

# An Effect of Nutrient Deficiency on Yield Estimation



Sushila Shidnal, Mrityunjaya V Latte

**Abstract-** By taking corrective measures to improve the farming quality, agricultural sector need a thoroughly explained and systematic theory for crop yield prediction. Any yield of the crop is usually depending on the crop unhealthy and healthy conditions. These conditions mainly occur due to major nutrients like nitrogen, Phosphorus and Potassium (NPK). Nitrogen deficiency will make the fields in some parts look Yellowish. Potassium deficiency may lead to have spots in the leaf and Phosphorous will make the fields some part look brownish. Hence segmenting this defected area is the major challenge to evaluate the total yield in the input paddy field image. The proposed model focus on segmentation of these regions using an efficient hierarchical model. This model uses segmentation methods like FCM and Color segmentation techniques there by improving the accuracy of the system and comparing with the ground truth values.

**Keywords –** Fuzzy C Means (FCM), GLCM, HSV Histogram, Hierarchical Colour Based Segmentation, NPK Regions.

## I. INTRODUCTION

The agriculture and advance technology combinations are both used for the improvement of the crop yield production and it becomes an interesting topic in recent years. Due to these rapid developments, efficient models of the crops and predictive tools might be expected to become a decisive component of precision agriculture. For producers, agricultural related organization and consultants yield of crop prediction has been a main topic of interest. Crop yield is a consolidated bio-socio-system consisting of complex or difficult interaction between the air, the water, the soil, and crops have been grown in it, where a model of comprehensive is needed which is only possible through the expertise of classical engineering.

With regarding the definition of Agriculture and Food by United Nations, forecasting of the crop is the prediction art for crop yield and also previously a few months in advance the production earlier the harvest actually takes place. This forecasting philosophy is usually based on different kinds of data collection from a different kind of sources like soil, meteorological data, remote sensed data and also agriculture statistics. Based on agronomic and meteorological data a number of indices are deemed which are all relevant in the perseverance of the yield of the crop for water satisfaction of instance crop, excess moisture, surplus and also average soil moisture.

Also quantifying changes in production systems of the crop overtime at the hierarchy of different levels are needed for the evaluation of the sustainability of the different strategies of management.

For the sustainability concepts development and yield forecasting different methods, implementation of long term field experiments are not conducted with the sufficient amount of detail in time and space; providing a critical tool for finding combinations in strategies of management for reaching the multiple goals are necessary for able to maintain a certain level in a crop.

These models help in providing land managers and makers in policy to experiment results from one location to another where there is a lack of response information. Many factors influence crop production, either indirectly or directly which affects crop performance. Some factors that are usually highlighted in the soil field the research are PH, texture, available nutrients, Organic matter, fertilizer application, and cultural practices, etc. have formed a complex system for agriculture since it deals with a large data set for each of the problems it faced with.

By taking into consideration all these problems and issues, proper technology and solution to overcome this are still not come. Multiple ways are there to expand the yield of crop and crops quality. As segmentation of the affected area in the field, will help in finding yield total, this paper focuses on segmenting the field area affected by NPK (Nitrogen, Phosphorus and Potassium) thereby evaluating the total yield. The assumption is based on that the if some field part looks yellowish then it is due to nitrogen deficiency, if the area looks brown then it is due to potassium deficiency and also if the leaf contains brown spots in it then it is due to phosphorous deficiency

## II. LITERATURE SURVEY

S. Sridevi et.al [01] has presented a systematic approach for Maize leaf nutrition deficiency identification. Based on this approach the feature extraction identifies the lack of nutrients in the crop image. This approach considered three steps for processing like pre-processing, regression and feature extraction. Processing of the HSV transformation and enhancement of histogram is done. Different features like color, shape, and texture are extracted. As maize image analyzer multivariate partial least square are used.

Hiteshwari Sabrol et.al [02] proposed an approach which gives a survey on different recognition of disease of the plant and classification by using the image processing techniques and soft computing. This paper aimed at focused on the pathology area of the plant for classification and recognition. In this work proposed an algorithm and it begins with the acquisition of the noninfected and infected plants, followed with image pre-processing, differentiation of the infected region due to disease using segmentation approach.

Manuscript published on 30 September 2019

\* Correspondence Author

**Sushila Shidnal\***, Assistant Professor, Department of Computer Science & Engineering,

Sir M Visvesvaraya Institute of Technology, Bangalore, 562157 India.

**Mrityunjaya V Latte**, Principal, JSS Academy of Technical Education, Bangalore, 560060 India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## An Effect of Nutrient Deficiency on Yield Estimation

Manoj Mukherjee et.al [03] have presented a systematic approach for image processing for the paddy leaf using the histogram for avoiding large scale effect of these diseases. Using this approach in the early stage the disease detection is done and hence helps in reducing the production loss. This algorithm implementation is done using Matlab software. By the help of agricultural experts, the consultative treatment module for the disease has been prepared as once the disease detection is done.

S. A. Ramesh Kumar et.al [04] has presented a systematic method for finding the unhealthy regions in Paddy Image. This work provides advances in different methods are used to study paddy diseases/traits using image processing and data mining. The methods studied are for throughput increments & reducing subjectiveness arising from human experts in detecting the paddy diseases are stored and gives the solution to the diseases. Different characteristics of paddy crops are grouped by an “associative rule mining”.

Asif Ali et.al [05] has developed a different method for image analysis for yield calculation of maize. The analysis of the relationship between the fresh biomass maize yield and the cover of the weed leaf at the sixth and the fourth stages where analyzed. This approach showed the total yield loss, which was linearly related to the weeds of the leaf cover. This is also used in the algorithm of decision for the intelligent boom sprayer.

Manickam Gopperundeivi et.al [06] had presented an efficient method for the identification and mapping of the crop by assessing its vigour. This work reports on estimating the paddy field taking different data. The study also involved in land images classification using the Vegetation index and supervised classification. Vegetation index like Normalized Difference Vegetation Index (NDVI) along with some Field data is used for finding the yield and NDVI relationship. This study was done to look over the general applicability of high temporal resolution Moderate Resolution Imaging Spectroradiometer (MODIS) of gridded vegetation product of 250m for monitoring the rice crop growth, mapping rice crop acreage and analyzing crop yield, at the province-level.

J. S. J. Wijesingha et.al [07] for the rice crop growth monitoring analyses on the general applicability of high temporal resolution Moderate Resolution Imaging Spectroradiometer (MODIS) grinned vegetation product. This work performed for data pre-processing and the analysis of the indices of Vegetation for rice acreage extraction for the overall yield analysis. Yield analysis of the point level showed the MODIS is correlated more with the rice yield for the prediction of yield using a maximum of Enhanced Vegetation Index in the cycle of yield prediction along with prediction error.

Mrs .K.R. Sri Preethaa M.E et.al [08] had presented a systematic approach for prediction of the yield of the crop and suggests the best crop by improving the profitability and quality of the agriculture sector. This work aimed at discussing the approach to expand the rate of the yield of the crop is based on the parameters like atmosphere and soil. Dataset involved different parameters like temperature, Soil Type, water Level, Depth, Spacing, Season, Month and Fertilizer which helped framers in choosing the specific crop for specific soil. Prediction is carried through using Bayesian algorithm good result will get.

Siti Khairunniza-Bejo et.al [09] presented a robust way for total yield calculation using ANN classifier. This work also presented a review of different ANN approaches for predicting yield using different factors of crop performance. This summarized on the ANN application and their basic concept on the architecture of the neural network is also presented. It is found that the ANN is giving good results when compared with other classification algorithms on total yield calculation in the crop field image.

Shreya S. Bhanose et.al [10] proposed an efficient approach for predicting the crops with its yield and hence supporting farmers to take up correct decisions to enhance the farming quality. This approach was done for clustering algorithms like modified K-means for extracting the useful information in the given prediction technique as Traditional clustering algorithms like k-Means, improved rough k-Means, and means++ makes the tasks complicated due to the random selection of initial center of cluster and decision of the number of clusters. Researchers in [11], [12], [13], [14], [15], [16] and [17] are also proposed efficient methods for prediction of crop technique.

