

Hybrid Method for Detection and Classification of Paddy leaf Deficiency Using Modified K-Means Image Segmentation

S. Sivagami

Assistant Professor, Department of Computer Science
Adhiyamaan College of Agriculture and Research, Hosur, Tamil Nadu
sivagamiacar@gmail.com

Dr. S. Mohanapriya,

Head of BCA Department, Sona College of Arts and Science, Salem
mohana_lks@yahoo.co.in

Abstract - Image processing widely used in agriculture sector for finding problems like disease identification, weed detection, fruit grading etc., In this paper, work is carried out based on automatic deficiency detection of Paddy leaves. Early detection of deficiency is most important to avoid qualitative and quantitative loss. Among all the grains rice is one of the most consumed grain in south India, but is easily affected by the nutrition deficiency. To increase the yield early identification of nutrient deficiencies of paddy crop is very essential. Paddy leaves color plays an important role in identifying micro deficiencies such as CSM (Calcium, Sulfur and Magnesium) during middle stage of its growth. Database of healthy, calcium defected, sulfur defected and magnesium defected leaves is created to identify deficient paddy leaves. HSV color model is used to extract Color features of both healthy and defected paddy leaves. Color features of test image is also extracted and compared against database properties. Comparison results are checked with the rules set to decide the specific deficiency. The rules are framed based on thorough experiment.

Keywords: Image segmentation, Classification, deficiency, Modified K-Means

1. Introduction:

Agriculture is the mother of all culture and paddy is the widely used cereal in south India. It is required to increase the crop yield to satisfy the food requirement of large population. The productivity and the quality of the cultivated crops heavily depend on the natural factors like soil type, nutrients available, water, seed quality etc. Sometimes It needs to apply externally fertilizers and nutrients if natural resources not fulfil the crop requirements. These extra supplements clearly increase the cost of productivity. Because of this, now a day it has been seen agriculture is not a profitable business, because of small difference between production costs and market price of crop. Paddy being the staple food of India, it is majorly affected by deficiency of primary and secondary nutrient elements like nitrogen, phosphorus, potassium, calcium, sulfur and magnesium. Leaves can be deficient with multiple nutrient elements at a same time. This can alter natural color of paddy leaves. Such leaves are considered as defective. Firstly, the database of healthy, calcium, sulfur and magnesium defected paddy leaves are created.

Natural factors like soil type, nutrients available, water and seed quality etc., affects the productivity and quality of the cultivated fruits and vegetables. Sometimes it is necessary to supplement extra nutrients if natural resources do not fulfil the crop requirement. For this extra supplement of nutrients, the productivity cost will increase. By utilizing technology development, we can use image processing for detection of plant disease by analysing its deficiency. Generally, we can observe the symptoms of deficiency on leaves, and stems. so here we use leaves for identification of deficiency in affected plants. Nitrogen, phosphorus and potassium are primary nutrients and very important to tomatoes Plants to grow. Nitrogen helps growth and chlorophyll production, and phosphorus helps tomatoes grow and cope with stress while aiding in energy production. Potassium is also important nutrient it provides immune to the plant to fights against disease, Potassium also improves the quality tomato and it regulates CO₂ intake so helps in photosynthesis. Tomatoes also need the "secondary nutrients," calcium, magnesium and sulfur, but in lesser amounts than the primary nutrients. Calcium improves is the most important secondary nutrient, it also provides immunity to the plant and main benefit of soil is it allow water penetration to the plant. Deficiency in sulfur will produce yellow leaves. Tomatoes also need little number of micronutrients too, the micro nutrients are Zinc, iron, boron, chloride, molybdenum, copper and manganese. Molybdenum, helps to efficiently use nitrogen, while zinc helps regulate growth and promotes proper sugar consumption. Boron helps the development of good quality fruit and seeds. Nutrients are observed from soil for growth of plants. the growth of

plant/crop affected by deficiency of these nutrients. Therefore, analysing nutrients status will play crucial role in agriculture and farming. The deficiency of nutrients in plants will reflect in its leaf. These symptoms are listed as follows, colour of the leaf's changes in to yellow colour, curled edges of the leaves, size of the leaf reduced to smaller size interveinal chlorosis, necrosis, distorted edges, One critical issue in identifying deficiency is similar symptoms present in old and younger leaf does not have same nutrient deficiency

