ESTIMATION OF CHLOROPHYLL CONTENT OF A PLANT USING FLUROSCENT IMAGING

MEGHANA KURUPALLI

Abstract: Chlorophyll content in leaves is related to leaf nitrogen and yield. Hence, estimating the amount of chlorophyll content in the plants can be an indirect measure for estimating the yield of a plant. The aim of this project is to determine how cotton plants react when subjected to drought stressed conditions, and to observe the differences between controlled and drought stressed plants. This can give us an idea of how cotton plants react to drought stress. The experiment is done with the fluorescent images of 10 different plants, all of the same genotype and data analysis is done to differentiate between stressed and non-stressed plants.

1 INTRODUCTION:

Ever increasing global food demand has put enormous pressure on humankind to look for ways to improve crop productivity. Due to the changing climatic conditions over the years, and the global warming, there might be a deficiency in the availability of water. According to a news article by BBC*, Water demand globally is projected to increase by 55% between 2000 and 2050. Much of the demand is driven by agriculture, which accounts for 70% of global freshwater use, and food production will need to grow by 69% by 2035 to feed the growing population. Hence, the world is running out of resources and we can only feed the exponentially growing population by finding methods which can offer a good trade-off between the use of resources and the yield. One of the most promising and emerging methodology in today's world is to use the power of computers in the field of agriculture to increase the yield by introducing the methods like high throughput plant phenotyping[3]. Estimating the yield of a plant through the observable properties of a plant is a lot helpful and faster than manually doing the job.

One of the major factors that determine plant health is the chlorophyll content of the plant. The principle underlying chlorophyll detection is that light absorbed by chlorophyll molecules in a leaf can undergo one of the following three processes. It can be used for photosynthesis, excess energy can be dissipated as heat, or it can be reemitted as light[1]. This is where the detection of chlorophyll by fluorescence comes into play. The above processes occur in competition, such that increase of one factor reduces the other two. Hence, by measuring the yield of chlorophyll fluorescence, information regarding the photosynthesis can be calculated, which in turn, indicates the health of the crop.

2 OBJECTIVES:

The objectives of this project are:

1. Plot the histograms of the fluorescent images of drought stressed plants, and classify the pixels into different ranges. This relatively gives the amount of chlorophyll at different areas of leaves.

2. Use K-means clustering algorithm to classify how different genotypes of plants respond to drought conditions.

3. Use Support Vector Machines classifier to classify a plant into either high or low stressed.

3 RELATED WORK:

FLUROSCENCE IMAGING

Chlorophyll fluorescence is a rapid, non-destructive technique that has been used successfully in the evaluation of plant photosynthetic activity. Chlorophyll Fluorescent Imaging has a useful potential to detect stresses before visual symptoms appear, which is ideal in screening of genotypes for the early identification of those with high tolerance to biotic and abiotic stress [2]. Hence, the fluorescent images are used I this project to estimate the variation of chlorophyll for stressed cotton plants.

A lot of work is done using the image analysis of fluorescent images in order to select the plants with high tolerance to pathogens. In this project, Image segmentation is used to divide the digital image into multiple sets of pixels to select the areas of interest, which is similar to dividing the different levels of chlorophyll in leaves.

In the earlier work done on this problem, the output is classified only into two classes, high content of chlorophyll and low content of chlorophyll. There was no chance for relatively measuring the amount of chlorophyll. What is intended to do in this project is to define a range of relative values for the amount of chlorophyll present in a part of leaf, rather than saying if it is high or low.

There are 2 types of clustering algorithms, supervised and un-supervised. K-means clustering comes under the catagoery of supervised machine learning algorithm. This project used K-Means algorithm because supervised learning is the best fit for this project[4]. We already have the data regarding which plants are stressed, and which plants are not. Hence, it is more suitable to use a supervised machine learning algorithm than an un-supervised technique.

SVM is also chosen for similar reason. It is robust, efficient and very suitable for high-dimensional data.

4 PROBLEM DEFINITION:

The aim of this project is to provide more accurate measure of chlorophyll content of the plant leaves. This can be helpful in determining when the plant is getting affected by the drought conditions more precisely. The input of this project is 20 cotton plants, out of which, 10 plants are under controlled conditions and 10 plants are drought stressed. The output of this project is the leaves showing different chlorophyll contents and also a classification model, which classifies a plant as either stressed or non-stressed.