The proposed method made use of different techniques like Fuzzy C means (FCM), Color segmentation based NPK region separation using hierarchical methods to get the NPK segmented region. Once the NPK regions are obtained they are again subjected to the validation process to know whether they are actually NPK segmented regions using ANN classifier. The validation step is followed by Yield calculation step for the evaluation of the total yield in the Paddy Field Image.

### III. METHODOLOGY

The block diagram of the proposed system is as shown in Fig. 1. The proposed system consists of two phases called as Testing Phase and Training Phase. Initially, Paddy Field image is taking as an input in the testing phase and it is then passed to the pre-processing block. Different steps like Noise removal, Resizing is done in the pre-processing block. Removal of noise is done using the median filter. Other than a field, this image may contain other regions which can be removed using manual ROI region selection. This noise free image is manually cropped and then down sampled by taking scaling factor. The resized image is then passed to an unhealthy and healthy region block. Separation of the healthy and unhealthy region is done using FCM clustering technique. The segmented cluster consisting of the much healthier region is chosen and it converted to binary to undergo Hierarchical approach of color segmentation for finding the suspected regions of NPK based on Colour thresholding to separate the final N, P and K regions in the original Paddy Field Image.

All these segments along with healthy regions are then passed to feature extraction block for extracting features like HSV color histogram and GLCM features. In the training phase defected samples due to NPK deficiency and also healthy samples are trained. Feature extraction from all these samples using the similar methods as carried through in testing phase is followed and the feature is extracted s then stored in Knowledgebase after training the ANN.

The features have been extracted during the testing and training phase is then passed to the ANN classification for validation. Once the validation is done on the segmented NPK and Healthy regions to know whether they actually defect or healthy regions, the NH, PH and KH healthy regions obtained from N, P and K regions after classification is also passed to the yield estimation block. Yield estimation step is carried out to estimate the Nutrient Deficiency is to find the total yield in the Input paddy field Image.

### 3.1 Pre-Processing and Manual ROI Selection

Input Paddy Field image initially as in Fig. 2 is passed to the pre-processing block. The input image in the pre-processing block is passed to the median filter. For filtering the weighted median filter (3x3) is made use. This is a non-linear digital technique and is often used for removing the noise. This type of preprocessing is done for improving processing result because the median filter is chosen in certain conditions and it preserves the edges while removing the noise. We suppose to evaluate the output of the weighted median filter, say  $v$  for the input  $I$  (1020x1020x3). The formula for this filter is as given below,

$$\min_v \sum_{i \in V} \sum_{l \in N_i} w_{il} |v_i - I_l| \quad (1)$$

Where the location of pixel  $V$  is in the image  $N_i$  denotes the set having pixel  $i$  and also the neighboring pixels of  $i$ .  $w_{il}$  denotes the non-negative weight [18]. The pre-processed image obtained is then subjected to Manual ROI selection block to manually crop the unwanted area which presents in input paddy field image as in Fig. 3.

### 3.2 Fuzzy C-Mean Clustering

The pre-processed block has an output and it is then passed to FCM clustering block. The approach of the clustering is partitioned  $N$  objects into  $C$  classes. In the proposed work  $N$  is equal to the number of pixels in the image. The number of clusters is denoted by  $C$  i.e. 3. The FCM algorithm utilizes the iterative optimization of a function objective based on a weighted measure of similarity between pixels present in the pixel and also between each cluster. The objective function's local extremism indicates an optimal clustering of the input data. The equation for minimized objective function is given by,

$$Q = \sum_{i=1}^C \sum_{j=1}^N (u_{ij})^m \|z_j - v_i\|^2 \quad (2)$$

Where always  $z_j \in Z$  and  $Z = \{z_1, z_2, z_3, \dots, z_N\}$  and  $v_i \in V$ , where  $V = \{v_1, v_2, \dots, v_c\}$ .  $\|*\|$  Are an expressing norm and similarity between the measured data value and also the center of the cluster.  $m \in [1, \infty]$  Depicts the weighting exponent and can be a real number is greater than 1. The best value for 'm' according to the calculation is between the intervals 1.5 to 2.5 and hence the best value of  $m$  is 2 is utilized as a good choice of the parameter of fuzzification. The fuzzy  $c$  partition for the given dataset is the fuzzy matrix of partition  $U = [u_{ij}]$  with  $i=1, 2, \dots, C$  and  $j=1, 2, 3, \dots, N$ . where  $u_{ij}$  depicts the membership degree of  $j^{th}$  pixel to the  $i^{th}$  cluster. The functions of membership are then subjected to satisfying the conditions given below [19].

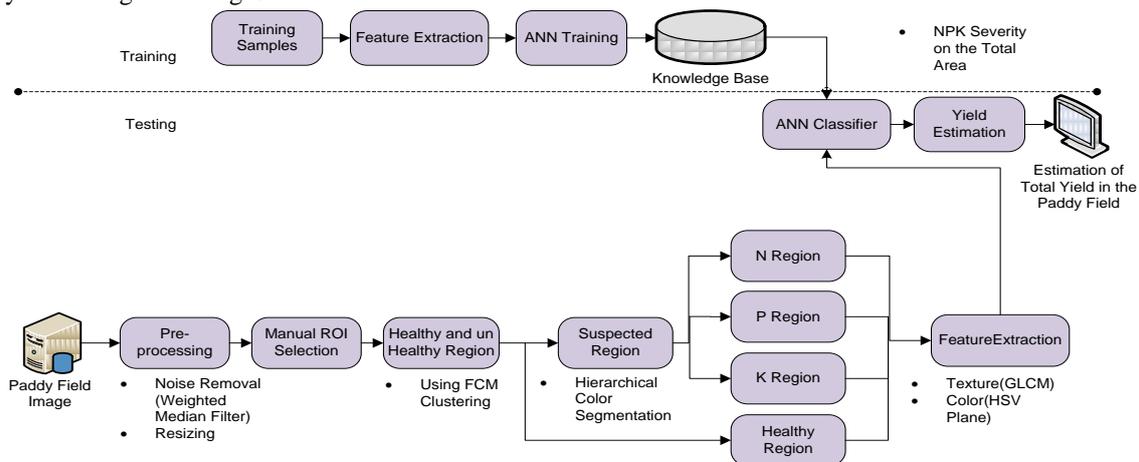


Fig. 1. Block Diagram of Proposed System



Fig. 2: Input Paddy Field Image



Fig. 3: Cropped Paddy Field Image

This algorithm main goal is finding the optimal  $C$  partition of Fuzzy by evolving the fuzzy matrix of partition  $U = [u_{ij}]$  repeatedly for finding the center of the cluster. This algorithm also tries to minimize the objective function as in eq. (2). This is done by updating the centers of the cluster iteratively and membership functions are using the equations given below,

$$v_i = \left( \sum_{j=1}^N (u_{ij})^m \right) / \left( z_j \sum_{j=1}^N (u_{ij})^m \right) \quad (3)$$

$$u_{ij} = \frac{1}{\left( \sum_{k=1}^c \left( \frac{\|v_i - u_j\|}{\|v_i - u_k\|} \right)^{\frac{2}{m-1}} \right)} \quad (4)$$

Once the clustering is done, each of the pixels is assigned to the cluster for its membership value is having a maximum value.

Based on the distribution of the intensity obtained using the image histogram threshold value calculation is done and also taking the maximum mean of cluster 1 and minimum of the cluster 2 or the maximum of the cluster 2 and minimum of the cluster 3. The intensity distribution in the image is to get two clusters and taking into account by using this thresholding technique as in Fig. 5. From the input image only two clusters (a) Healthy and (b) unhealthy are created.