2. Literature Survey

N. Minni et al. (2016) have proposed a new method for Detection of Nutrient Deficiencies in Plant Leaves using Image Processing. In this paper first plant leaf images with macro nutrient deficiency symptoms were collected, then in pre-processing step the image is converted into HSI format and H value taken into account for their research. After that image segmentation was done by masking green pixels, remaining pixels are considered as useful segments. Texture and color features are extracted from segmented image, using this features database was created. K-Nearest Neighbour classifier was used to classify the image. Nurbaity Sabri et al. have presented a new method for Nutrient deficiency detection in maize (*Zea mays L.*) leaves using image processing. In this research they consider three main nutrient deficiency like nitrogen, potassium and magnesium. Maize plant leaf images with these three nutrient deficient were collected, the images were converted into 500x500 size to reduce execution time then median filter was applied to remove noise from the image. After pre-processing the image shape and texture features were extracted then classification was done with random forest classifier. Han Yuzhu et al. (2011) have proposed strategy for nitrogen assessment for pepper in blossoming and fruiting utilizing color image processing. Various elements of RGB were thought of and connected with nitrogen. Relapse examination for inorganic nitrogen in soil, absolute nitrogen, nitrogen fixation and SPAD meter readings were done and all shows negative coefficient for the thought about capacity. N applications were given in various medicines. Gautam, R.K. and S. Panigrahi (2007) created image processing method for multispectral images alongside R, G, and B groups. NDVI and Green vegetation list were determined. Multilayer perceptron and outspread premise work neural network were sent for leaf nitrogen assessment of corn utilizing conventional spectrums and determined 26 band list. Radial basis work indicated upgraded yield among all the neural organization models.

Adán Mercado-Luna et al. (2010) have proposed R, G, B color index esteems that were related with nitrogen and chlorophyll of tomato. Estimations of N and chlorophyll which were estimated by compound investigation and SPAD meter separately. Red and Blue index indicated high negative connection contrasted with Green index. Trial was completed for various medicines of nitrogen. SPAD meter connection was not suggested by numerous scientists as it relies upon nitrogen as well as on temperature.

3. Materials and Methods

The paddy leaf images were classified into two classes in this proposed methodology as healthy or defected. First images are classified into healthy and defected. Defected leaves are further classified as calcium, sulfur and magnesium deficient paddy leaf images. Three steps are used to classify Leaf image of paddy, namely, Image acquisition, Segmentation, Feature extraction and Classification as shown in figure1.

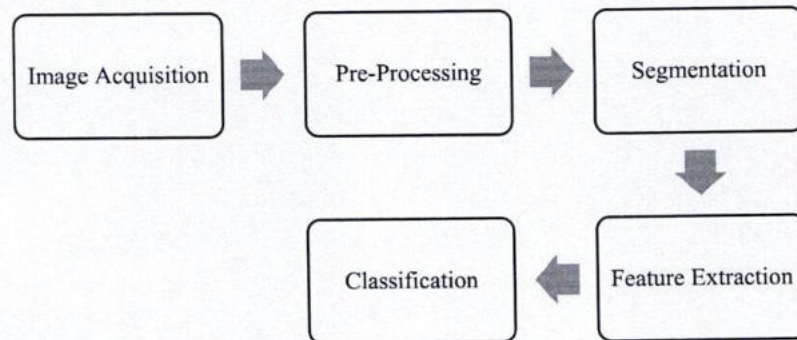


Fig. 1. Step flow of the proposed work

3.1 Image Acquisition:

Nearly 200 images are used from image database available in Tamil Nadu Agriculture University, Coimbatore. Among the 200 images 40 images are healthy, and 40 images are calcium deficient, 40 images are sulfur deficient and 40 images are magnesium deficient and remaining images are used for testing. Tips of youngest leaves become white or bleached, rolled and curled are the symptom of calcium deficiency. Yellowing or pale green color of the whole plant is the symptom of sulfur deficiency. Pale grayish green interveinal chlorosis spreads from the tip to the leaf base is the symptom of magnesium deficiency. Sample leaf images of healthy and deficient leaf images are shown as in figure 2.