5 DATASET

The image data sets used for this project are visible and fluorescent images of 20 cotton plants. Two experiments are conducted on 10 plants each. In the experiment 1, 5 plants are under controlled conditions and 5 are stressed. For the stressed plants in experiment-1, stress is induced on day-13 and lasted for 8 days. For the stressed plants in experiment-2, stress is induced on day-20, and lasted for 9 days. The images are captured in five side views (SV_0, SV_72. SV_144, SV_216, SV_288) and one top view. Hence, the total number of images is 2*(6*20*30), i.e., 7200 images approximately. All the images were captured in controlled conditions at Beadle HTC, UNL. Fig.1 shows a sample image of how a fluorescent image looks like. For classification and clustering, 10 plants are used, 5 from experiment-1 and 5 from experiment-2.

Fig.1. Fluorescent image of cotton plant, in SV_0 and SV_216

6 APPROACH:

The brief flow of the project is as follows:

(a). Segment the images to get the plant part

(b). Classify the pixels

(c). Apply k-means clustering and SVM classification.

7 PROCEDURE:

MATLAB is used for the implementation of this project as it has a fantastic collection of libraries for image processing.

A few assumptions have been made for this project.

1. The plants grown under controlled conditions are not affected by stress

2. For the plants in which stress is induced, the range of all the plant pixels before the day stress is induced, are considered to be control pixels.

The following steps show the procedure followed for the project:

1. Selection of images – Of all the views of images, top view of the plant showed the most leaf area. Hence, the top view images are selected for this project. Fig.2 shows an example of how a top view cotton plant.

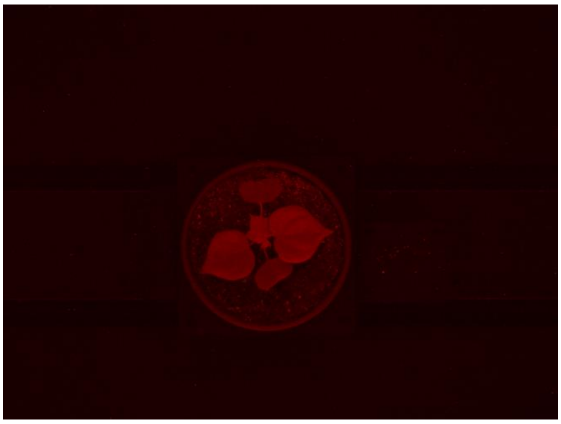


Fig.2. Top view image of cotton plant, Day 7

2. Segmentation – Image segmentation using color thresholding gives pretty good results for the RGB images. Hence, I tried to use the plant part of RGB images as a mask for getting the plant part in the fluroscent images. But ufortunately, the results are not very promising. The following figure Fig.3 show the results of using SURF feature detection algorithm.

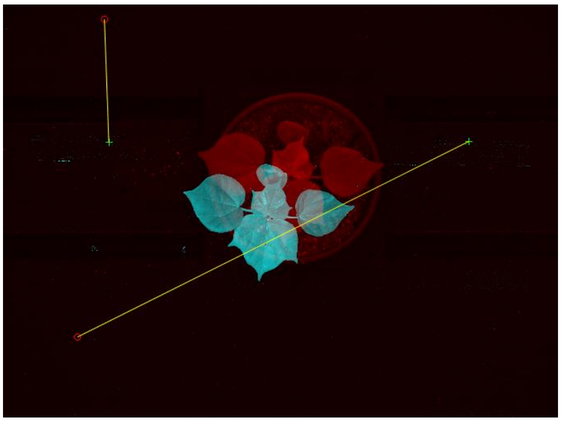


Fig.3. Matching points using SURF detection algorithm

The plant part is segmented from the background using the color thresholder application in MATLAB. Fig. 4 shows the result after the thresholding is done.

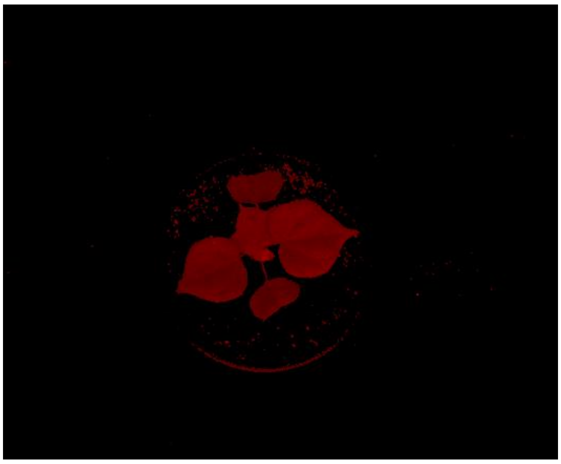


Fig.4. Image after color thresholding

The image still had some undesired areas which had to be removed. So, the image is binarized and erosion and dilation are performed on the image to generate a mask for obtaining the plant part. Fig.5 shows the binarized image and Fig.6 shows the image after performing erosion and dilation.