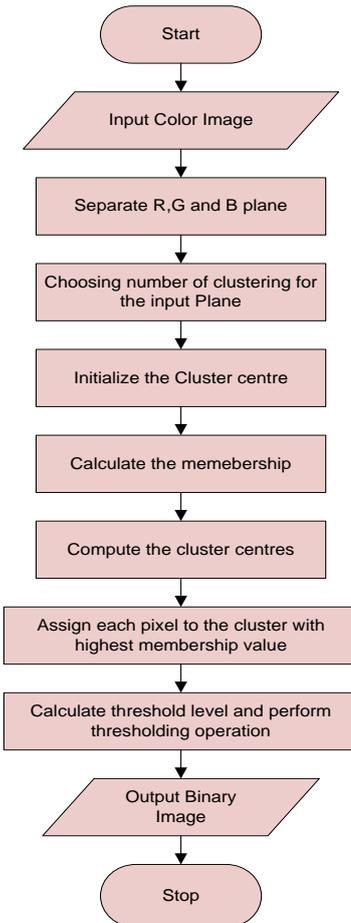


Fig. 4: Flow Chart for Fuzzy C Means Clustering

The algorithms flowchart is given in Fig. 4. Algorithm for this is given in Algorithm 1.

**Algorithm 1**

**Input:**  $Z = \{z_1, z_2, z_3, \dots, z_n\}$  set of data points and  $V = \{v_1, v_2, v_3, \dots, v_c\}$  set of centers.

**Output:** fuzzy membership matrix  $U$ .

- 1: Randomly select 'c' cluster centers.
- 2: **For**  $j = 1$  to  $c$ , fuzzy membership '  $u_{ij}$  ' using,

$$u_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij} / d_{ik})^{(2/m-1)}}$$

- 3: Calculate the fuzzy centers ' $v_j$ ' using:

$$v_j = \left( \sum_{i=1}^n (u_{ij})^m z_i \middle| \sum_{i=1}^n (u_{ij})^m \right), \quad \forall_j = 1, 2, \dots, c$$

- 4: Repeat step 2 and 3 until the minimum 'j' value is achieved or  $\|u^{k-1} - u^k\| < \beta$

Where

'k' is the iteration step.

'β' is the termination criterion between [0, 1].

'U(u<sub>ij</sub>)<sub>n×c</sub>' Fuzzy membership matrix.

'J' is the objective function.

**5: End For**

Using this approach once the partitions of the healthy and unhealthy region are done; Unhealthy regions are then passed to Suspected Region Selection Based on Hierarchical Colour Segmentation.



(a)



(b)

Fig. 5: (a) Healthy Cluster; (b) Unhealthy Cluster

3.3 Suspected Region Selection Based on Hierarchical Colour Segmentation

Once the segmented clusters for healthy and unhealthy regions are obtained they are passed to suspect region block for separating the NPK regions. The proposed methodology is shown in Fig. 6.

A flowchart is shown in that initially unhealthy image is taken as input and the binary conversion is done on it. This binary converted image will contain the number of components which are connected for the defected part in it. The centroid for a component is calculated and that particular component's bounding box area is upsampled by scaling factor which is chosen initially during the pre-processing step.

TABLE 1: THE THRESHOLD FOR NPK REGIONS

	Red Plane	Green Plane	Blue Plane
<b>N</b>	40 to 90	40 to 70	30 to 50
<b>P</b>	130 to 160	110 to 150	40 to 70
<b>K</b>	100 to 160	60 to 110	20 to 60

This upsampled area is then located in the original image and that particular area is cropped in the original image. We call this approach as a hierarchical approach. As the original image field is very large in size, processing of this image will become difficult; hence the hierarchical technique is followed to simplify the segmentation approach.

These segmented blocks are then passed to color segmentation step for NPK separation based on NPK threshold values. Similar steps are carried out for the remaining connected components to segment NPK regions in all the blocks on the Original Image. The chosen Threshold for NPK region is as in Table 1. Every time they obtained NPK regions from every block is placed in different planes to get N, P and K planes separately. This segmented N, P, and K regions are again validated by comparing the features extracted from it to the features already stored in the knowledge base using ANN classifier to get the final Validated NPK say VN, VP and VK segments as shown in the Fig. 15, 16, 17. Healthy regions in N (HN), P (HP) and K (HK) are also obtained from these validated regions after the classification. This is acquired by subtracting VN, VP, and VK from N, P and K.

### 3.4 Feature Extraction

The measurements of one or more functions are called features. The object has a quantifiable property and that is specified by these features and only significant information's are later picked from these features. In this work, the methods for feature extraction are given below.

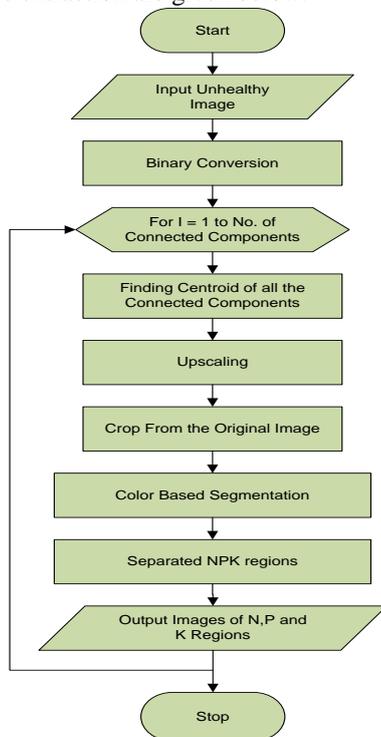


Fig. 6: Flowchart for Hierarchical Approach of Colour Segmentation

#### 3.4.1 Grey-Level Co-Occurrence Matrix

The basis on the statistical distribution of pixel in a statistical analysis of texture and characteristics of texture were calculated in a given position that is relative to others in a matrix consisting of pixels representing the image. Statistical parameters are considered for each combination depends on the pixel number, i.e., first, second or higher-order. Based on the GLCM the second order statistics are used by analyzing the image as a texture. GLCM approach is the tabulation of the frequency or how in an image often a combo of pixel brightness values occurs. The Fig.10 given below represents the GLCM formulation of the gray level of four levels in an image of distanced = 1 and the 0° direction.

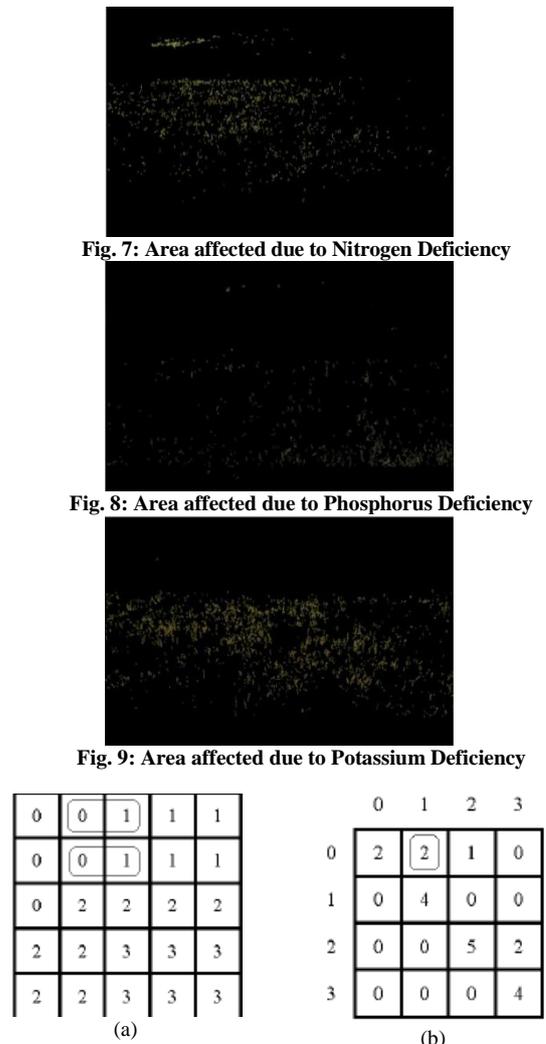


Fig. 10: (a) Example Matrix of the pixel intensity; (b) GLCM matrix

Fig. 10 (a) depicts the example matrix of the pixel intensity which represents the image with four levels of grey. 1 and 0 are the intensity levels are marked within a thin box. The thin box represents the pixel intensity 0 with the intensity of pixel 1 as its neighbor. There are usual occurrences of two pixels of such types. Hence the GLCM matrix is formed as shown in Fig. 10 (b) with value two in 0 rows, column 1. Similarly, matrix GLCM column 0 and row 0 is also has been given a value of two, because of two occurrences in which the pixel of value 0 has 0 pixels as its neighbor in the horizontal direction. Due to which the pixels matrix in (a) have been transformed into a GLCM as (b). It is not only to the horizontal direction, but GLCM also formed for different directions like 90°, 45° and 135° as depicted in Fig. 11 [20].

