Plant Disease Detection Using K-means Clustering

Shraddha Shrivastav[a], Priyansha Mishra[b]*

[a]*Technical Associate, Genpact Headstrong Capital Markets, Noida, India*

[b]*Associate System Engineer, TATA Consultancy Services, Noida, India*

*mishrapriyansha@gmail.com, Shraddhashrivastav02@gmail.com

Received: *14.11.2018,* ***Accepted:*** *22.12.2018*

Abstract

The objective of this project is to identify the infected area in the plant leaf, to extract the interested region with the help of segmentation and to identify the diseases with the help of feature inspection. This project aims to design and evaluate software that uses image processing technique so that we can perform early detection of plant diseases. We present a software based solution that is faster, cheaper and much accurate.

Digital Image Processing uses computer algorithm to implement image processing while taking digital image as input. It gives us functionality to implement wider range of algorithms on the input data and avoids several problems like noise, signal distortion etc. It can model images into multidimensional systems. Image Segmentation is done in order to separate region of interest in the leaf so that we can analyze it easily. It gives the contrast in intensity between a pixel and its neighboring pixel in the complete image. A constant image exhibits a contrast value of zero.

Keywords- Segmentation, feature inspection, plant leaf disease, K-means clustering.

Introduction

Diseases found in plants contain symptoms which are primarily visible in leaves. Naked eye judgment of these diseases is difficult. The problem is even severe when the farmers rely on their experiences and use the extensive fertilizers which give a temporary treatment but it is possible that the disease is different from the one predicted or the fertilizer used was a harmful chemical ,breaching into the crops leading to future adverse effects to plant and animal life. It is necessary for a healthy crop that whenever a disease is encountered it is dealt with right diagnosis which can be ensured with a expert but since farmers cannot contact the experts every time a disease is spotted it becomes a decision solely dependent on the knowledge of the farmer. Meeting with experts is also an expensive activity, keeping in mind the financial status of the farmers. As there are no devices or technical mechanisms for the disease detection available with the farmers, they use naked eye observations to guess the diseases and use the techniques with which they are familiar. These techniques may or may not be absolutely the right treatment.

A contaminated plant life is worth causing a lot of destruction and devastations. The primary and direct loss is caused to the farmer which due to crop loss suffers huge debts and if consumption of a unhealthy crop occurs, it

causes adverse health hazards. The treatment of cropland without experts may cause wrong detection of disease. Even if it results in temporary solution to the crop loss, it may be an extensive and overuse of harmful fertilizers. These fertilizers cause even more severe and fatal diseases in the future. Expert advices not only correct remedy but also environment friendly methods of curing crop loss. It is not right to always expose the fields to the chemicals where there are new advances made in the field of agriculture every day. The unawareness of the farmers is a hindrance to these advances and there is a need of a much easier manner for the farmers to contact the experts.
Today when there are automated technologies in all the fields, there can be an automated technological advancement in the disease detection of the crop.

Literature review and related work

The proposed algorithm starts working by creating a device independent color space. A color space as the name suggest is the special orientation of colors. Color profiling is the description of the parameters of the device capturing source image (Dhaygude *et al.,* 2013). Digital image processing is advance as it allows production of various representation of color space which can be reproduced whenever needed. It can be chosen as some labels mapped to physical colors or it can be randomly chosen or it can be mapped according to some standards like RGB (Arivazhagan *et al.,* 2013).

Next we apply device independent color space transformation and convert the color according to the color specified in the color space transformation structure. In the color transformation structure various parameter for transformation are provided. A device dependent color system like RGB system is not used because the RGB values altered as brightness and contrast will depend on device capturing the image. The color system that we use must be device independent so as to improve the precision and accuracy of the disease detection. The K means clustering comes into picture.

K means clustering algorithm tries to divide the objects (pixels) into K number of classes on the basis of features. In our approach we divide the leaf image into four clusters or segments with the help of K means clustering where those segments containing diseases are separated from those free of any disease. If more than one segment contains diseases, the leaf is said to have more than one disease. The best outcome of the algorithm is when leaf is classified for 3-4 segments (Amoda *et al.,* 2014).

