Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

# Assessing Impact of Nitrogen Deficiency on Paddy Field Yield Estimation: A Hierarchical Segmentation and SVM Approach

**Sharanamma M. Hugar**
*Research Scholar, Dept. of CSE, VTU CPGS RO kalaburagi, Affiliated to Visvesvaraya Technological University, Belagavi – 590018, INDIA*
**Dr. Mohammed Abdul Waheed**
*Professor, Department of CSE, VTU CPGS, RO, Kalaburagi., Affiliated to Visvesvaraya Technological University, Belagavi – 590018, INDIA*

***Abstract -*** In the agriculture sector, it's crucial to provide detailed explanations and methodically work towards predicting crop yields. This involves making informed decisions to enhance the quality of the analysis. Crop yield largely depends on the health of the crops, influenced significantly by key nutrients like nitrogen (N). A lack of nitrogen can lead to yellowish fields, potassium deficiency might result in leaf blotches, and phosphorus scarcity can turn fields brownish. Identifying these nutrient-deficient areas in paddy fields is a major challenge in estimating total yield. To address this, we use an efficient hierarchical model to segment these problem areas accurately. This approach has demonstrated impressive results in system accuracy.

***Keywords*** *- GLCM, Hierarchical Colour Based Segmentation, Nitrogen Regions, Support Vector Machine.*

## 1 Introduction

Integrating modern technology with traditional farming practices is increasingly crucial for enhancing agricultural productivity. As technology evolves, sophisticated crop models and predictive tools are becoming vital for precision farming. Accurate crop yield prediction is essential for farmers, agricultural bodies, and consultants, given that complex interactions among air, water, soil, and plants influence crop yield. To effectively model these interactions, advanced engineering methods are necessary.

Such models are invaluable for land managers and policymakers, especially when they must extrapolate outcomes from one location to another without specific response data. Various factors, both direct and indirect, affect crop production. Soil science, for example, examines aspects like pH, texture, nutrients, organic matter, fertilization, and farming practices. Agriculture is complex, with each issue requiring substantial data analysis.

Despite these challenges, current technology lacks comprehensive solutions for optimizing agricultural output and quality. One way to enhance these is by analyzing the total yield by segmenting field areas affected by nutrient deficiencies. Yellow areas in fields often indicate nitrogen deficiency, brown areas suggest potassium deficiency and brown leaf spots can signify phosphorus deficiency. This study focuses on these aspects to improve yield predictions.

## 2 Literature Survey

The field of agriculture has seen significant advancements through the integration of image processing and soft computing techniques in various research studies, each contributing unique methods and insights:

Hiteshwari Sabrol et al. focused on identifying and classifying plant diseases. Their method involved collecting images of healthy and diseased plants, preprocessing them, and segmenting the diseased areas. This approach classifies plant diseases and contributes to the academic understanding of plant pathology.

S. A. Ramesh Kumar et al. developed a strategy for identifying risk areas in paddy fields using image processing and data mining. Their research aimed to identify diseases and other factors affecting paddy production. The proposed solution increased efficiency and reduced the subjectivity of human experts in detecting faults in paddy. They used "associative rule mining" to categorize paddy characteristics.

Manickam Gopperundevi et al. created a method to estimate and map crop vigor. This involved using various factors to assess paddy fields' condition, supervised classification, and the Vegetation Index. Their study also explored the relationship between yield and the Normalized Difference Vegetation Index (NDVI), utilizing MODIS data to monitor rice crop growth, map the area, and measure productivity.

K.R. Sri et al. developed a technique for determining optimal crop yield and recommending the best crops for maximizing agricultural profitability and quality. Their research focused on enhancing agricultural output using environmental and land data. Farmers collected data on various factors, such as temperature, soil type, and water level, to select the most suitable crops for their soil. Their predictions relied on the accuracy of the Bayesian algorithm.