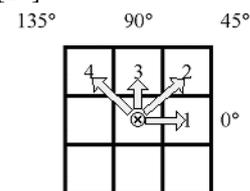


Fig. 11: Different Directions Formed by GLCM

From centre to the pixel 1 representing direction = 0° with distance 'd' = 1, to the pixel 2 direction = 45° with distance 'd' = 1, to the pixel 3 direction = 90° with distance d = 1, and to the pixel 4

direction = 135° with distance 'd' = 1. Even though the co-occurrence matrices extract the properties of texture, it's not directly used as an analysis tool. This matrix is again extracted to fetch the numbers that are usually used for texture classification.

Mathematically, given for an image  $I(N, P, K \text{ Regions})$  of size  $K \times K$  ( $256 \times 256$ ), the elements of a  $G \times G$  ( $8 \times 8 \times 4$ ) co-occurrence matrix of gray-level MCO for a displacement vector  $d = (dx, dy)$  is defined as in eq. (5). The defined  $d$  value for angle 0 is  $[0 \ d]$ , for angle 45 is  $[-d \ d]$ , angle for 90 is  $[-d \ 0]$  and for degree 135 is  $[-d \ -d]$ .

$$M_{co} = \sum_{x=1}^k \sum_{y=1}^k \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + d_x, y + d_y) = j \\ 0 & \end{cases} \quad (5)$$

In the proposed work four important features like contrast, energy, entropy, and Correlation are extracted. Contrast calculates the level of intensity difference between the adjacent pixels in a considered image and is given as follows by the equation below,

$$Contrast = \sum_{i,j} |i - j|^2 p(i, j) \quad (6)$$

Where  $p(i, j)$  denotes the position of the GLCM in that the value represents the sum of co-occurrence between adjacent pixels of  $i$  and its neighbor  $j$ .

Correlation measures the dependency level between pixels against the remaining pixels in the input image. Also, Correlation measures the linear dependency of the grey levels of neighbouring pixels and is it is given by,

$$Correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (7)$$

Energy measures the summation of the squared element in the entire GLCM. It is denoted by the eq. (8),

$$Energy = \sum_{i,j} p(i, j)^2 \quad (8)$$

The information which compresses the considered image and it contains the loss of information from the image for GLCM calculation is referred to as entropy and it is given as,

$$Entropy = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} -p_{ij} * \log p_{ij} \quad (9)$$

The formulation and extraction of the input NPK segments are extracted using MATLAB for calculating GLCM to get the feature vector of about 88.

### 3.4.2 HSV Histogram

In this histogram approach, HSV color space is considered for histogram generation, where each pixel contributes either its Hue or its Intensity. Extraction of the color histogram as the feature is extracted vector contains two steps: 1) Hue range representation between 0 to  $2\pi$  which is quantized after a transformation: 2) To obtain the quantized set of gray values [21]. Extracted feature vector components based on Hue is calculated using the equation given below,

$$N_h = [2\pi QL + 1] + 1 \quad (10)$$

QL represents the quantization level for the Hues. The QL value taken is 8. Number of Components denoting the gray value is given by the equation below,

$$N_g = \left\lceil \frac{I_{\max}}{QGL} \right\rceil + 1 \quad (11)$$

$I_{\max}$  Denotes the maximum intensity value, usually, 255 is chosen and QGL denoted the number of levels of gray

which are quantized. Usually, QGL is chosen as 16 [22]. The quantized value of the Hue always ranges from 0 to  $2\pi$ . Also, both the endpoints will be red. Thus the feature vector obtained is thus a combination of two different vectors of size 32. These extracted 32 features are then passed to ANN block.

### 3.5 ANN Classifier

The obtained 88+ 32 features are then trained to ANN every time when a query image is passed. The neural network model is a branch of the artificial intelligence which is generally referred to as Artificial Neural Network (ANNs) [23], [24]. This model can be precisely and quickly find the buried patterns in the data that would replicate the useful Knowledge. The artificial neural network usually did the many artificial neurons which are correlated together with the explicit neural network. As ANN consider classification is the most dynamic research in different areas of application. On individual output nodes, the desired signals are given by an external teacher in the supervised learning approach.

#### 3.5.1 Perceptron Learning

It is another form of a neural network and the weights and the bias are trained for producing a correct target vector. The Technique used here for learning is called perceptron learning. They are likely suitable for simple problems in pattern classification.

#### 3.5.2 Backpropagation Learning

Only linearity independent problems or linearity separable problems are handled by simple perceptron. The partial derivative of the network error is taking with respect to weight, here we learn little about the direction of the error moving in the network. The error rate has changed the value of the weights i.e. the negative of the derivative if we take then proceeding for adding it to the weights make the error decrease until it reaches local minima. It tells us that the error is increasing when the weight is increasing. And adding negative value to the weight is a thing and vice Versa if in case of the negative derivative.

When you consider that the taking of those partial derivatives after which applying them to every of the weights takes position, establishing from the layer of output to hidden layer weights, then the hidden layer to input layer weights because it turns out, that is integral on the grounds that changing these set of weights requires that we all know the partial derivatives calculated within the layer downstream, this algorithm is known as the backpropagation algorithm. Neural network training different modes of are followed: Batch and online mode. The weights of number which updates the two approaches for the equal number of information shows it is very difficult. The online process updates of weight are calculated for each and every input data pattern and the weights are modified after each sample. A replacement answer is to evaluate the weight replaces for each and every input sample, but store these values during one pass by means of the training set has been referred to as an epoch.

At the epoch ends, the addition of all the contributions, and weights might be made up to date with the composite value.

This process adapts the cumulative update with weights and it is known as the batch-training mode. Training involves training samples feeding as input vectors through a neural network, error calculation of output layer and weight adjustment to minimize error in the network. The ANN used in the proposed work containing 3 layers. A number of neurons are used is 120 in the input layer, 60 in the middle layer and 4 in the output layer. The backpropagation approach of neural training is followed in our proposed work for classification. The working flow of this approach is explained in Algorithm 1 given below,

*Steps of the backpropagation algorithm are:*

We are assuming that dealing with a network that's having a single input and a single output unit.

*Algorithm 1 Backpropagation algorithm.*

A network is considered with a single real input  $x$  and network function  $F$ . The derivative  $F'(x)$  is calculated in two different phases as given below,

*Feed-forward:* The input  $x$  initially is fed into the BPNN [25] network. The primitive functions of the nodes and their derivatives are calculated at every node. The derivatives are then stored.

*Backpropagation:* The constant 1 is fed into the output unit and then network run backward. The information is incoming from a node is then added and the result is multiplied by the stored value in the left part of the unit. The transmission of the result is done to the left of the unit. The result collected at the input unit is the derivative of a network function with respect to  $x$ .

### 3.6 Yield Estimation

Once the validation on NPK and Healthy segmented regions are done, they are passed to yield estimation block. In the



Fig. 12: Input Paddy Field Image



Fig. 13: Manually Cropped Image



Fig. 17: VK Region



Fig. 18: Total Yield Estimated for the Input Image



Fig. 19: Closer View of the Unhealthy Area in Field Image in Fig. 12



Fig. 20: VN Region in the Closer View of Field Image in Fig. 12

Yield, Estimation block Total yield on Validated Healthy Image (H) is done by dividing it with Input Paddy Crop Image. The total area (H) is added with NH, PH, KH regions. Where NH is the healthy region obtained after the validation of the N region using ANN classifier, similarly KH and PH regions. Calculating Total Yield in terms of percentage as in eq. (12),

$$Total\ Yield = \frac{H + NH + PH + KH}{Total\ Paddy\ Field\ Area} * 100 \quad (12)$$

## IV. EXPERIMENTAL RESULT

### Database

This proposed system works on five different databases, these filed images are captured by a different standard camera. Each filed image has its own property. Image analysis techniques are designed to improve the system performance in deficiency identification in field images.