Image acquisition in image processing can be broadly defined the action of retrieving an image from source (Arivazhagan *et al.,* 2013).

Preprocessing images means if there are any disturbances in the image like reflections of particles on other particles, are removed. The intensity of the particular particle (pixel) and low frequency noise in the background of the image are also removed. Certain portions of the image are masked if required. Preprocessing as the name suggests is for enhancing the quality of the image before we start with actual computations (Kulkarni *et al.,* 2012).

Image Segmentation means we segment the image into multiple segments or pixels or super pixels so that it becomes easier for the analyzer to grasp more knowledge from that piece of segment. We can classify the objects and shapes, their boundary and curves from those segmented pixels. This is done by assigning labels to pixels. Pixels sharing same label also share same characteristics.

Feature Extraction means we generate data after applying image processing techniques and train the system so that it becomes self capable to analyze the disease by correlating the features of inspecting image with trained data.
Statistical is a part of data analytics. It includes collecting and searching all the information in the features extracted data. Conclusion of the disease found can be generated with some probability range **(**Amoda *et al.*, 2014**).**

Methodology

The proposed approach step by step of the image segmentation and recognition processes is illustrated in Algorithm Algorithm: Basic steps describing the proposed algorithm.

1. First step is Image acquisition
2. Second step is Preprocessing of images
3. Third step is Masking green-pixels
4. Fourth step is to remove the masked cells inside the boundaries of the infected clusters.
5. Next step is to apply K-means clustering
6. Next and important step is Image Segmentation and to obtain the useful segments
7. Final step is Features Extraction

In our project, the K-means clustering is set to use squared Euclidean distances.
The next step in algorithm is the feature extraction.

In this paper color co-occurrence method has been used for feature extraction (Badnakhe *et al.*, 2011).

Euclidean distance metric *has been used f*or K means clustering and *K*-means minimizes within-cluster point scatter:

$$W(C)=\frac{1}{2}\sum_{k=1}^{K}\sum_{C(i)=k}\sum_{C(j)=k}\left\|x_i-x_j\right\|^2=\sum_{k=1}^{K}N_k\sum_{C(i)=k}\left\|x_i-m_k\right\|^2$$

Where

m_k is the mean vector of the k^{th} cluster
N_k is the number of observations in k^{th} cluster
$C(i)$ denotes cluster number for the i^{th} observation
$x_1,\ldots, x_N$ are data points

A software program was written in MATLAB that would take in .mat files representing the training and testing data, train the classifier using the 'train files', and then use the 'test file' to perform the classification task on the test data. In the experimental results, the threshold value for each of the above categories is constant for all samples infected with the same disease. We have decided the threshold value by taking the various samples of leaves of specific plants.

The Matlab code provided is able to successfully detect the disease present in the plant leaf with the use of feature inspection technique. The respective preventive measures for the same diseases will also appear along with the result.

High level module structure & functional specifications of each module in details.

The basic module structure of the proposed image processing- based disease detection solution

Functional specifications of each module:

Image acquisition in image processing can be broadly defined the action of retrieving an image from source.
Preprocessing images means if there are any disturbances in the image like reflections of particles on other particles, are removed. The intensity of the particular particle (pixel) and low frequency noise in the background of the image are also removed. Certain portions of the image are masked if required. Preprocessing as the name suggests is for enhancing the quality of the image before we start with actual computations.

Image Segmentation means we segment the image into multiple segments or pixels or super pixels so that it becomes easier for the analyzer to grasp more knowledge from that piece of segment. We can classify the objects and shapes, their boundary and curves from those segmented pixels. This is done by assigning labels to pixels. Pixels sharing same label also share same characteristics.

Feature Extraction means we generate data after applying image processing techniques and train the system so that it becomes self capable to analyze the disease by correlating the features of inspecting image with trained data.
Statistical analysis is a part of data analytics. It includes collecting and searching all the information in the features extracted data. Conclusion of the disease found can be generated with some probability range.

