Diseases Detection of Various Plant Leaf Using Exploratory Data Analysis and Pytorch

Manvi Ahuja[1] and Nishtha[2]
[1]Department of Botany (Biosciences),Branch Botany , UIBT , Chandigarh University NH-95, Ludhiana – Chandigarh state Highway , Punjab 140413, India.
manviahuja1997@gmail.com

Abstract

Livelihhood is the achieved by major source of agriculture. In India employment can also be provided by Agricultural sources on larger scale. Developing countries are providing the employment as crucial pandemic time is going on so that many prople are getting benefitted by agricultural means. Developing countries are having major source like 70% people are dependent on agricultural sources for their survival[1]. Due to non availability of technical education they are dependent on the cultivation of various variety of crops and vegetables. They are using the tradition ancient methods of cultivation. For better cultivation they are required to provide the knowledge of latest tools and techniques. They do not have enough guidance what climate and other means are required to get the maximum yield. They need to be provided sessions to provide them awareness about the lates tools and which crop can be cultivated in which season and what are the factors responsible so that maximum yield can be achieved. When plants are affected by many diseases then cultivation is affected a lot. Chemical spraying effects the human health directly which is used by majority of farmers for disease protection. Identification of forthcoming diseased plants is also a big challenge for farmers. Detection of diseased plants are done by latest techniques. We have done the surveys for the identification of the diseased plants data.

Keywords :- Pesticidal chemicals, crops, yield, seeds, harvesting.

I INTRODUCTION:

II RELATED WORK

Data Set Description

The data holds the images of apple - Plant leaf with healthy and infected conditions

Files train.csv - the training set metadata.
image - the image ID.
labels - the target classes, a space delimited list of all diseases found in the image. Unhealthy leaves with too many diseases to classify visually will have the complex class, and may also have a subset of the diseases identified.
sample_submission.csv - A sample submission file in the correct format.

1. image
2. labels

train_images - Set Images of Training Data

test_images - The test set images. This competition has a hidden test set: only three images are provided here as samples while the remaining 5,000 images will be available to your notebook once it is submitted.
Table 1 shows the data about different category of leafy patches.we have collected the data for the healthy and other types of patches occurred on the leaves like eye_leaf, Spot etc[2].

image	labels
800113bb65efe69e.jpg	healthy
8002cb321f8bfcdf.jpg	scab frog_eye_leaf_spot complex
80070f7fb5e2ccaa.jpg	scab
80077517781fb94f.jpg	scab
800cbf0ff87721f8.jpg	complex

Submission data
Image & labels

We have used the following labels while our implementation of disease detection.

Category of Labels:-

Healthy

Complex One

frog_eye_leaf_spot

complex frog_eye_leaf_spot

mildew_ powdery

complex

powdery_mildew

rust

Rust Complexity

frog_eye_leaf_spot rust

scab

frog_eye_leaf_spot scab

frog_eye_leaf_spot complex scab

scab	4826
healthy	4624
frog_eye_leaf_spot	3181
rust	1860
complex	1602
powdery_mildew	1184
scab frog_eye_leaf_spot	686
scab frog_eye_leaf_spot complex	200
frog_eye_leaf_spot complex	165
rust frog_eye_leaf_spot	120
rust complex	97
powdery_mildew complex	87