Additionally, the proposed plan included techniques like color segmentation-based nitrogen (N) area separation using hierarchical methods. After collecting the N-segmented areas, an SVM classifier was applied to identify whether it is an N-partitioned region. Finally, the total yield within the Paddy Field Image was estimated following the validation stage, exemplifying the potential of integrating advanced computational methods in agriculture.

## 3 Methodology

In the system depicted in Figure 1, a comprehensive approach involving training and testing phases is used to analyze paddy field images. During the testing phase, an image of the paddy fields is read and passed through a pre-processing block. This pre-processing includes scaling the image to the appropriate size and reducing noise, achieved using a median filter. After noise reduction, the image may still contain several irrelevant sections. These are manually eliminated by selecting the Region of Interest (ROI), effectively akin to cropping and downsampling the image to focus solely on noise-free areas. Once pre-processed, the image is applied to differentiate between healthy and unhealthy regions in the paddy fields. The segmented cluster from the markedly unhealthy region is then chosen and converted into a binary format, enabling hierarchical color segmentation. This step is crucial as it identifies potential nitrogen-deficient areas (N areas) through color thresholding, isolating actual N-deficient regions in the original paddy field image. This systematic process enhances the accuracy and efficiency of agricultural analysis, crop health, and management decision-making.

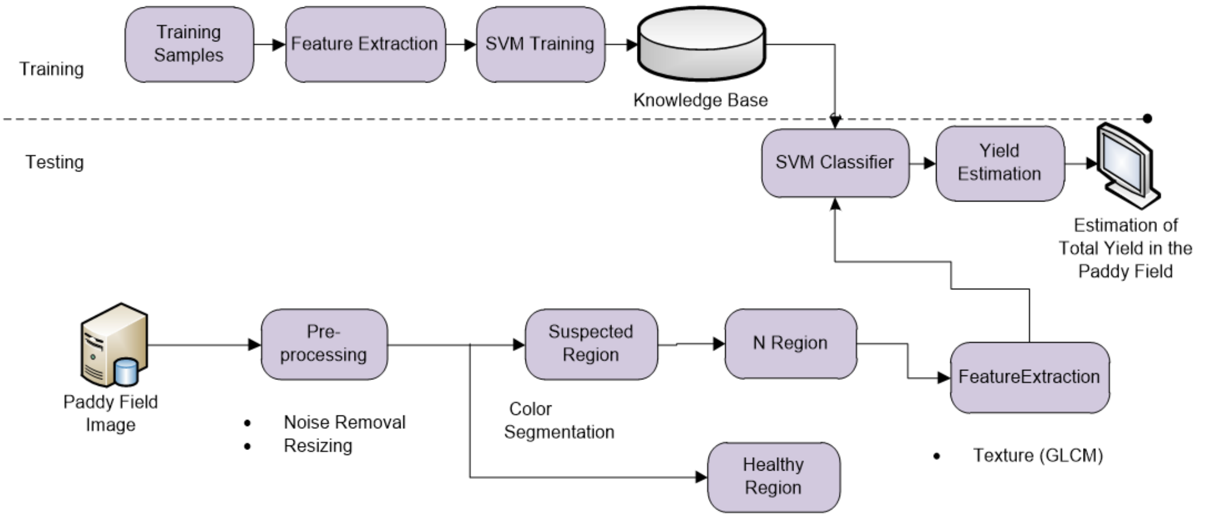



Fig. 1: Block Diagram of Proposed System

The feature extraction block is pivotal in the described system, particularly in extracting Gray Level Co-occurrence Matrix (GLCM) features from the segments and healthy zones identified in the paddy field images. This extraction process is a key component in the program's training and testing phases, as it involves analyzing both unhealthy and normal (healthy) crop samples.

Once the Support Vector Machine (SVM) training is complete, feature extraction proceeds similarly during the testing phase. The features extracted during testing and training are fed into an SVM classifier. This step is crucial as it ensures the accuracy of the classification process by effectively differentiating between healthy and unhealthy regions based on the extracted features.