### Experimental Set up

The proposed model is designed by using MATLAB Tool. The whole experimental setup is partitioned in two parts i.e. Training and Testing Phase. In the training phase, five different data sets are trained and stored as a knowledge base. In the testing phase, the real-time paddy input image is considered as input. Preprocessing, Clustering and Segmentation algorithms are used to analyze the input clearly. Further techniques for feature extraction are designed to compare the input field image with a knowledge base using ANN decision-based Classifier.



Fig. 14: (a) Healthy Region



Fig. 14: (b) UnHealthy Region



Fig. 15: VN Region



Fig. 16: VP Region



Fig. 21: VP Region in the Closer View of Field Image in Fig. 12



Fig. 22: VK Region in the Closer View of Field Image in Fig. 12

Detailed results obtained at every stage of a proposed system are explained below. Initially, the input paddy field image as shown in Fig. 12 is taken and is passed to the pre-processing block. Once the pre-processing steps like noise removal and resizing are done of an input image they are

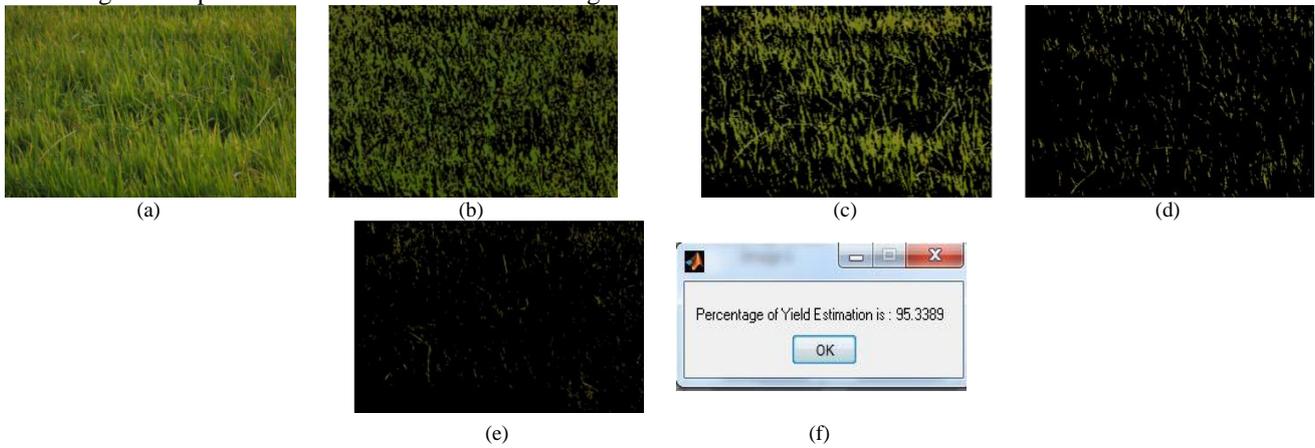
checked on whether the Input Field image contains the unwanted area.

## An Effect of Nutrient Deficiency on Yield Estimation

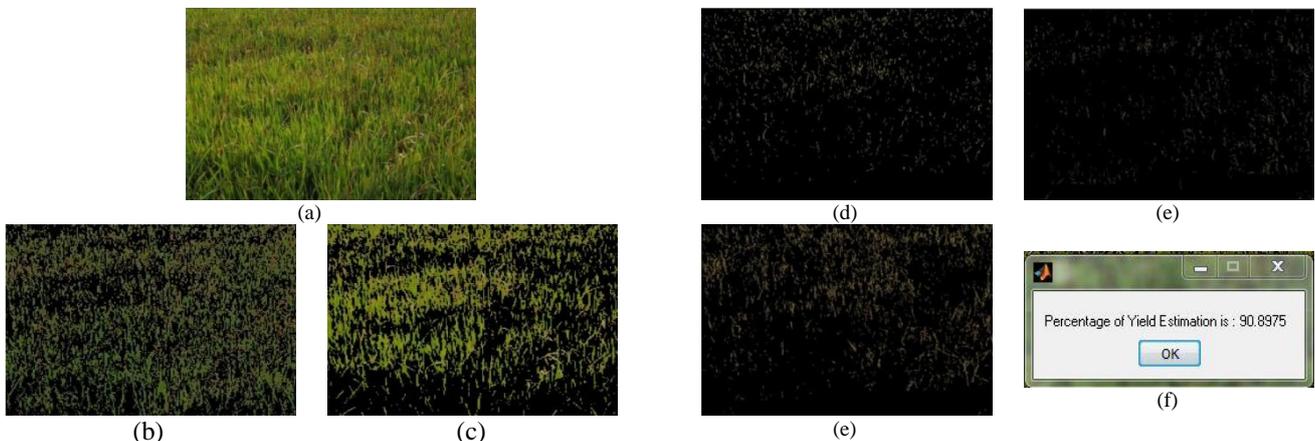
If yes the unwanted area is cropped from that image as shown in Fig. 13 then they are subjected to clustering by using FCM. The clustering output which gives the clusters consisting of healthy regions and unhealthy regions as shown in Fig. 14 (a) and (b). The unhealthy region obtained in this stage is then passed to Hierarchical Color Segmentation block for proper NPK region separation based on threshold values. These segmented regions may contain regions which are healthy or NPK segmentation may not be carried out properly, so to validate this features like GLCM and GLRLM are extracted and is passed to ANN classifier for Comparing it with the features already stored in the Knowledgebase obtained from extracting features from the Manually cropped NPK segments. The obtained validated Segments VN, VP, and VK are as shown in Fig. 15, 16 and 17.

To see the obtained results in close view, the more area defects in the Field image of Fig. 12 as shown in Fig. 19 is picked. Fig. 20 depicts the Validated Defected N region

(VN) obtained in the area inside the red rectangular block of Fig. 19. Similarly Fig. 26 depicts the Validated Defected P region obtained (VP); Fig. 27 depicts the Validated Defected K (VP) region obtained. Fig. 23 (a) depicts the Input Paddy Field Image. The Obtained Healthy and Unhealthy region are given in Fig. (b) and (c). As the Image contains more of affected areas by N and K regions, the obtained N defected and K defected area is as shown in the Fig. (d) and (e). The total estimated yield for this image is about 95.338% as in (f). Fig. 24(a) depicts the Input Paddy Field Image and Obtained Healthy and Unhealthy region is given in Fig. (b) and (c). As the Image contains areas affected by N, P and K regions, the obtained N defected P defected and K defected area is as shown in the Fig. (d), (e) and (f). The total estimated yield for this image is about 90.89% as in (g). Similarly Fig. 25 (a), (b), (c), (d), (e), (f) and (g) are the Input, Healthy, Unhealthy, N Region, P Region, K region and Total Yield Estimated for the input image (a).



**Fig. 23 (a) Input; (b) Healthy Image; (c) Unhealthy Image; (d) N Defected Region; (e) K Defected Region; (f) Estimated Yield.**



**Fig. 24: (a) Input; (b) Healthy Image; (c) Unhealthy Image; (d) N Defected Region; (e) P Defected Region; ; (f) K Defected Region; (f) Estimated Yield**

Table 2 Depict the Comparison table for the existing and proposed System. The table shows that the proposed system gives good Result when compared to the existing systems.