Name: labels, dtype: int64

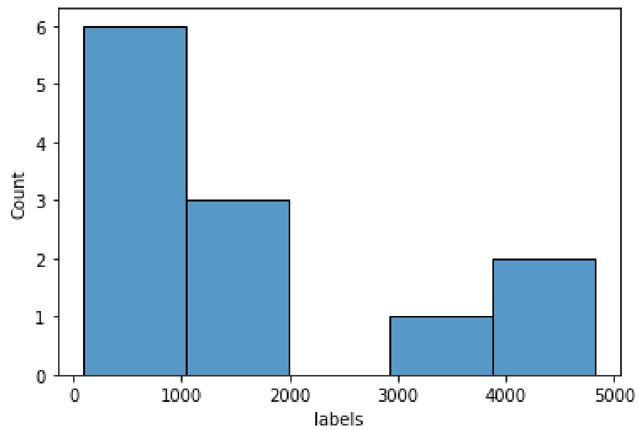


Figure 1 Labelled data

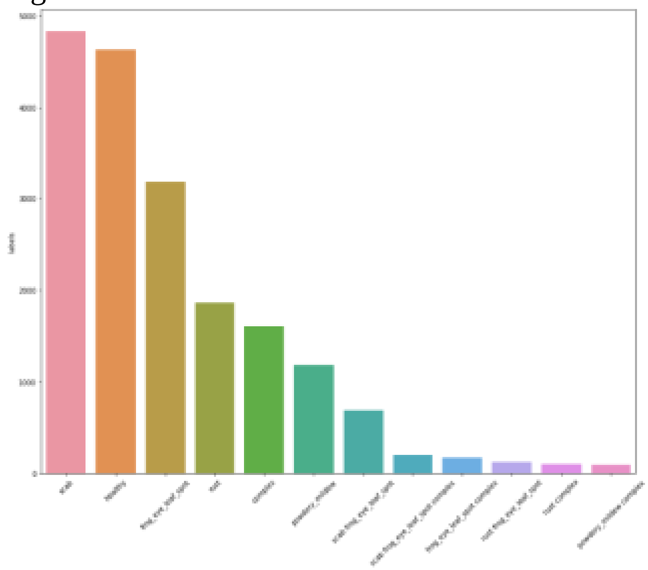


Figure 2 RGB graph of labelled data

- Dataset is pretty unbalanced as per above pie chart
- Need to chose the appropriate sampling strategy to sort out this issue

Figure 3 Pathological Images of Plant Leaves

We have plotted the few images in the training data above (the RGB values can be seen by hovering over the image)[2][9]. The green parts of the image have very low blue values, but by contrast, the brown parts have high blue values. This suggests that green (healthy) parts of the image have low blue values, whereas unhealthy parts are more likely to have high blue values. This might suggest that the blue channel may be the key to detecting diseases in plants[7].

	image	labels
0	800113bb65efe69e.jpg	healthy
1	8002cb321f8bfcdf.jpg	scab frog_eye_leaf_spot complex
2	80070f7fb5e2ccaa.jpg	scab
3	80077517781fb94f.jpg	scab
4	800cbf0ff87721f8.jpg	complex
...	...	...
18627	fffb900a92289a33.jpg	healthy
18628	fffc488fa4c0e80c.jpg	scab
18629	fffc94e092a59086.jpg	rust
18630	fffe105cf6808292.jpg	scab frog_eye_leaf_spot

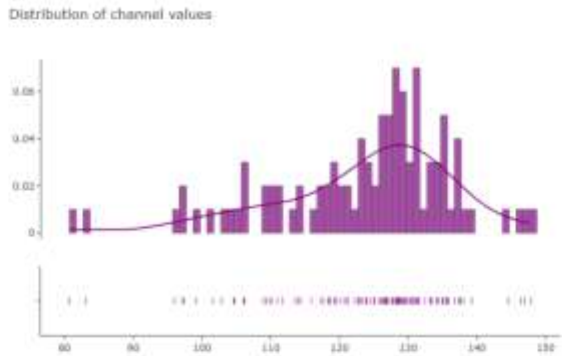


Figure4 Distribution of Channel Values

Histograms are a graphical representation showing how frequently various color values occur in the image i.e frequency of pixels intensity values. In a RGB color space, pixel values range from 0 to 255 where 0 stands for black and 255 stands for white. Analysis of a histogram can help us understand thee brightness, contrast and intensity distribution of an image. Now let's look at the histogram of a random selected sample from each category.

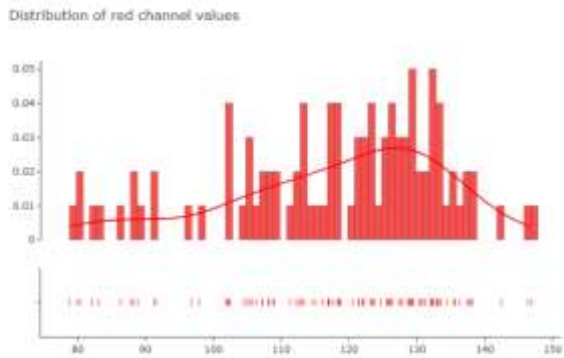


Figure 5 Distribution of Red Channel Values

The red channel values seem to roughly normal distribution, but with a slight leftward (Negative skew). This indicates that the red channel tends to be more concentrated at higher values, at around 100[10]. There is large variation in average red values across images.