After the classification process is concluded, the identified non-healthy (NH) regions, derived from the nitrogen-deficient (N) areas, are directed into the yield estimate block. This is an important step after the segmented N-deficient and healthy areas have been analyzed to confirm their health status. At this stage, the total yield in the paddy field image is calculated, providing a crucial insight into the impact of nutrient deficiencies on crop production.

By assessing and quantifying the effect of nutrient deficiencies, particularly nitrogen, on the paddy yield, this system offers a comprehensive approach to analyzing agricultural health and productivity. This allows for more informed decision-making in agricultural practices, potentially leading to improved crop management and yield optimization.

### 3.1 Pre-Processing

Figure 2 in the referenced documentation illustrates the process where the input image of a paddy field undergoes pre-processing analysis. A key component of this stage is the application of a median filter, specifically a weighted median filter, to the input image. The weighted median filter is a nonlinear digital filtering technique predominantly used for noise removal in image processing.

The effectiveness of this pre-processing step, particularly the use of the median filter, is noteworthy because it significantly improves the processing results. One of the notable advantages of the median filter, especially in the context of agricultural image analysis, is its ability to preserve the edges of features in the image while effectively reducing noise. This quality is crucial in maintaining the integrity of important details in the image, such as the boundaries between healthy and unhealthy regions of the crop.

In the process of applying this filter, the objective is to understand how the input image, denoted as I, influences the output of the weighted median filter, represented by v. The relationship between the input image and the filter's output is described by a specific formula, referred to here as Eqn (1). This equation likely details the mathematical operation of the weighted median filter, considering the intensity values of the pixels in the input image and applying the filter's weighting criteria to produce the filtered output. Understanding and applying this formula is essential for effectively reducing noise in the paddy field images, thereby enhancing the accuracy of subsequent analysis stages, such as feature extraction and disease classification.

$$\min_{v} \sum_{i \in V} \sum_{l \in N_i} w_{il} |v_i - I_l| \tag{1}$$

Fig. 2: Input Paddy Field Image

$N_i$ is the position of pixel $V$ in the picture. '$w_{ii}$' denotes the set containing pixel '$i$' and its surrounding pixels. The non-negative weight [12] is denoted.

### 3.2 Hierarchical Color Segmentation for Region Selection

In the process outlined in Figure 3, the system employs a suspicious region block to isolate nitrogen-deficient (N) regions in paddy fields, focusing on segmenting clusters indicative of unhealthy and healthy areas. This approach starts with a binary conversion of the diseased image, simplifying the data to highlight features associated with the disease or deficiency. Following this conversion, the system identifies and calculates the centroids of various components within the binary image, each representing a diseased portion of the crop. This step is crucial for determining the distribution and severity of the disease. Finally, the areas within the bounding boxes encapsulating these diseased components are upsampled. This upsampling involves applying a scaling factor determined during the pre-processing stage, enhancing the resolution of the specific areas for more detailed and accurate analysis. Following these steps, the system efficiently segments and analyzes nitrogen-deficient regions, facilitating precise disease detection and management in agricultural practices.

TABLE 1: THE THRESHOLD FOR N DEFICIENCY REGIONS

| | **Red Plane** | **Green Plane** | **Blue Plane** |
|---|---|---|---|
| **Nitrogen** | 43 to 85 | 34 to 67 | 32 to 54 |

In the described process, following the upsampling of a specific region in the primary image of a paddy field, the system locates this region and trims its edges. This step is part of what is known as a hierarchical approach, which is essential due to the complexity and size of the original image. The large field size of the original image makes processing challenging, thus necessitating the use of the hierarchical method to simplify the segmentation operation.

After segmenting the blocks, they are advanced to the color segmentation stage. They are divided into N groups based on N threshold values, as indicated in Table 1. This segmentation process also applies to other essential components found throughout the original image, segmenting these components into N distinct areas.