Fig.5.Binarized image

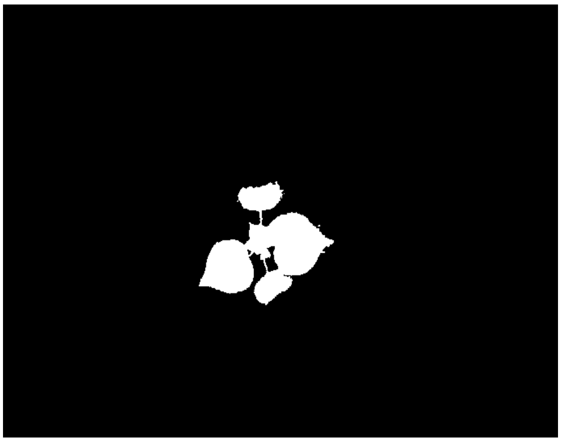


Fig.6. Image after erosion and dilation

The black and white image is used as a mask for the original plant image to get the plant part. Fig.7 shows the binary image used as a mask. Fig.8 shows the pant part obtained.

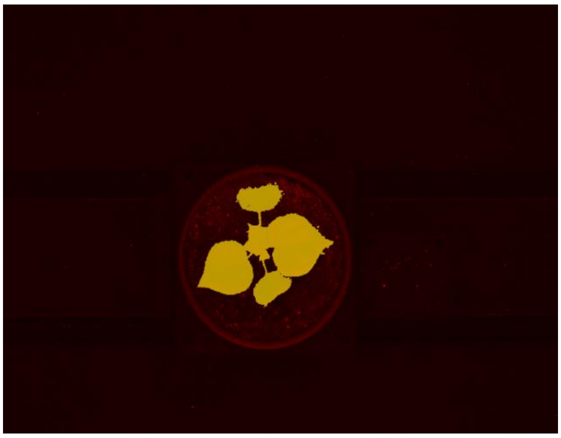


Fig.7. Binary image used as a mask for obtaining plant part

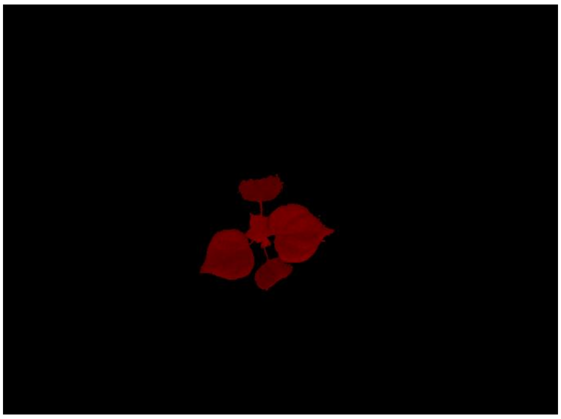


Fig.8. Image after the mask is applied

Histograms are drawn for the plant parts. The minimum and the maximum values of pixels are taken from the histogram.

According to the assumptions 1 and 2, the pixels lying between minimum and maximum values for the histogram are taken as non-stressed pixels for the controlled plants. For the stressed plants, the pixel range till the day the stress is induced, are taken as controlled pixels. The Fig.9 shows the histogram of a non-stressed plant.