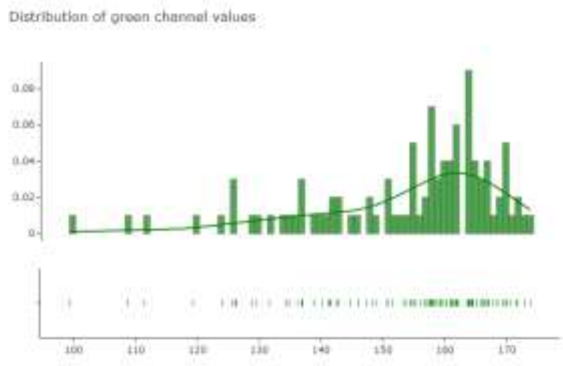


Figure 6 Distribution of Green Channel Values

The green channel values have a more uniform distribution than the red channel values but its right skewed, with a smaller peak. The distribution also has a right skew (in contrast to red) and a larger mode of around 160. This indicates that green is more pronounced in these images than red, which makes sense, because these are images of leaves!

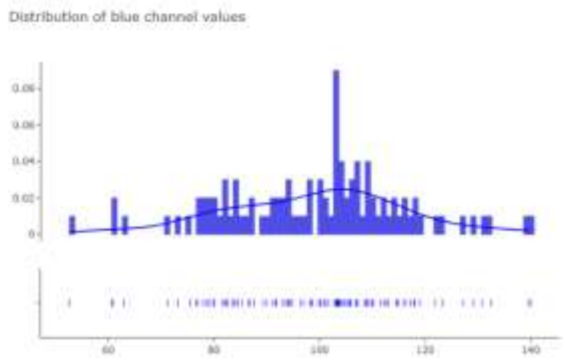


Figure 7 Distribution of Blue Channel Values

The blue channel has the most uniform distribution out of the three color channels, with minimal skew (slight leftward skew)[17]. The blue channel shows great variation across images in the dataset.