Once the N areas are defined, each block is placed on a separate plane, creating N-independent planes. These segmented N sections are then subject to verification through an SVM classifier. The classifier compares the features of the segmented areas with the previously recorded characteristics in the knowledge base. This comparison results in the generation of validated N or VN segments, as shown in Figure 4.

The process then focuses on identifying healthy zones in the N areas, referred to as HN (Healthy in N). This is achieved by subtracting the VN from the segmented image. This step is crucial for distinguishing between the areas affected by nitrogen deficiency and those not, ultimately contributing to a more accurate crop health assessment and aiding in effective agricultural management. The hierarchical method, along with the subsequent segmentation and classification steps, provides a comprehensive and detailed analysis of the paddy field image, enhancing the precision of agricultural practices.

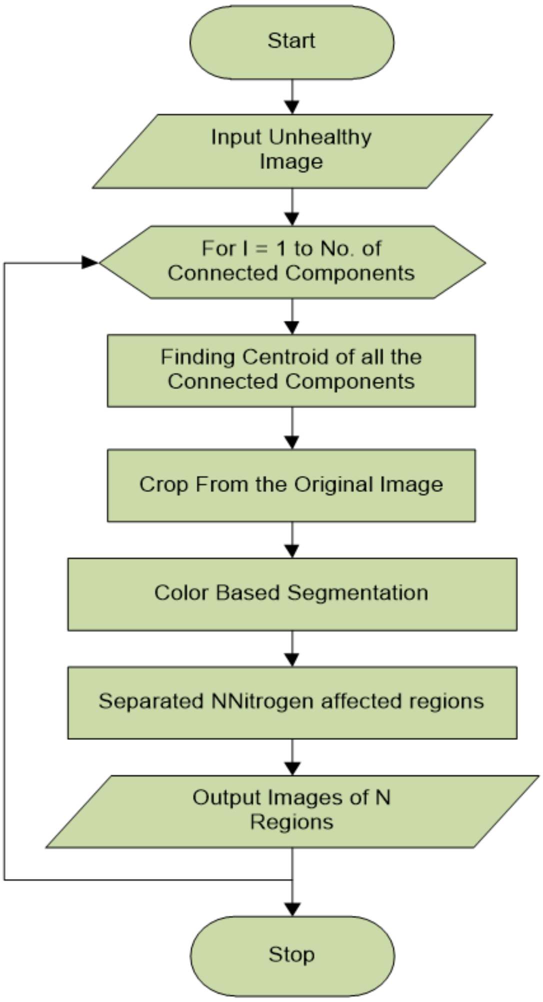



Fig. 3: Flowchart for Hierarchical Approach of Colour Segmentation

### 3.3 Feature Extraction

Quantitative evaluations, commonly referred to as features, are used to assess one or more characteristics of an object. These features represent the measurable attributes of the object, capturing only the crucial information. In the next section, the strategies used for feature extraction in this study are detailed, focusing on how these essential attributes are isolated and analyzed.

#### *3.3.1 Grey-Level Co-Occurrence Matrix*

In the field of statistical texture analysis, the process begins with selecting a specific location within the pixel matrix of an image. This chosen location is the starting point for determining the basis of pixel distributions and texture characteristics. The significance of each pixel combination in the image is determined by the number of pixels present in each combination. The analysis delves into first, second, and higher-order statistics to accurately represent the image's texture.

Second-order statistics based on the Gray Level Co-occurrence Matrix (GLCM) are essential for a detailed texture analysis. The GLCM technique tracks how frequently a particular arrangement of pixel brightness values occurs within the image. This method is crucial for understanding the spatial relationship between pixels, reflecting the image's texture.

Figure 5 demonstrates the GLCM formulation for four grey levels. In this specific example, the image is analyzed with a distance parameter d = 1 and an orientation of 0 degrees. This orientation and distance parameter are pivotal in determining how the pixel values are paired for the GLCM calculation, ultimately influencing the texture features extracted from the image. Such detailed texture analysis using GLCM is a fundamental part of image processing, especially in applications where the texture provides significant insights, such as in agricultural image analysis, medical imaging, and pattern recognition.