3.2 Pre-Processing:

Pre-processing is the process of improving the appearance of the image by denoising method that is filters used to remove noise in image, contrast enhancement used to improve the quality of the image and image resizing.

All the images are resized into 256X256 format for easy handling of image. Images are converted into HSI format

3.3 Image Segmentation:

This is the process of split the image into no of groups such that within the group the pixels are more related and between the group the pixels are departed a lot. There are lots of algorithms available for segmentation, among them in this work modified K-means segmentation algorithm was proposed to segment the image. In K-Means segmentation algorithm, it is necessary to manually assign the value for k and initial centroids are selected randomly. Final segmentation results mostly based on these initial values. Sometimes it is possible for wrong segmentation because of incorrect values assigned to k and initial centroid. To overcome these problems new K-means algorithm was proposed, in this algorithm it is not necessary to manually assign any value, it automatically segments the given image based the following modified K-means segmentation algorithm.

3.4 Feature Extraction:

Feature extraction is the process of retrieving important attributes from segmented image. There are three types of features like texture feature, shape feature and color feature. In this proposed method statistical and texture feature are extracted using GLCM feature extraction method. The following GLCM features are used for this proposed method Contrast, energy, entropy, Homogeneity, uniformity, etc.,

3.5 Classification:

Final step in the disease identification is classification. In our proposed work we did the classification based on five classifiers.

3.5.1 **Multiclass Support Vector Machine:** SVM was introduced by Cortes and it is mostly used for classification purposes. It is a supervised learning model along with learning algorithms that are used to analyze the data and mainly used to recognize the patterns for classification. Basic SVM is a binary linear classifier that takes a set of input and for each input, it predicts which of the two classes it belongs. In a linear classifier best hyperplane is the one that separates between two classes. If the best hyperplane exists for the given data set then it is called maximum margin hyperplane and the linear classifier is known as a maximum classifier. Svm minimizes experimental classification error as well as maximize the margin so it is called maximum margin classifier.