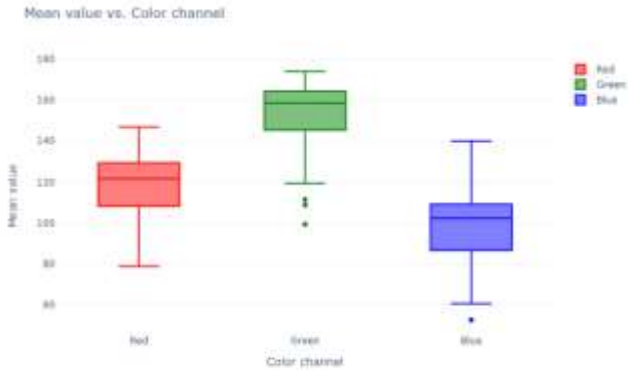


Figure 8 Distribution of all RGB Channel Values

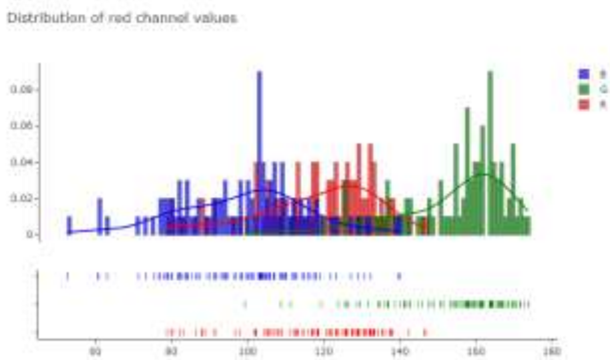


Figure 9 Histrogram of all RGB Channel Values

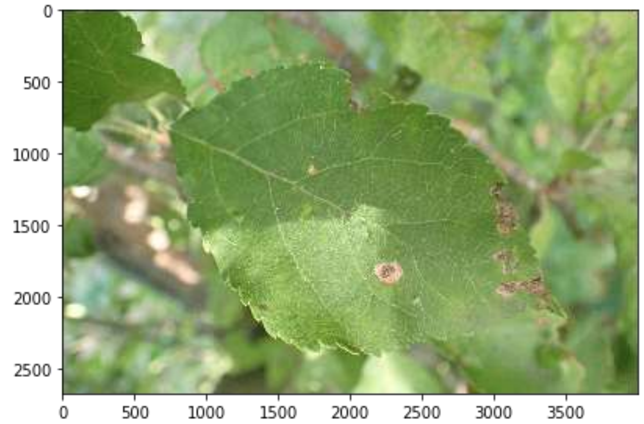


Figure 10 Diseased Leaf impression

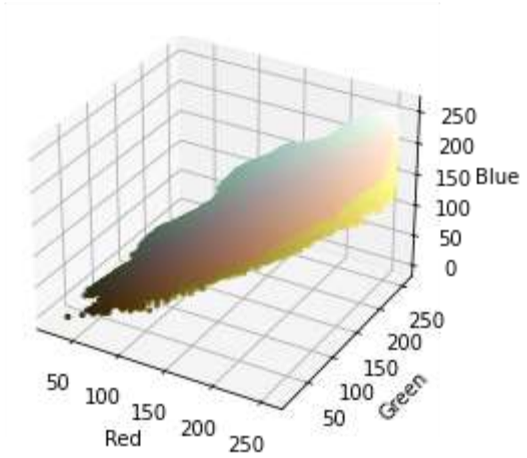


Figure 11 3D scatter plot for the image in RGB

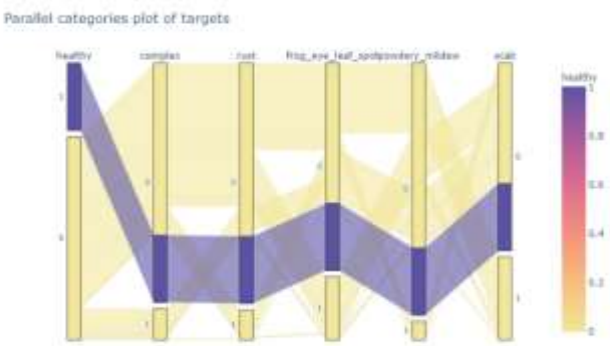


Figure 12 Parallel categories of plot data

In the above plot, we can see the relationship between all 6 categories. As expected, it is impossible for a healthy leaf to have scab, rust, or multiple diseases. Also, every unhealthy leaf has one of either scab, rust, or multiple diseases. The frequency of each combination can be seen by hovering over the plot[18][19].

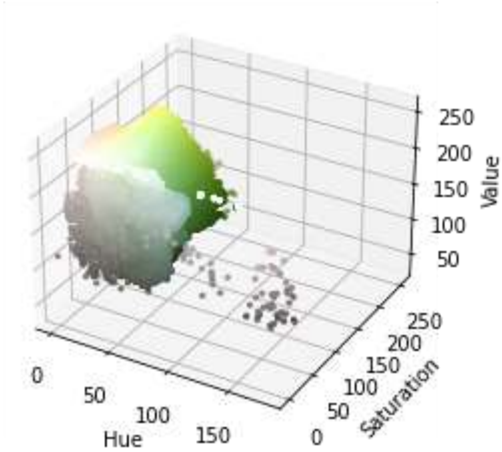
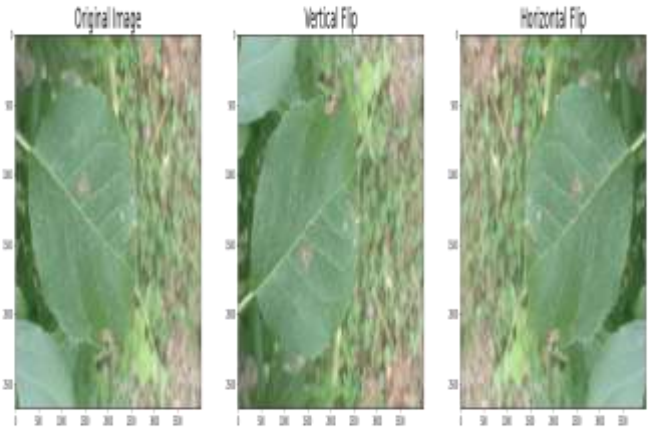


Figure 13 3D scatter plot for the image in HSV



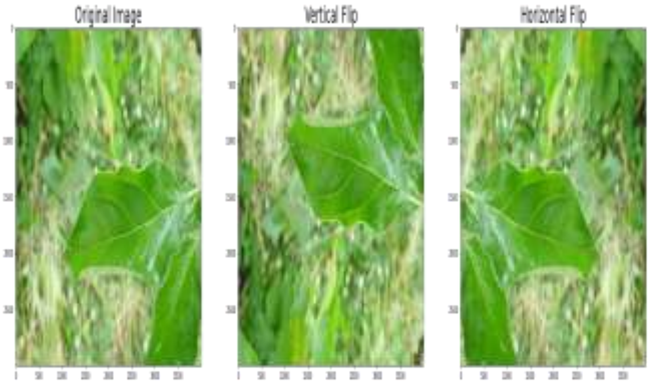


Figure 14 Leafy images of different patterns

The second column of images above contains the Canny edges and the third column contains cropped images. I have taken the Canny edges and used it to predict a bounding box in which the actual leaf is contained. The most extreme edges at the four corners of the image are the vertices of the bounding box. This red box is likely to contain most of if not all of the leaf. These edges and bounding boxes can be used to build more accurate models[20].