Fig. 4: Area affected due to Nitrogen Deficiency

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 1 |
| 0 | 0 | 1 | 1 | 1 |
| 0 | 2 | 2 | 2 | 2 |
| 2 | 2 | 3 | 3 | 3 |
| 2 | 2 | 3 | 3 | 3 |

(a)

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 2 | 2 | 1 | 0 |
| 1 | 0 | 4 | 0 | 0 |
| 2 | 0 | 0 | 5 | 2 |
| 3 | 0 | 0 | 0 | 4 |

(b)

Fig. 5: (a) Example Matrix of the pixel intensity; (b) GLCM matrix

The methodology described focuses on extracting key features from images, specifically contrast, energy, entropy, and correlation. To quantify the intensity difference between adjacent pixels in the analyzed image, an equation like the following is employed:

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

The term "value" in this context refers to the cumulative co-occurrence of adjacent pixels labeled 'i' and their neighboring pixel 'j,' where p(i,j) denotes the specific position in the Gray Level Co-occurrence Matrix (GLCM).

"Correlation," as used here, measures the extent to which individual pixels in an image are dependent on other pixels. It specifically analyzes the linear relationship between the grey levels of two adjacent pixels. This correlation is quantified by a specific formula, which assesses the inter-pixel relationship within the image, providing insights into the textural characteristics of the image based on how pixel values are linearly related.

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i \sigma_j} \tag{3}$$

The energy calculation in this context involves summing up the squared values of all elements in the Gray Level Co-occurrence Matrix (GLCM). An equation denotes Eq. (4), effectively quantifying the energy metric. In image texture analysis, the energy metric reflects an image's uniformity and texture regularity, with higher values indicating more uniform and consistent patterns. This calculation is a key part of analyzing the textural properties of images, especially in applications like medical imaging, remote sensing, and agricultural monitoring.

$$Energy = \sum_{i,j} p(i,j)^2 \tag{4}$$

Entropy, in the context of image analysis and GLCM (Gray Level Co-occurrence Matrix) computation, represents the information required to compress the analyzed image. It also accounts for the loss of image information during the GLCM computation process. Entropy quantifies the randomness or complexity in the image's texture, with higher entropy values indicating more complex or less predictable patterns. This metric is crucial for understanding the textural characteristics of an image, as it provides insights into the level of detail and variation present in the image's pixel intensity distribution.

$$Entropy = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} -p_{ij} * log p_{ij} \tag{5}$$

MATLAB, a widely used software for numerical computing and advanced data analysis, calculates the Gray Level Co-occurrence Matrix (GLCM) and derives a feature vector. The feature vector comprising approximately 88 elements is crucial for formulating and extracting the input data's nitrogen-deficient (N) segments. This process involves analyzing the textural characteristics of the images, with MATLAB providing the computational tools necessary to perform these sophisticated image-processing tasks. The extraction of such a detailed feature vector enables a comprehensive analysis of the N segments, facilitating accurate identification and assessment of areas in the paddy fields affected by nitrogen deficiency.

### 3.4 SVM Classifier

Support Vector Machine (SVM) is a renowned supervised learning technique for classification and regression tasks. It is particularly effective in separating target classes in n-dimensional or multidimensional spaces. The primary objective of SVM is to establish the optimal decision boundary, characterized by the largest significant margin, to categorize new data points.

In n-dimensional space, while there could be multiple potential lines or decision boundaries, the aim is to identify the simplest decision boundary that effectively categorizes the data. The dimensions of the SVM hyperplane, the decision boundary in this context, are determined by the dataset's features. The construction of hyperplanes in SVM focuses on maximizing the margins, where the margin measures the distance between the closest data points of different classes.