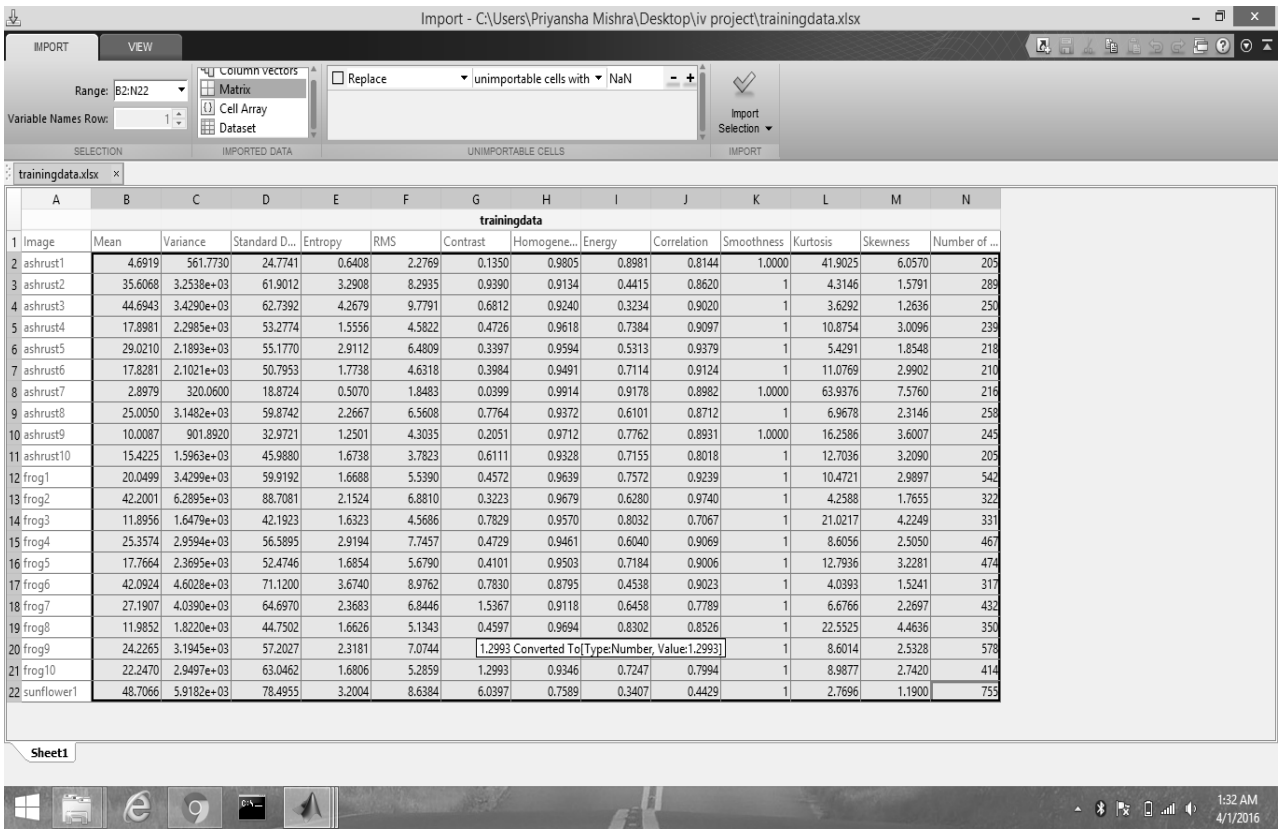

	Image	Mean	Variance	Standard D...	Entropy	RMS	Contrast	Homogene...	Energy	Correlation	Smoothness	Kurtosis	Skewness	Number of ...
2	ashrust1	4.6919	561.7730	24.7741	0.6408	2.2769	0.1350	0.9805	0.8981	0.8144	1.0000	41.9025	6.0570	205
3	ashrust2	35.6068	3.2538e+03	61.9012	3.2908	8.2935	0.9390	0.9134	0.4415	0.8620	1	4.3146	1.5791	289
4	ashrust3	44.6943	3.4290e+03	62.7392	4.2679	9.7791	0.6812	0.9240	0.3234	0.9020	1	3.6292	1.2636	250
5	ashrust4	17.8981	2.2985e+03	53.2774	1.5556	4.5822	0.4726	0.9618	0.7384	0.9097	1	10.8754	3.0096	239
6	ashrust5	29.0210	2.1893e+03	55.1770	2.9112	6.4809	0.3397	0.9594	0.5313	0.9379	1	5.4291	1.8548	218
7	ashrust6	17.8281	2.1021e+03	50.7953	1.7738	4.6318	0.3984	0.9491	0.7114	0.9124	1	11.0769	2.9902	210
8	ashrust7	2.8979	320.0600	18.8724	0.5070	1.8483	0.0399	0.9914	0.9178	0.8982	1.0000	63.9376	7.5760	216
9	ashrust8	25.0050	3.1482e+03	59.8742	2.2667	6.5608	0.7764	0.9372	0.6101	0.8712	1	6.9678	2.3146	258
10	ashrust9	10.0087	901.8920	32.9721	1.2501	4.3035	0.2051	0.9712	0.7762	0.8931	1.0000	16.2586	3.6007	245
11	ashrust10	15.4225	1.5963e+03	45.9880	1.6738	3.7823	0.6111	0.9328	0.7155	0.8018	1	12.7036	3.2090	205
12	frog1	20.0499	3.4299e+03	59.9192	1.6688	5.5390	0.4572	0.9639	0.7572	0.9239	1	10.4721	2.9897	542
13	frog2	42.2001	6.2895e+03	88.7081	2.1524	6.8810	0.3223	0.9679	0.6280	0.9740	1	4.2588	1.7655	322
14	frog3	11.8956	1.6479e+03	42.1923	1.6323	4.5686	0.7829	0.9570	0.8032	0.7067	1	21.0217	4.2249	331