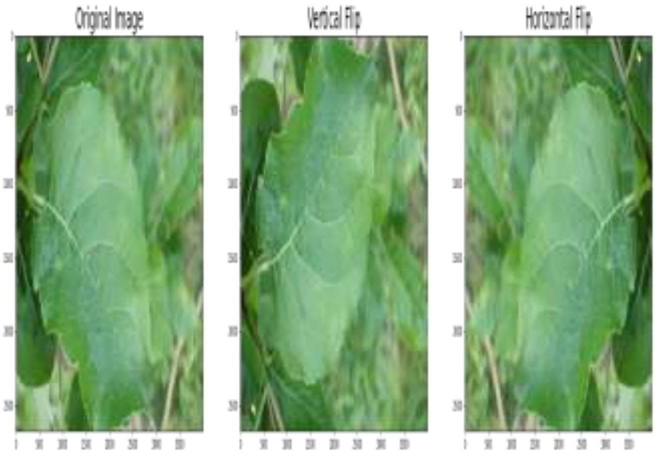
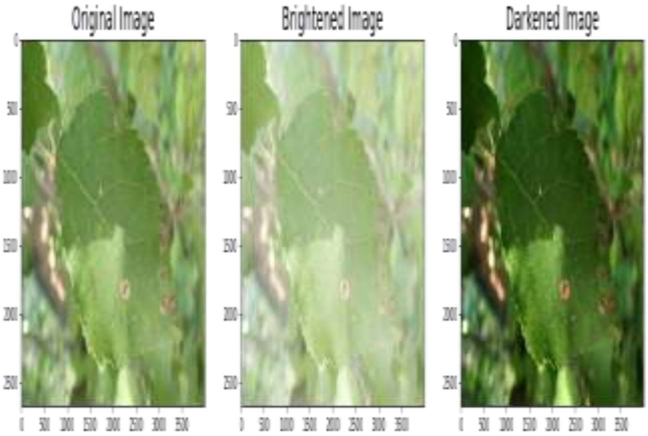


Figure 15 Leafy images of different patterns

Convolution is a rather simple algorithm which involves a kernel (a 2D matrix) which moves over the entire image, calculating dot products with each window along the way. The GIF below demonstrates convolution in action[18].



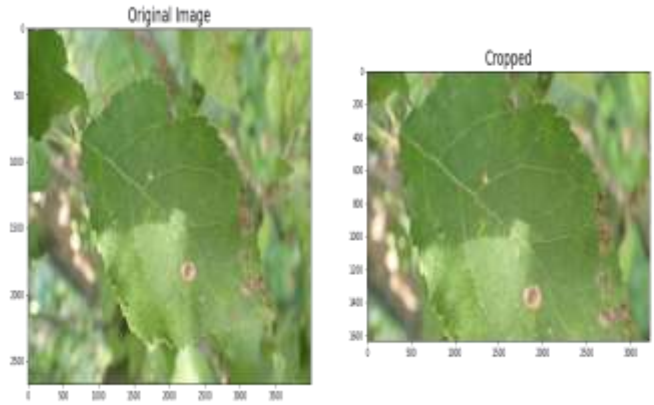
The convolution operator seems to have an apparent "sunshine" effect of the images. This may also serve the purpose of augmenting the data, thus helping to build more robust and accurate models.

Blurring

Blurring is simply the addition of noise to the image, resulting in a less-clear image. The noise can be sampled from any distribution of choice, as long as the main content in the image does not become invisible. Only the minor details get obfuscated due to blurring. The blurring transformation can be represented using the equation below[17[20].

$$A_{ijk} = A_{ijk} + \mathcal{N}(0, 0.1)$$

The example uses a Gaussian distribution with mean 0 and variance 0.1. Below I demonstrate the effect of blurring on a few leaf images:



CONCLUSION AND SCOPE

Our presented work has the detailed explanation of diseased leafy plants with different planes like convoluted, skewed, dim, brightened etc. We have very well mentioned the importance of identification of diseased plants so that maximum yield can be produced for the mankind. Impact of the above said is clearly illustrated in he labels part and patterns of the different diseased leaves are also clearly shown in our research work. We have suggested to follow the pathological approach for broader prospective approach for better cultivation. Paper depicts the deep learning methods Pytorch using the explanatory data analysis to understand the data part very well. We have used the deep learning keras and tensor flow libraries for implementation of Explanatory Data analysis approach.