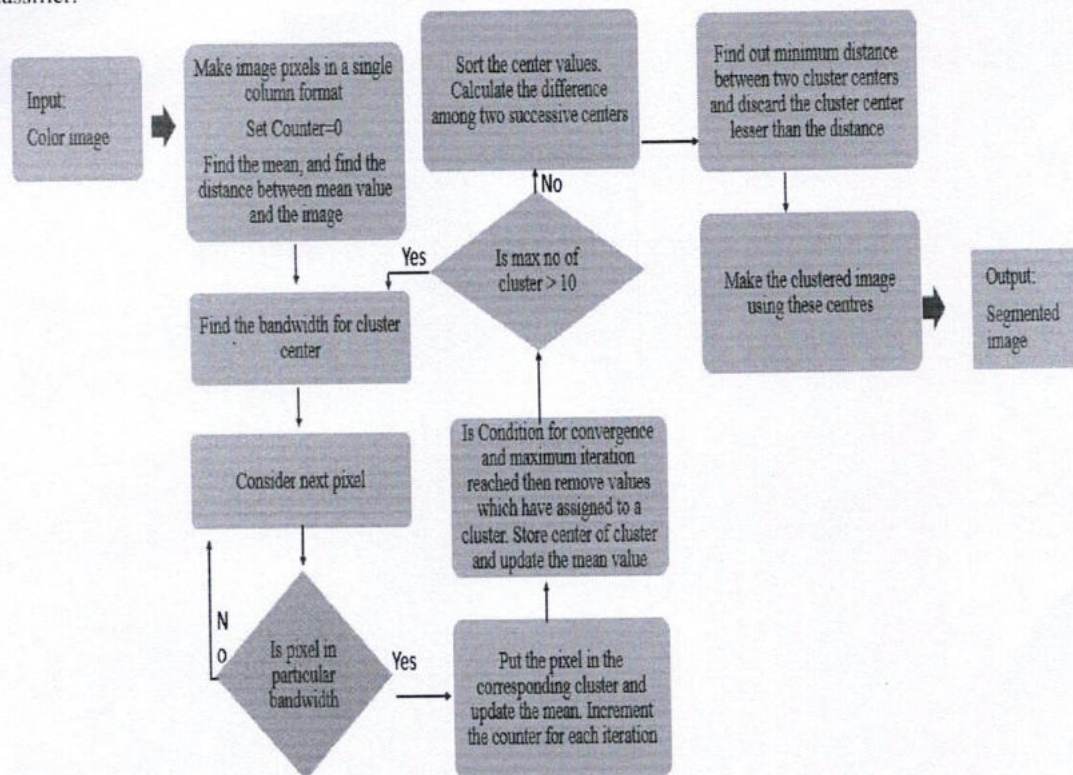


Fig. 2. Step flow of Modified K-Means Segmentation algorithm

3.5.2 Decision Tree:

It is a supervised learning algorithm mainly used for classification and regression problems. It is a tree-like structure it starts with the root node and expands on further branches. Each leaf node represents the output, internal nodes denotes dataset features and branches denote the decision rules. Classification and regression tree algorithm is used to build the tree-like structure. The below diagram describes the outline of the decision tree structure.

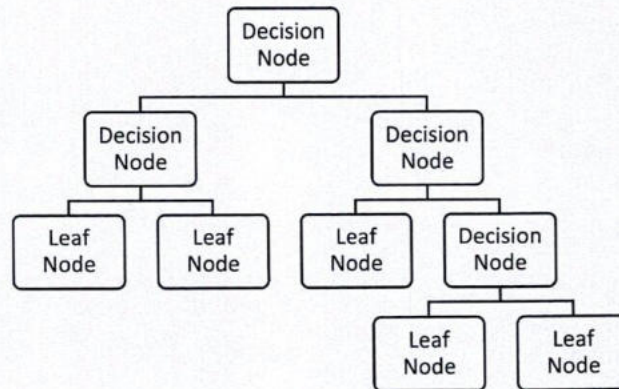


Fig. 3. Decision Tree Structure

3.5.3. K-Nearest Neighbor (KNN):

It uses the similarity between features to predict the new data points. The value will be assigned to the new data points based on the concept that how it closely matches the points in the training set. The following steps explain the working flow of KNN.

- Load training and testing dataset.
- Choose the value of K.
- For each test data do the following:
 - Determine the distance between test data and each row of training data.
 - Sort them in ascending order.
 - Choose the top K rows from the sorted array.
 - The test point will be assigned to the most frequent class of the rows.

3.5.4. Naïve Bayes Classifier:

It works based on Bayes theorem. For each class membership probabilities are calculated for data points associated with a particular class. The most suitable class will be chosen based on the class having maximum probability. This is also called Maximum A Posteriori (MAP). It can be summarized as the following equation:

$$k = \operatorname{argmax}_k P(c_k) \prod_{i=0}^n P(x_i/c_k)$$

Where the denominator $P(x)$ is omitted since the value is same for all class. The MAP for a hypothesis is:

$$\operatorname{MAP}(H) = \max P(H/E)$$

$$\operatorname{MAP}(H) = \max(P(H/E) * P(H)) / P(E)$$

$$\operatorname{MAP}(H) = \max(P(E/H) * P(H))$$

3.5.5. PSO-ANFIS Classifier

The following figure shows the particle swarm optimized Adaptive neuro Fuzzy Inference system.