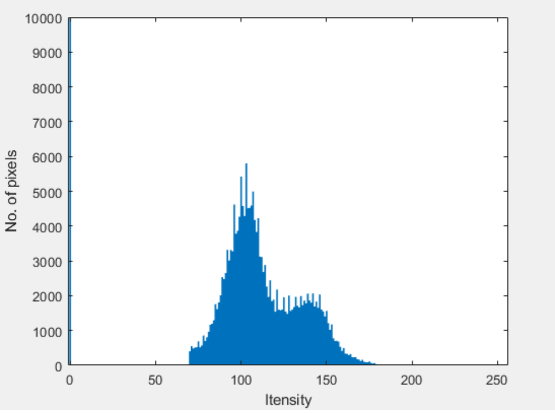


Fig.9 Histogram of non-stressed plant

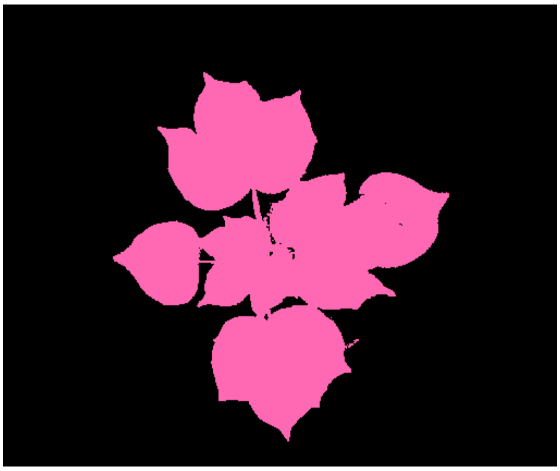


Fig.10.Non-stressed plant

In the stressed plants, the pixels which lie outside the range of controlled pixels are taken as stressed pixels. The Fig.10 shows the image of a stressed plant.

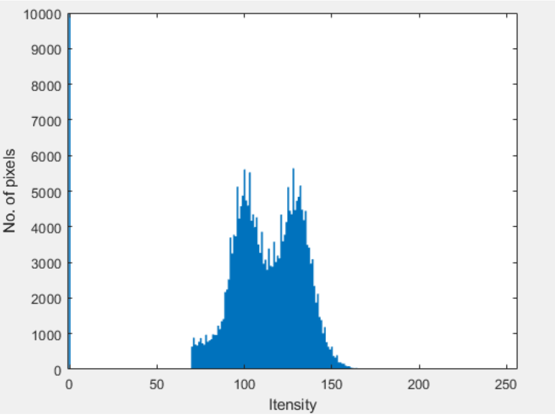


Fig.10. Histogram of a stressed plant

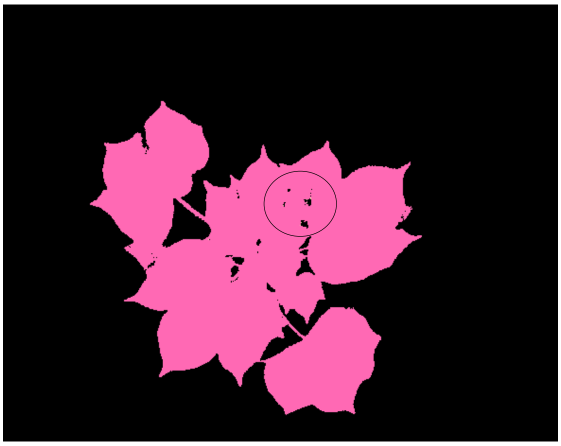
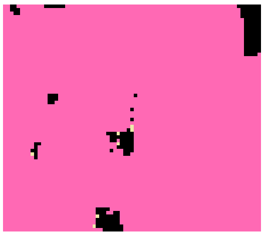


Fig.11.Stressed plant

From the above images, it can be seen that the spread of pixels is almost same for both the stressed and non-stressed plants, starting from 70 and going all the way till 190. The percentage of stressed pixels and the percentage of non-stressed pixels are calculated for each plant.

In the observed plant images the differences in maximums of stressed and non-stressed plant pixels are calculated. The pixels above the range of highest non-stressed plant pixel range are considered to be stressed pixels. The following images show the stressed(Fig.12) and non-stressed plants(Fig.11), with different colors for stressed and non-stressed plants. Since the ranges of stressed and non-stressed pixels are so close, instead of dividing the pixels into 3 classes, they are classified only into 2 classes. The non-stressed pixels are represented by pink color and the stressed pixels are represented by yellow color.

We can see that there is no significant difference between the stressed and controlled images. However, if we can observe closely, there a few yellow pixels, the stressed pixels. The Zoomed image of yellow pixels is shown in Fig.12.



Magnified circled part of image-12.

All the above information is stored in an excel sheet along with a column, which specifies whether the plant is stressed or is under normal conditions. This data is used to perform the analysis tasks like clustering and classification. A sample of how the data looks like is shown in table-1.

Min_value	Max_value	%_low_stressed	%_high_stressed	Stressed
0	185	0	100	0
0	192	0	100	0
0	194	0.00044	99.99956	0
0	193	0	100	0
0	194	0.000394	99.99961	0
0	194	0.000369	99.99963	1
0	185	0	100	1

Table 1: Plant data

In the above table(Table 1),

- min_value represents the minimum value of the histogram
- max_value represents the maximum value of the histogram
- %_low_stressed represents the percentage of low-stressed pixels in the plant.
- %_high_stressed represents the percentage of high-stressed pixels in the plant.
- Stressed says if the plant is stressed or not. 0 represents non-stressed and 1 represents stressed.

