'));
A(:,:,14) = rgb2gray(imread('mango3.tiff'));
A(:,:,15) = rgb2gray(imread('mango4.tiff'));
A(:,:,16) = rgb2gray(imread('mango5.tiff'));
A(:,:,17) = rgb2gray(imread('neem1.tiff'));
A(:,:,18) = rgb2gray(imread('neem2.tiff'));
A(:,:,19) = rgb2gray(imread('neem3.tiff'));
A(:,:,20) = rgb2gray(imread('neem4.tiff'));
A(:,:,21) = rgb2gray(imread('neem5.tiff'));
A(:,:,22) = rgb2gray(imread('neem6.tiff'));

%determining the feature parameters
for k=1:22
    B=im2bw(A(:,:,k),graythresh(A(:,:,k)));
    [n N]=size(B);
    %average length is taken as parameter 1
    l=0;
    for i=1:n
        C(i) = 0;
        for j=1:N
            if(B(i,j) == 0)
                C(i) = C(i) + 1;
            end
        end
        if(C(i)>0)
            l=l+1;
            cc(l) = C(i);
        end
    end
```

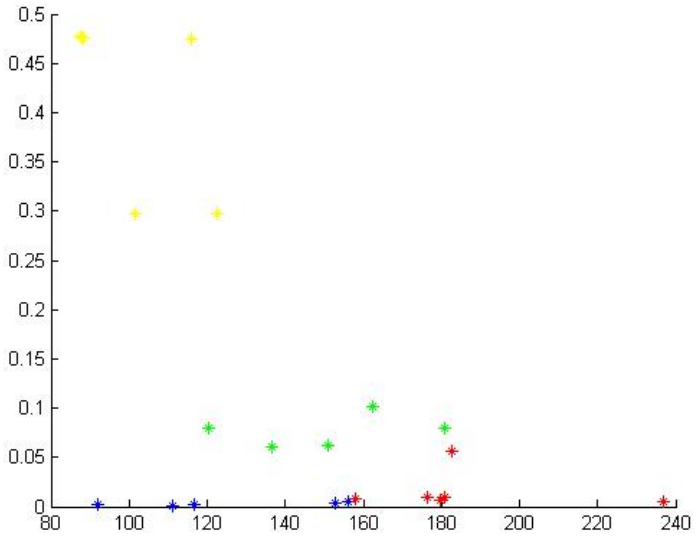
```
    param1 = mean(cc);
   %average breadth is taken as parameter 2
    m=0;
    for j=1:N
        D(j) = 0;
        for i=1:n
            if(B(i,j) == 0)
                D(j) = D(j) + 1;
            end
        end
        if(D(j)>0)
            m=m+1;
            dd(l) = D(j);
        end
    end
    param2 =  mean(dd);
    %parameter 3
    param3 = (sum(C))/max(max(C),max(D));
    %parameter 4
    param4 = sum(C)/sum(D);
    E(k,:) = [param1 param2 param3 param4];
end
%using k-means clustering
for p=1:1000
    IDX(p,:) = kmeans(E,4);
end
%finding the repititive solution as the optimal solution
rep=zeros(1000,1);
for i=1:1000
    for j=i:1000
        if(IDX(i,:) == IDX(j,:))
            rep(i) = rep(i)+1;
        end
    end
end
occur_max = max(rep);
for i=1:1000
    if(rep(i) == occur_max)
        IDX_opt = IDX(i,:)
    end
end
```

Three parameters are used as the features of the leaves

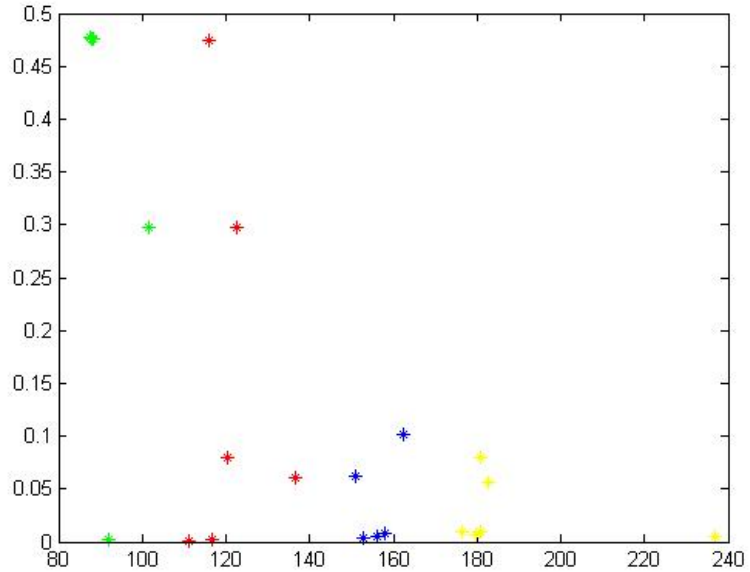
1. Mean of number of black pixels along the row where the pixels exist after using threshold to convert the gray scale image to binary image
2. Mean of number of black pixels along the column where the pixels exist after using threshold to convert the gray scale image to binary image
3. Total number of black pixels per length (longest length considered).
4. Length by breadth ratio

The new category is obtained by k-means clustering by iterating it a huge number of times and checking for repeatability of solution.

Using just the first two parameters to visualize a 2-D plot



TRUE PLOT



CLUSTERING PLOT

We can clearly see that the two parameters did not do that good the job

RESULTS OBTAINED WHEN ALL THE FOUR PARAMETERS WERE USED

LEAF	TRUE CATEGORY	CLUSTERED CATEGORY
ashoka leaf 1	3	3
ashoka leaf 2	3	3
ashoka leaf 3	3	4
ashoka leaf 4	3	2
ashoka leaf 5	3	2
banayan leaf 1	1	1
banayan leaf 2	1	1
banayan leaf 3	1	2
banayan leaf 4	1	1
banayan leaf 5	1	1
banayan leaf 6	1	1
mango leaf 1	2	2
mango leaf 2	2	2
mango leaf 3	2	3
mango leaf 4	2	2
mango leaf 5	2	2
neem leaf 1	4	4
neem leaf 2	4	4
neem leaf 3	4	4
neem leaf 4	4	4
neem leaf 5	4	4
neem leaf 6	4	4

Only 5 out of the 22 leaves were wrongly categorized as in 22% error exists in the model

Hence it is an acceptable estimate. But solution can be improved by tweaking the parameters such that the training set is satisfied and then use a test set to assess the quality of the parameters.