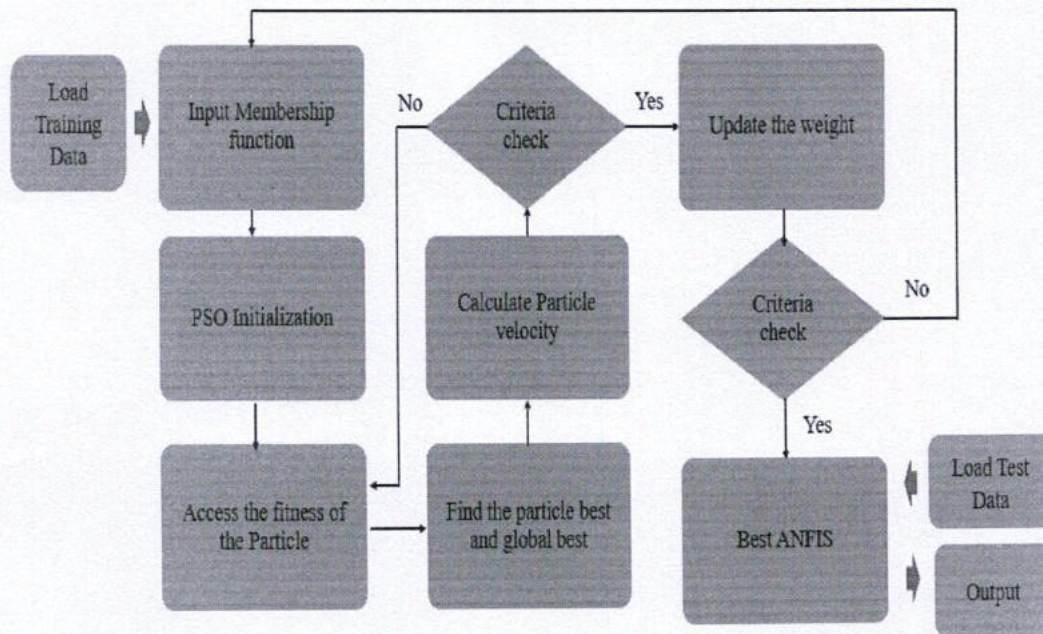


Fig 4. PSO-ANFIS Classifier work-flow

4. Result and Discussion

In our proposed method we consider three micro nutrients like Calcium, sulfur and magnesium in paddy plant. The input paddy leaf images are of different sizes, so the first image resize was done then the color image is converted into the gray image after that image denoising was done using the Gaussian filter which is followed by segmentation by using proposed modified K-means segmentation algorithm. After performing segmentation useful features are extracted from ROI using the GLCM feature extraction method. After getting useful features classification was done using M-SVM, K-NN, Decision Tree, Naïve Bayes and PSO-ANFIS classifiers. The following table 1 shows number of training and testing images used in our proposed work.

Table 1. Paddy Leaf Dataset

| Type of Leaf | No. of Training Images | No. of Testing Images | Total No. of images |
|--------------------|------------------------|-----------------------|---------------------|
| Healthy images | 40 | 10 | 50 |
| Calcium defected | 40 | 10 | 50 |
| Magnesium defected | 40 | 10 | 50 |
| Sulfur defected | 40 | 10 | 50 |
| Total | | | 200 |

The following figure 5, shows the calcium deficient leaf image and its segmented part

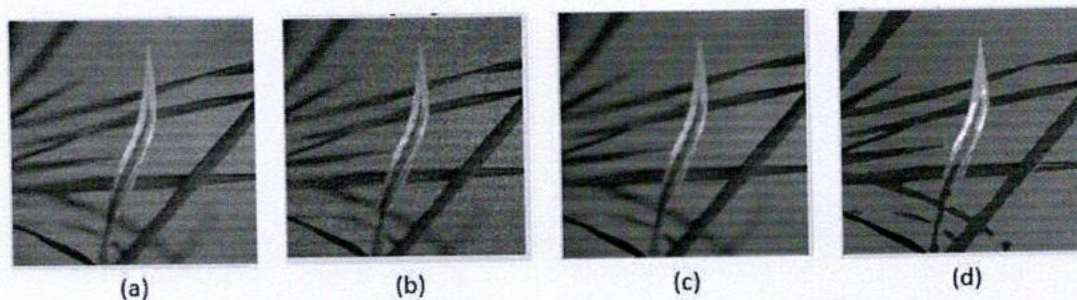


Fig. 5. (a) Original image (b) Noised Image (c) De-noised image (d) Segmented image

The following table shows the performance of different classifier after Modified K-Means segmentation