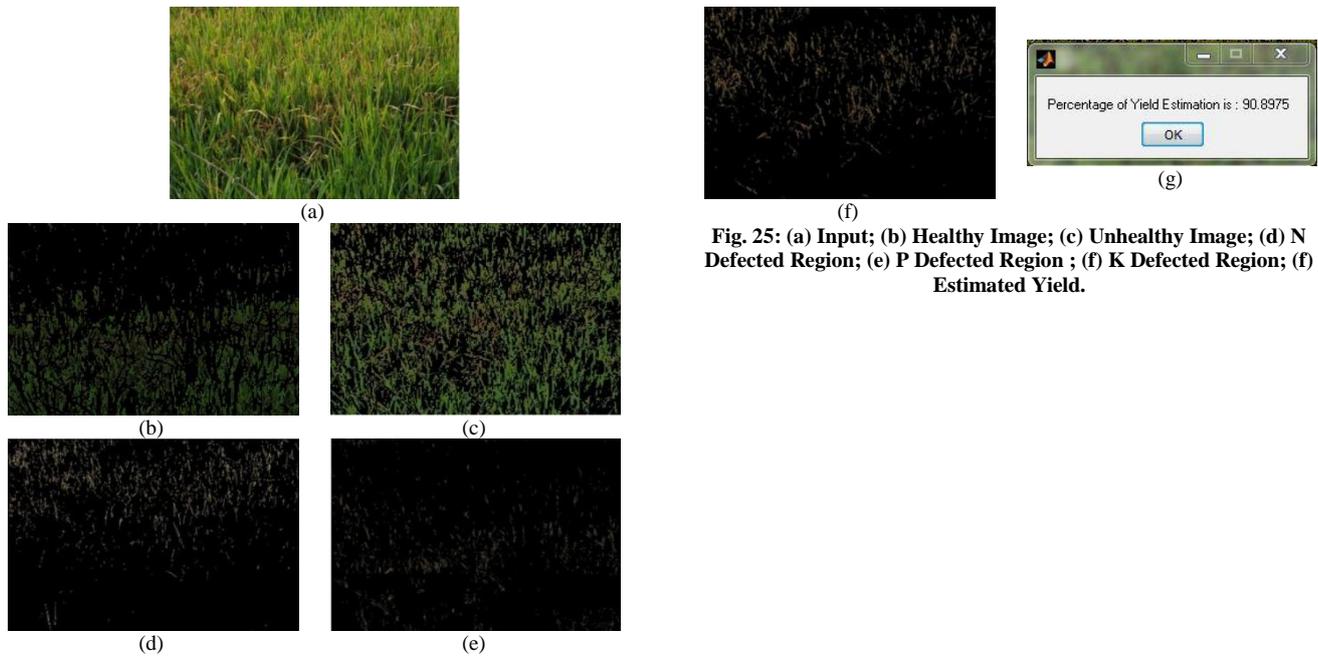
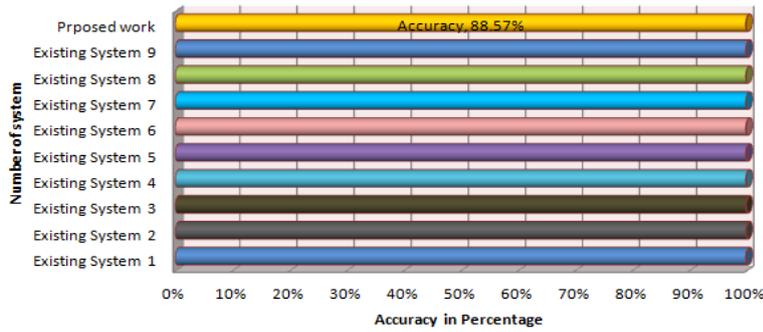


Fig. 25: (a) Input; (b) Healthy Image; (c) Unhealthy Image; (d) N Defected Region; (e) P Defected Region ; (f) K Defected Region; (g) Estimated Yield.

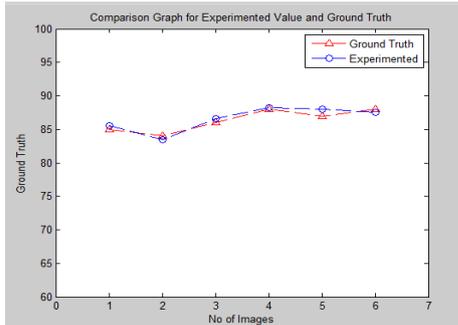
TABLE 2: COMPARISON TABLE FOR PROPOSED AND EXISTING SYSTEMS

Paper	Dataset	Yield Prediction Accuracy
The yield of Crop estimation model for Iowa using remote sensing and surface parameters (Existing System 1) [11].	The algorithm is worked on Normalized difference vegetation index (NDVI), vegetation condition index (VCI) and temperature condition index (TCI) for Iowa, Soybean, Corn Crops	78%
Yield assessment and Rice crop monitoring with MODIS 250m gridded Vegetation product: a case study in Sa Kao Province, Thailand (Existing System 2) [26]	The algorithm is evaluated utilizing the MODIS 250m Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) time-series data, crop calendar information and field data.	86%
Evaluation of a Process-based Agro Ecosystem Model (Agro-IBIS) across the U.S. Corn Belt: Simulation of the Interannual Variability in Maize Yield(Existing System 3) [27]	Model validation and Interannual Variability	36%
Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. Agr Forest Meteorol (Existing System 4) [28]	GCM predictor selection and rainfall, Hindcasts, Maize hindcast scenarios, and k-nearest neighbor-weighted analogs	36%
Efficiency change in Thailand rice production: Evidence from panel data analysis(Existing System 5) [29]	Time-varying production frontier model, Inefficiency model and Determinants of technical efficiency	85.67%
Research on Rice Acreage Estimation in Fragmented Area based on Decomposition of Mixed Pixels(Existing System 6) [30]	Extraction of non-arable mask images, Pure-pixel rice mapping, Rice acreage estimation, Sub-pixel rice mapping and Survey data for validation of rice acreage results	83.74%
Rice Crop Monitoring and Yield Estimation through Cosmo Skymed and Terrasar-X: A Sar-based experience in India(Existing System 7) [31]	Basic Processing of SAR Data for Multi-Temporal Analysis, Rice Map Accuracy Assessment, Rice yield estimation, and Multi-Temporal $\sigma^0$ Rule-Based Rice Detection	87%
Application of Artificial Neural Network in Predicting Crop Yield: A Review(Existing System 8) [9]	ANN in Predicting Crop Yield, ANN in Soil and Soil-Plant Hydrology, ANN in Sensing Technologies and Backpropagation algorithm.	47%
The behavior of HSI Color Co-Occurrence Features in Variety Recognition from Bulk Paddy Grain Image Samples(Existing System 9) [32]	Recognition Using Reduced Color Texture Features, Recognition Using Color Texture Features, co-occurrence matrix method, SGLDM, GLCM, and backpropagation algorithm.	78- 84% (considering mid-value i.e. 81%)
Proposed system	The algorithm is evaluated on Taking Paddy Field Images of different Locations collected by us.	88.57%

## Comparison Graph



**Fig. 26: Comparison Graph for Existing and Proposed System**



**Fig. 17: Comparison Graph for Ground Truth and Experimented Values obtained from the Proposed System.**

Fig. 26 depicts the Comparison Graph for Proposed and the Existing Systems. The proposed system provides good results when it is compared to the existing systems. Fig. 27 gives the comparison graph for the Ground Truth value for the Total Prediction Accuracy to the Proposed Values. The ground truth values are collected according to the farmer's prediction of the total yield of the field based on his experience. The proposed system gives good experiment results when they are compared to the existing system.

The confusion matrix comprises knowledge regarding actual and predicted values. In this system, the actual and predicted values are categorized. The difficulty in statistical categorization and even machine learning field could overcome by this matrix method and is also known as the error matrix. This is a precised table design which permits the idea of the performance of an algorithm. The confusion matrix is supervised learning and the unsupervised learning is defined by matching matrix. Every column in the matrix is predicted class and row is actual class. It is a two-dimensional unique sort of possibility table having the same sets of classes from each dimension. The matrix is a test data set with true identified values. This matrix is moderately easy to know but the associated terms are confusing. The associated terms are defined as follows.