3. Clustering and classification - Clustering is done on the images to see which plants have similar behaviour when stress is applied on them. For performing the clustering, k-means clustering algorithm is used. This is because we can specify the number of clusters required even before starting the algorithm. The fig.13 shows how the data looks when plotted with minimum values against the maximum values. Fig.14 shows the result of clustering the data.

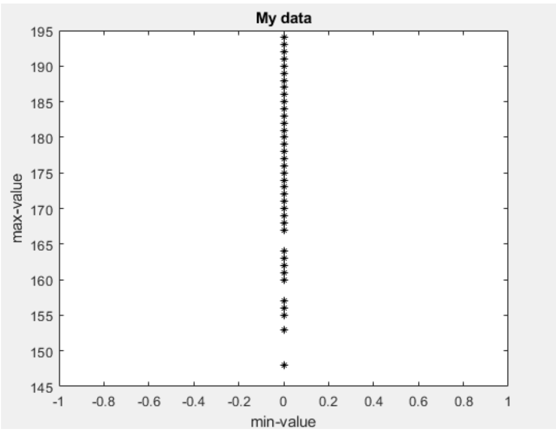


Fig.13.Figure shows the original data.

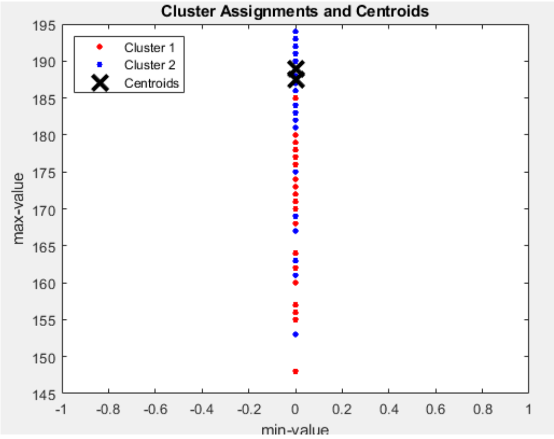


Fig. 14. Output of clustering the data.

Since the ranges of controlled and the drought stressed pixels are very similar, k-means clustering algorithm could not do a good job in clustering the data. The graph plotted is the minimum values in the range of histogram against the maximum value.

The same data is used for classification, to classify whether a plant is stressed or a control plant. Support Vector Machines are used for classification. The classification is done using 5-fold cross validation in order to prevent the overfitting of the data. The following figures show the results of classification. The following figure shows the result of classification. From Fig.15,

- The blue points indicate the plants correctly classified as low-stressed.
- The blue crosses indicate the plants wrongly classified as non-stressed plants.
- The orange points indicate the number of plants correctly classified as high stressed plants.
- The orange crosses indicate the plants wrongly classified as high stressed plants.

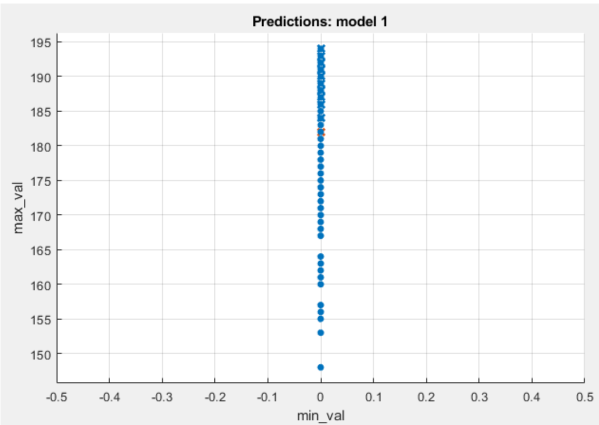


Fig.15. Output of classification

To look at it more clearly, Fig.16 shows the blue points, Fig.17 shows the orange crosses, Fig.18 shows the blue crosses, Fig.19 shows the orange dots.