Table 2. Performance of classifier after K-means and Modified K-Means segmentation

| Classifier | Precision | Recall | Specificity | False positive Rate | Accuracy (%) |
|---------------|-----------|--------|-------------|---------------------|--------------|
| KNN | 0.7047 | 0.6933 | 0.8587 | 0.0953 | 88.11 |
| Decision Tree | 0.6809 | 0.6667 | 0.8533 | 0.1191 | 87.22 |
| Naïve Bayes | 0.6379 | 0.6167 | 0.8433 | 0.1621 | 85.56 |
| M-SVM | 0.7219 | 0.7133 | 0.8627 | 0.0781 | 91.78 |
| PSO-ANFIS | 0.8443 | 0.8333 | 0.9667 | 0.1557 | 94.44 |

The comparison of different classifiers is shown in the following figure 6 and 7.

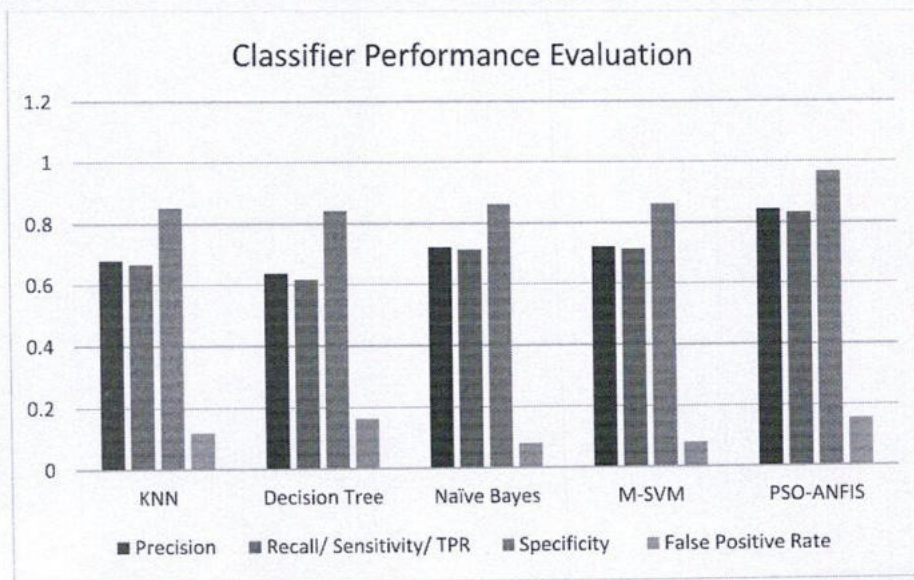


Fig. 6. Performance evaluation of different classifiers

The following formulas are used to calculate Accuracy, precision, recall, specificity and False positive rate (FPR)

$$Accuracy = \frac{\text{Number of data Correctly Classified}}{\text{Total number of data Classified}} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$False\ Positive\ Rate(FPR) = \frac{FP}{FP+FN} \quad (5)$$

Where TP – True Positive, FP – False Positive, TN – True Negative and FN – False Negative