15	frog4	25.3574	2.9594e+03	56.5895	2.9194	7.7457	0.4729	0.9461	0.6040	0.9069	1	8.6056	2.5050	467
16	frog5	17.7664	2.3695e+03	52.4746	1.6854	5.6790	0.4101	0.9503	0.7184	0.9006	1	12.7936	3.2281	474
17	frog6	42.0924	4.6028e+03	71.1200	3.6740	8.9762	0.7830	0.8795	0.4538	0.9023	1	4.0393	1.5241	317
18	frog7	27.1907	4.0390e+03	64.6970	2.3683	6.8446	1.5367	0.9118	0.6458	0.7789	1	6.6766	2.2697	432
19	frog8	11.9852	1.8220e+03	44.7502	1.6626	5.1343	0.4597	0.9694	0.8302	0.8526	1	22.5525	4.4636	350
20	frog9	24.2265	3.1945e+03	57.2027	2.3181	7.0744	1.2993 Converted To[Type:Number, Value:1.2993]				1	8.6014	2.5328	578
21	frog10	22.2470	2.9497e+03	63.0462	1.6806	5.2859	1.2993	0.9346	0.7247	0.7994	1	8.9877	2.7420	414
22	sunflower1	48.7066	5.9182e+03	78.4955	3.2004	8.6384	6.0397	0.7589	0.3407	0.4429	1	2.7696	1.1900	755

Figure 1: Database used to classify the diseases

Results & Discussion

Symptoms of frogeye leaf spot are most visible and typically seen on leaves, but can also occur on stems, pods, and seeds. Ash rust (A fungal disease) is responsible for the appearance of swollen yellow to orange spots on upper leaf surfaces during summer. For any disease image we calculate three segmented images. As depicted below is an example of a source image of ash rust disease and its three segmented images. We then chose one of the segmentation suitable for disease detection.

We then compute the features like mean, variance, standard deviation, entropy to match the feature of a disease and detect the disease.