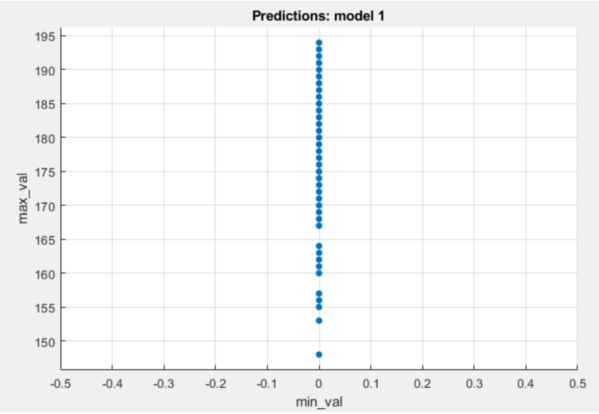


Fig.16.correctly classified controlled plants

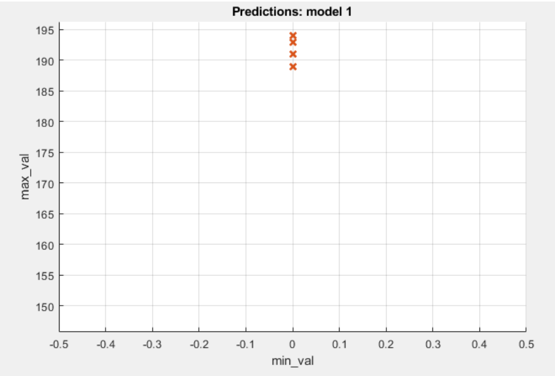


Fig.17.Incorrectly classified stressed plants

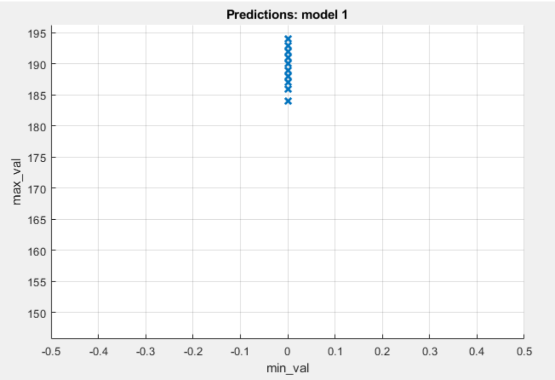


Fig.18.Wrongly classified controlled plants.

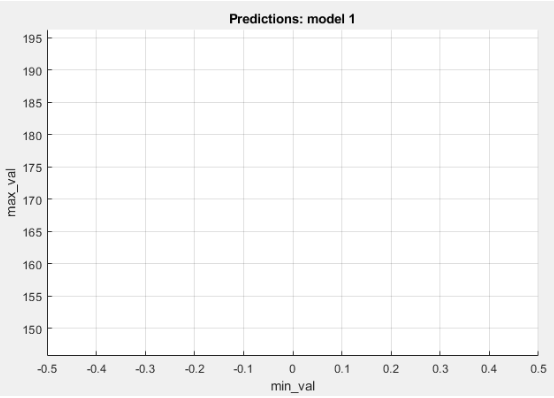


Fig.19.Correctly classified stressed plants.

From the above figures, it can be clearly seen that the model has difficulty in classifying the plants with high stress because the range of pixels in controlled plants and drought stressed plants is very similar.

EVALUATION:

There is no strict evaluation metric for this project as there is no ground truth available for the amount of chlorophyll present in the plant. However, the accuracy measure is taken for classification and the model could achieve 81.4% accuracy for being able to classify a plant, as stressed or non-stressed.

IMPLEMENTATION PLAN AND TIMELINE:

Spring break – End of March

Selecting the best view to proceed with the project work and separating plant part from the fluorescent images using visible light images.

Start of April – Mid of April

Plotting histograms of images and classifying the range of pixels into different groups. Assigning new pixel values based on where a pixel belongs on the range of total number of pixels.

Mid of April - End of April

Applying clustering and classification techniques on plant data. Generate the final report with results.

REFERENCES:

[1].Chlorophyll fluorescence-a practical guide by Kate Maxwell and Giles N. Johnson

[2]Gorbe, Elisa, and Angeles Calatayud. "Applications of chlorophyll fluorescence imaging technique in horticultural research: a review." *Scientia Horticulturae* 138 (2012): 24-35.

[3] Fahlgren, Noah, Malia A. Gehan, and Ivan Baxter. "Lights, camera, action: high-throughput plant

phenotyping is ready for a close-up." *Current opinion in plant biology* 24 (2015): 93-99.

[4] Wagstaff, Kiri, et al. "Constrained k-means clustering with background knowledge." *ICML*. Vol. 1. 2001.

*http://www.bbc.com/future/story/20170412-is-the-world-running-out-of-fresh-water

[5]Rousseau, Céline, et al. "High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis." *Plant Methods* 9.1 (2013): 17.