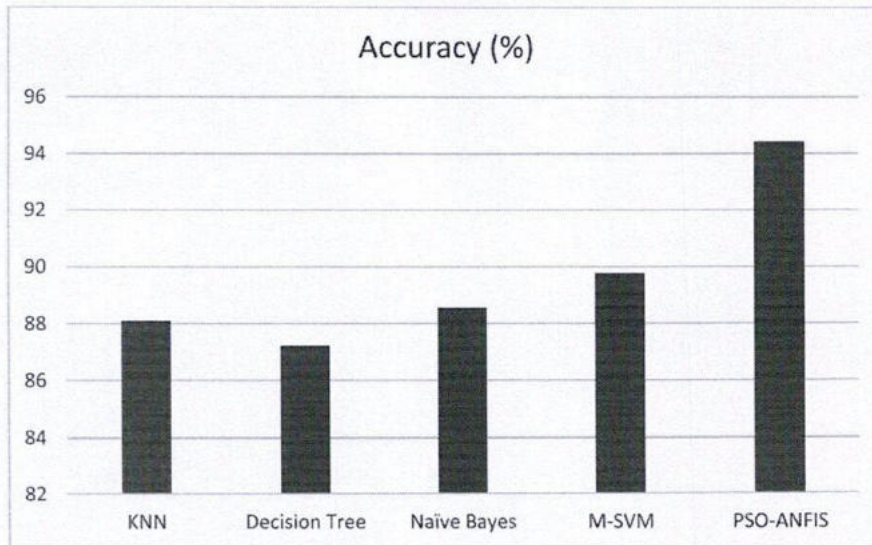


Fig. 7. Accuracy comparison of different classifier

5. Conclusion

This work mainly focussed on the detection of deficiency in paddy plant with the proposed modified K-means segmentation algorithm. In this modified K-Means segmentation algorithm it is not necessary to initialize cluster number and cluster center. Our work also compares the accuracy of different classifiers. Among these five classifiers, PSO-ANFIS provides more accuracy than others, and our proposed Modified K-means and PSO-ANFIS classifier produced a better accuracy of 94.44%. In the future, we can extend the work to other plants as well as a other nutrients too.

References:

- [1] Gautam, R.K. and Panigrahi, S. (2007), "Leaf nitrogen determination of corn plant using aerial images and artificial neural networks", *Canadian Biosystems Engineering*, volume 49, pp-7.1 - 7.9.
- [2] Simone Graeff, Judit Pfenning, Wilhelm Claupein, and Hans-Peter Liebig (2008), "Evaluation of Image Analysis to Determine the N-Fertilizer Demand of Broccoli Plants", *Advances in Optical Technologies*, Article ID 359760, Volume 2008, pp-1-8.
- [3] S. Jeyalakshmi and R. Radha, A Review on Diagnosis of Nutrient Deficiency Symptoms In Plant Leaf Image Using Digital Image Processing, *ICTACT Journal On Image and Video Processing*, MAY 2017, Vol. 07, ISSUE: 04 pp. 1515-24
- [4] Y. Karimi, S.O. Prasher, R.M. Patel, S.H. Kim (2006), "Application of support vector machine technology for weed and nitrogen stress detection in corn", *Computers and Electronics in Agriculture* 51, pp- 99–109.
- [5] Xiao-li Li and Yong He (2008), "Chlorophyll Assessment and Sensitive Wavelength Exploration for Tea (*Camellia sinensis*) Based on Reflectance Spectral Characteristics", *HORTSCIENCE*, Vol 43(5), pp- 1586–1591.
- [6] Adán Mercado-Luna, Enrique Rico-García, Alfredo Lara-Herrera, Genaro Soto-Zarazúa1, Rosalia Ocampo-Velázquez, Ramón Guevara-González, Gilberto Herrera-Ruiz and Irineo Torres-Pacheco (2010), "Nitrogen determination on tomato seedlings by color image analysis (RGB)", *African Journal of Biotechnology*, Vol. 9(33), pp- 5326-5332.
- [7] N. Minni and N. Rehna. (2016), "Detection Of Nutrient Deficiencies In Plant Leaves Using Image Processing." *International Journal of Computing Algorithm* 5.2 (2016): 84-87.
- [8] Nurbaity Sabri et al. (2020) Nutrient deficiency detection in maize (*Zea mays L.*) leaves using image processing, *IAES International Journal of Artificial Intelligence (IJ-AI)* Vol. 9, No. 2, , pp. 304–309 ISSN: 2252-8938, DOI: 10.11591/ijai.v9.i2.pp.304-309
- [9] Marian Wiwart, Gabriel Fordon' ski, Krystyna Z' uk-Gołaszewska, El 'zbieta Suchowilska (2009), " Early diagnostics of macronutrient deficiencies in three legume species by color image analysis", *Computers and Electronics in Agriculture* 65, pp- 125–132.
- [10] Han Yuzhu, Wang Xiaomei and Song Shuyao (2011), "Nitrogen determination in pepper (*Capsicum frutescens L.*) plants by color image analysis (RGB)", *African Journal of Biotechnology*, Vol. 10(77), pp- 17737-17741.