- True positive (TP)
- True negative (TN)
- False-positive (FP)
- False negative (FN)

For considering any data set the confusion matrix can be evaluated by above four conditions. True positive always predicted by yes, true negative always predicts no, false-positive predicts yes but actually it is not and false-negative predicts no but actually it is yes. Along with this row and column, totals are added. The false-positive and false-negative is also called type 1 and type 2 error.

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (18)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (19)$$

$$Specificity = \frac{TN}{FP + TN} \quad (20)$$

Totally five different data set considered in proposed experimental work. Individual data set for corresponding system performance is explained by using the confusion matrix as represented in below Table 3. Performance accuracy of each data set is represented in Fig. 28. The designed system and its performance are analyzed by using Precision, Recall, Accuracy, Sensitivity and Specificity measurement parameter of the confusion matrix. System performance parameters i.e. Precision, Recall or Sensitivity, Specificity of the proposed system is listed in Table 4. The mathematical equations used in parameter calculation are shown below Eq. (16), to Eq. (20). Fig. 29 represent the performance graph of Precision, Recall, and Specificity of different data set. The proposed systems performance graph is shown in Fig. 30.

**TABLE 4: PERFORMANCE ANALYSIS OF THE PROPOSED USING CONFUSION MATRIX USING DIFFERENT DATASET**

Database	Number Images worked	Confusion Matrix			System Accuracy
Data set 1	35	N = 35	Predicted: No	Predicted: Yes	87.5%
		Actual: No	TN = 8	FP = 2	
		Actual: Yes	FN = 2	TP = 23	
Data set 2	32	N = 32	Predicted: No	Predicted: Yes	89.47%
		Actual: No	TN = 7	FP = 2	
		Actual: Yes	FN = 2	TP = 21	
Data set 3	38	N = 38	Predicted: No	Predicted: Yes	86.48%
		Actual: No	TN = 10	FP = 2	
		Actual: Yes	FN = 2	TP = 24	

**TABLE 3: TABLE: ACCURACY TABLE**

	Accuracy
Existing System 1	78 %
Existing System 2	86 %
Existing System 3	36 %
Existing System 4	36 %
Existing System 5	85.67 %
Existing System 6	83.74 %
Existing System 7	87 %
Existing System 8	47 %
Existing System 9	81 %
Proposed System	88.57 %

Data set 4	37	N = 37	Predicted: No	Predicted: Yes	89.47%
		Actual: No	TN = 12	FP = 3	
		Actual: Yes	FN = 2	TP = 20	
Data set 5	38	N = 38	Predicted: No	Predicted: Yes	87.8%
		Actual: No	TN = 6	FP = 2	
		Actual: Yes	FN = 2	TP = 28	
Including all Data set	180	N = 180	Predicted: No	Predicted: Yes	88.57%
		Actual: No	TN = 43	FP = 11	
		Actual: Yes	FN = 10	TP = 113	

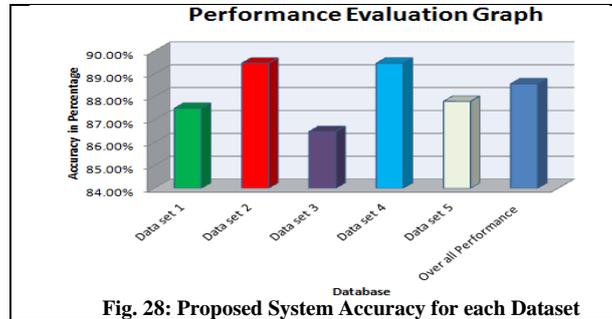


TABLE 5: PERFORMANCE ANALYSIS TABLE FOR THE DIFFERENT DATA SET

Database	Number of Images worked	Precision	Recall or Sensitivity	Specificity
Data set 1	35	0.913	0.913	0.777
Data set 2	32	0.923	0.923	0.833
Data set 3	38	0.8695	0.909	0.88
Data set 4	37	0.9335	0.933	0.75
Data set 5	38	0.908	0.916	0.806
Total Number of Images = 180		0.92	0.92	0.8

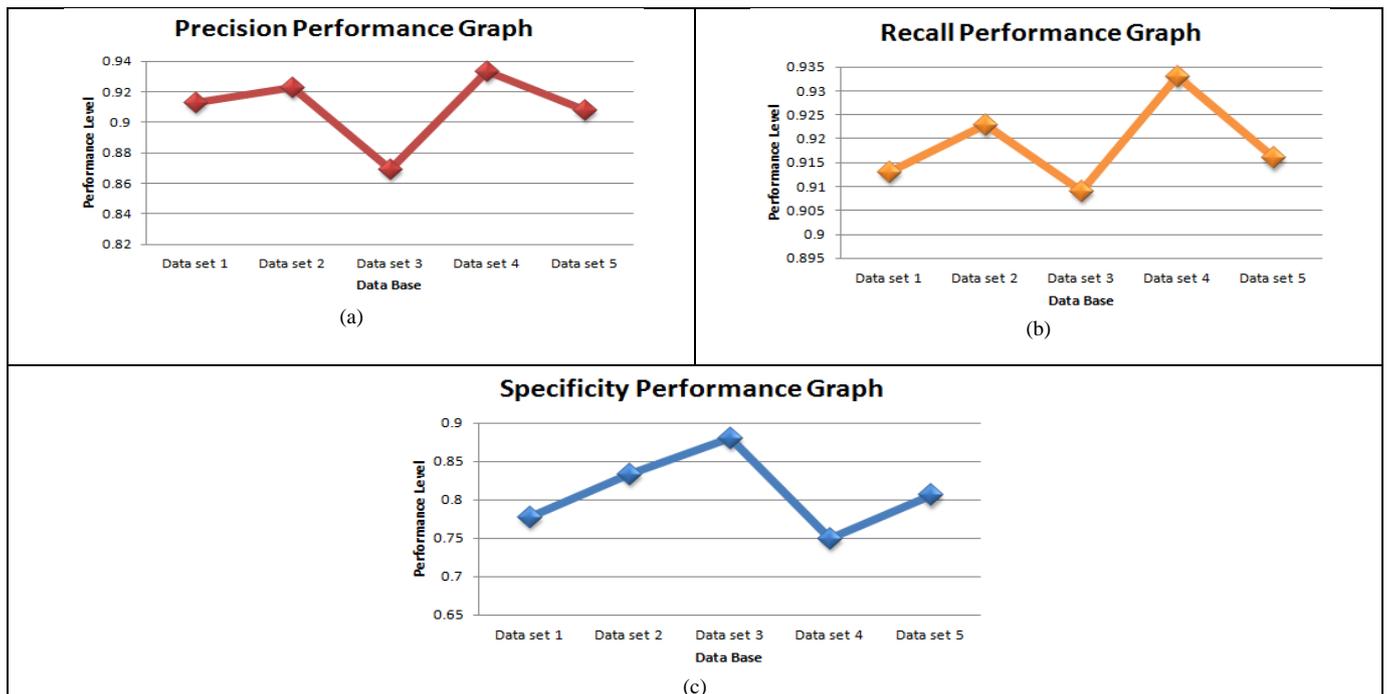


Fig. 29: (a) Precision Performance Graph; (b) Recall Performance Graph; (c) Specificity Performance Graph

# An Effect of Nutrient Deficiency on Yield Estimation

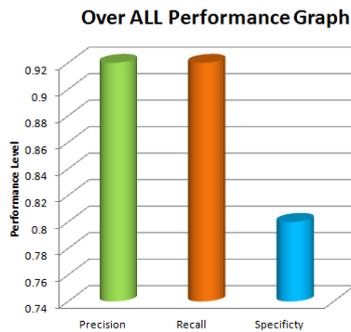


Fig. 30: Over All System Performance Graph

## V. CONCLUSION

In this paper an efficient method is proposed to segment the affected region in the paddy field image. This approach considers the 3 nutrient deficiency like NPK and segmented area affated by NPK. This approach makes use of efficient algorithms like FCM and Suspected Region Selection Based on Hierarchical Colour Segmentation approach along with ANN classifier. The algorithm is used for effectively calculating the Nutrient Deficiency, Effect on Yield Estimation and to evaluate the Total Yield in Input Paddy Field Image. Proposed methodology result indicates a systematic and a strong way of disease intensity assessing for total yield calculation in a more précised way. The algorithm is estimated on different Field images resulting in a good performance. Future work involves adding an efficient prediction algorithm for Total Yield calculation in the Paddy Field Image.