References:

[1]. "India agriculture economy". Available :http://statistics times.com/ economy/ sectorwise-gdp-contribution-of India.

[2 "PlantVillage", Plantvillage.psu.edu, 2020. [Online]. Available: https://plantvillage.psu.edu/. [Accessed: 31- Jan- 2020].

[3]. "iNaturalist", iNaturalist, 2020. [Online]. Available: https://www.inaturalist.org/. [Accessed: 30- Jan- 2020].

[4] I. Ltd., "Success-stories - PlantSnap: Training the world's largest plant recognition classifier. | Imagga Technologies Ltd.", Imagga.com, 2020. [Online]. Available: https://imagga.com/success-stories/plantsnap-case-study. [Accessed: 31- Jan- 2020].

[5]. "Smartphone users worldwide 2020 | Statista", Statista, 2020. [Online]. Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. [Accessed: 22- Apr- 2020]

[6] Prof.Sonal, P.Patil, Rupali, Zambre,"classification of cotton leaf spot disease using SVM," intetnational journal of engineering research and applications Vol.4,pp92-97, May 2015

[7 K. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis, "Machine Learning in Agriculture: A Review", Sensors, vol. 18, no. 8, p. 2674, 2018.

[8] Pragya adhikari , yeonyee Oh, dilipR. Panthee"current status of early blight resistance in tomato: An update, international journal of molecular scuence , september 2017

[9].Akansha pandey, Sanjeev Dubey , "evaluations of brinjal germplasm for resistance to fusarium wilt disease", international journal of scientific and rrsearch publications, volume 7 , issue 7 , july 2017

[10]. Gittaly Dhingra , Vinay kumar , Hem Dutt joshi study of digital image processing techniques for leaf disease detection and classification , springer - science , 29 November 2017

[11]. Shitala Prasas , sateesh K. Peddujo,Debashish Ghost ," multi - resolution mobile vision system for plant leaf disease diagnosis" , pp.379-388, springer-verlag london 2015

[12]. Shanwen Zhang, Zhuhong you , Xiaowei Wu, "plant disease leaf image segmentation based on superpixel clustering and EM algorithm" , springer , june 2017

[13]. Keyvan Asefpour Vakilian and Jafar Massah , "An artificial neural network approach to identify fungal diseases of cucumber (cucumis sativus L.). Plants using digital image processing" , Vol.46, No.13, 1580-1588 Taylor and Francis , 2013

[14]. Mohammed Brahimi, Kamel Boukhalfa and Abdelouahab Moussaoui ," Deep Learning for Tomato Disease : Classfication and symptoms Visualization", Vol.31, no.4, 299-315, Taylor and Francis,2017
[15]. H.Al-Hairy , S. Bani- Ahmad, M. Reyalat , M.Braik and Z.AIRahammeh," Fast and Accurate Detection nd classification of plant Diseases, international journal of computer applications" , Vol.17,No.1,pp.31-38.March2011
[16]. Yuanyuan Shao, Guantao Xuan , Yangyan Zhu,Yanling Zhang, Hongxing Peng , Zhongzheng Liu and Jialin Hou , Redearch on automatic identification systen of tobacco diseases, vol.65, no.4, 252-259, Taylor and Francus , 2017
[17]. Vijai Singh, A.K. Misra ," detection of plant leaf diseases using image segmentation and soft computing techniques" , Information Processing In Agriculture 4 (2017) 41-49, science direct , 2017
[18]. Shanwen Zhang , Xiaowei Wuc, Zhuhong You, Liquing Zhang, leaf image based cucumber disease recognition using sparse representation classification , computers and electronics in agriculture 135-141 , science direcr ,2017
[19]. Amar Kumar Dey, Manisha Sharma , M.R.Meshram," image processing Based leaf Rot Disease; Detection of Betel Vine(piper Betlel)", Procedia computer science 748-754 , Science direct , 2016
[20]. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, DubravkoCulibrk and Darko Stefanovic," Deep Neural Netwroks Based Recognition of plant Diseases by Leaf Image classification ", Hindawi publishing corporation computational intelligence and neuroscience , vol 2016, Article ID 3289801, 11 pages
[21]. Manisha Bhange, H.A. Hingoliwala," Smart Farming : Pomegranate disease detection using image processing , procedia computer science 280-288, science direct ,2015