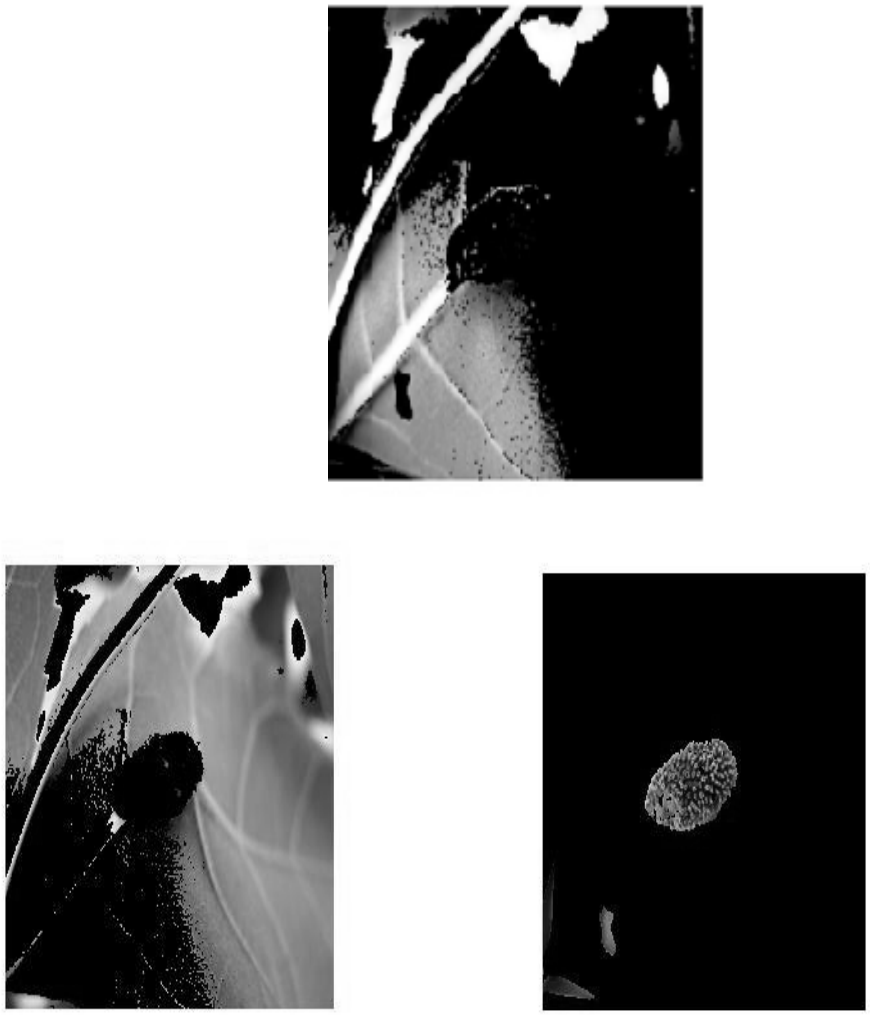

Figure 2: Ash rust three segmentation images

Conclusion and future work

We would like to conclude that it is an efficient plan where plant diseases are detected and preventive measures are provided as well. The Feature based matching technique used by us is effectively detecting the disease found in leaf. Further work can be extended by increasing the database of sample images and involvement of more advance features of color extraction which can lead to better result. More training samples are needed in order to cover more cases and to predict the disease more accurately

References

Amoda, N., Jadhav, B., Naikwadi S. 2014. Detection and classification of plant diseases by image processing. *International Journal of Innovative Science, Engineering & Technology,* 1(2), 70-74.

Arivazhagan, S., Shebiah, R., Newlin, A. S., Varthini S.V. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agriculture Engineering International: CIGR Journal*, 15(1), 211-217.

Badnakhe, M. R., Deshmukh, P. R. 2011. An Application of K-Means Clustering and Artificial Intelligence in Pattern Recognition for Crop Diseases. *International Conference on Advancements in Information Technology*, 20, 134-138.

Dhaygude, S. B., Kumbhar, N. P. 2013. Agricultural plant Leaf Disease Detection Using Image Processing. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2(1), 599-602.

Kulkarni, A. H., Patil, A., K. R. 2012. Applying image processing technique to detect plant diseases. *International Journal of Modern Engineering Research*, 2(5), 3661-3664.