## REFERENCE

1. S. Sridevi and Dr. Anna Saro Vijendran, "An Evolving Expert System for Maize Plant Nutrient Deficiency using Image Processing Technique", International Journal of Science and Research (IJSR), Vol. 5, 2014.
2. Hiteshwari Sabrol and Satish Kumar, "Recent Studies of Image and Soft Computing Techniques for Plant Disease Recognition and Classification", International Journal of Computer Applications, Vol. 126, No. 1, 2015.
3. Manoj Mukherjee, Titan Pal, and Debabrata Samanta, "Damaged Paddy Leaf Detection Using Image Processing", Journal of Global Research in Computer Science, Vol. 3, No. 10, 2012.
4. S. A. Ramesh Kumar and Dr. K. Ramesh Kumar, "A Study On Paddy Crops Disease Prediction Using Data Mining Techniques", International Journal of Data Engineering (IJDE), Vol. 7, No. 1, pp. 336 - 347, 2015.
5. Asif Ali, Jens.C. Streibig, Svend Christensen and Christian Andreasen, "Estimation of weeds leaf cover using image analysis and its relationship with fresh biomass yield of maize under field conditions", HAICTA, pp. 41- 49, 2011.
6. Manickam Gopperundeivi and Vijayaraghavan Kannan, "Paddy Yield Estimation Using Remote Sensing and Geographical Information System", Journal Of Modern Biotechnology, Vol. 1, Issue 1, pp. 26 - 30, 2012.
7. Wijesingha, J. S. J., N. L. Deshapriya, and L. Samarakoon, "Rice Crop Monitoring and Yield Assessment with MODIS 250m Gridded Vegetation Products: A Case Study of Sa Kaeo Province, Thailand", The International Archives of Photogrammetry, Vol. 40, Issue 7, 2015.
8. Mrs. K.R. Sri Preethaa M.E, "Crop Yield Prediction", International Journal of Engineering Technology and Sciences, Vol. III, Issue III, 2016.
9. Siti Khairunniza-Bejo, Samihah Mustaffha, and Wan Ishak Wan Ismail, "Application of Artificial Neural Network in Predicting Crop Yield: A Review", Journal of Food Science and Engineering, Vol. 4, pp. 1 - 9, 2014.
10. Shreya S. Bhanose, Kalyani A. Bogawar, Aarti G. Dhotre, and Bhagyashree R. Gaidhani, "Crop and Yield Prediction Model", International Journal Of Advance Scientific Research And Engineering Trends, Vol. 1, Issue 1, 2016.

11. Anup K. Prasad, Lim Chai, Ramesh P. Singh, and Menas Kafatos, "Crop yield estimation model for Iowa using remote sensing and surface parameters", Elsevier, Vol. 8, pp. 26 - 33, 2006.
12. Andres Vina, Anatoly A. Gitelson, Anthony L. Nguy-Robertson, and Yi Peng, "Comparison of different vegetation indices for the remote assessment of green leaf area index of crops", Elsevier, Vol. 115, pp. 3468-3478, 2011.
13. Manickam Gopperundeivi and Vijayaraghavan Kannan, "PaddyYield Estimation Using Remote Sensing and Geographical Information System", Journal Of Modern Biotechnology, Vol. 1, Issue 1, pp. 26 - 30, 2012.
14. David B. Lobella and MarshallB. Burke, "On the use of statistical models to predict crop yield responses to climate change", Vol. 150, Issue 11, 2010.
15. Rajesh S. Sarkate, Khanale P.B and Thorat S. B, "BPNN Approach in Pixel Classification based Precision Segmentation for Agriculture Images", International Journal of Computer Applications, pp. 25 - 27, 2014.
16. V. Radha Krishna Murthy, "Crop Growth Modelling and Its Applications In Agricultural Meteorology", Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, pp. 235-261, 2004.
17. Kishore C. Swain and Qamar Uz Zaman, "Rice Crop Monitoring with Unmanned Helicopter Remote Sensing Images", INTECH Open Access Publisher, 2012.
18. Ray and Nilanjan, "Median filter with absolute value norm spatial regularization", IEEE, Pp. 1437 - 1440, 2011.
19. Ammara Masood and Adel Ali Al-Jumaily, " Fuzzy C Mean Thresholding based Level Set for Automated Segmentation of Skin Lesions", Journal of Signal and Information Processing, Vol. 4, pp. 66-71, 2013.
20. Aswini Kumar Mohanty, Swapnasikta Bebertha and Saroj Kumar Lenka, "Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram", International Journal of Engineering Research and Applications (IJERA), Vol. 1, Issue 3, pp. 687 - 693, 2011.
21. Shamik Sural, Gang Qian and Sakti Pramanik, "Segmentation And Histogram Generation Using The Hsv Color Space For Image Retrieval", Vol. 2, pp. 589 - 592, 2002.
22. M. Bertini, A. Del Bimbo, C. Torniai, C. Grana, R. Vezzani and R. Cucchiara, "Sports Video Annotation Using Enhanced HSV Histograms in Multimedia Ontologies", pp. 160 - 170, 2007.
23. Saravanan K and S. Sasithra, "Review On Classification Based On Artificial Neural Networks", International Journal of Ambient Systems and Applications (IJASA), Vol. 2, No. 4, 2014.
24. Mohd Adzhar Abdul Kahar, Sofianita Mutalib, and Shuzlina Abdul-Rahman, "Early Detection and Classification of Paddy Diseases with Neural Networks and Fuzzy Logic", Conferences, Malaysia, 2015.
25. G. Venkata Narasimhulu and Dr. S. A. K. Jilani, "Back Propagation Neural Network-based Gait Recognition", International Journal of Computer Science and Information Technologies, Vol. 3, Issue 5, pp. 5025 - 5030, 2012.
26. J. S. J. Wijesingha, N. L. Deshapriya, and L. Samarakoon, "Rice Crop Monitoring and Yield Assessment with Modis 250m Gridded Vegetation Product: A Case Study in Sa Kaeo Province, Thailand, The International Archives of the Photogrammetry, Vol. XL-7/W3, 2015.
27. Christopher J. Kucharik, "Evaluation of a Process-based Agro Ecosystem Model (Agro-IBIS) across the U.S. Corn Belt: Simulation of the Interannual Variability in Maize Yield", Earth Interactions, Vol. 7, pp. 7 - 14, 2003.
28. J.W. Hansen and M. Indeje "Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. Agr Forest Meteorol", Agricultural and Forest Meteorology, Elsevier, Vol. 125, pp. 143 - 157, 2004.
29. Orawan Srisompun and Somporn Isvilanonda, "Efficiency change in Thailand rice production: Evidence from panel data analysis", Journal of Development and Agricultural Economics, Vol. 4, pp. 101 - 108, 2012.
30. Huanxue Zhang, Qiangzi Lia, Fan Lei, Xin Du, and Jide Wei, "Research on Rice Acreage Estimation in Fragmented Area based on Decomposition of Mixed Pixels", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 7, pp. 134 - 139, 2015.
31. S. Pazhanivelan, P. Kannan, P. Christy Nirmala Mary, E.Subramanian and S. Jeyaraman, "Rice Crop Monitoring and Yield Estimation through Cosmo Skymed and Terrasar-X: A Sar-based experience in India", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 7, pp. 85-92, 2015.
32. Basavaraj S. Anami, Naveen N. M and N. G. Hanamaratti, "Behavior of HSI Color Co-Occurrence Feature in Variety Recognition from Bulk Paddy Grain Image Samples", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, pp. 19 -30, 2015.