To achieve this, SVM finds an n-dimensional hyperplane that best separates the data points. The kernel function within SVM plays a critical role in this process. It calculates the distances between data points (x-n and x-m), with closer points receiving higher scores in the kernel's computation. Figure 6 in the referenced material likely illustrates how the SVM kernel operates, showing its capability to handle complex data distribution and effectively classify data points in the multidimensional space. This makes SVM a powerful tool for tasks involving high-dimensional feature spaces, such as image classification, text categorization, and bioinformatics.

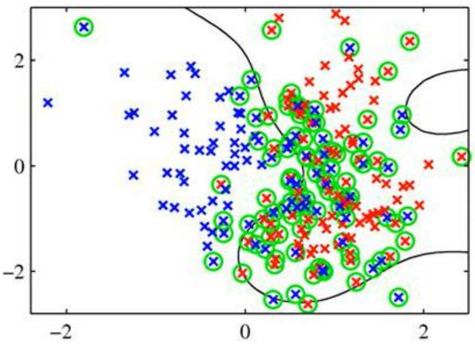



Figure 6. Non-Linear SVM Kernel

The Radial Basis Function (RBF) Kernel, utilized in this context, shares similarities with the K-Nearest Neighbors (K-NN) algorithm's functionality. The RBF Kernel is a popular choice in various kernelized machine learning methods, especially in Support Vector Machine (SVM) classification, due to its ability to handle non-linear data distributions effectively.

One of the key advantages of using the RBF Kernel, akin to the K-NN algorithm, is its simplicity in storing only the support vectors during the training phase. This approach significantly reduces the space complexity, a critical consideration in machine learning models dealing with large datasets.

The mathematical workings of the RBF Kernel are encapsulated in Equation 6. This equation likely details the computation involved in the RBF Kernel, which typically includes measuring the distance between data points in a feature space and applying a Gaussian function to these distances. The RBF Kernel's ability to map data points to a higher-dimensional space where they can be linearly separated makes it a powerful tool in SVM classification, allowing for more accurate and effective modeling of complex and non-linear relationships in the data.

$$k(\vec{x}_i, \vec{x}_j) = \exp(-\gamma \|\vec{x}_i - \vec{x}_j\|^2) \; for \; \gamma > 0 \tag{6}$$

where,

$\vec{x}_i, \vec{x}_j =$ feature vectors

$\gamma = \frac{1}{2\sigma^2}$

$\sigma =$ free parameter

In building a knowledge base for training in machine learning, a foundational dataset is established from which the system can learn and identify patterns for effective classification. The crucial stage following the creation of this knowledge base is to evaluate the model's classification performance, which is done using test data or images that are distinct from those used in training. This ensures the model's ability to generalize to new data. The testing involves several steps, starting with preprocessing the test data to ensure uniformity and suitability for analysis. Subsequent feature extraction isolates relevant characteristics or attributes from the data using techniques consistent with those applied during training. Finally, these features are tested against the knowledge base, employing the learned patterns and insights to classify or predict outcomes for the test data. This comprehensive process is vital for assessing the effectiveness of the machine learning model and identifying areas for potential improvement, ensuring its accuracy and reliability in practical applications.

### 3.5 Yield Estimation

Following verification, the segmented nitrogen-deficient (N) and Healthy areas are forwarded to the yield estimate block for calculation. This crucial step involves determining the estimated yield block based on the Validated Healthy Image (H). To do this, the value derived from the Validated Healthy Image is divided by the area of the Input Paddy Crop Image. This computation measures the yield relative to the total area under consideration.

Additionally, the scope of the general healthy area (H) is expanded by including the NH areas. NH refers to the healthy zone identified after applying the Support Vector Machine (SVM) classifier to the N region and verifying its results. Essentially, NH areas are portions of the field initially identified as nitrogen-deficient but later confirmed as healthy through the classification process.

The total yield is then represented as a percentage, as outlined in equation (7). This equation likely provides a formula to quantify the total yield about the overall area of the paddy field, considering both the healthy and recovered (NH) areas. By calculating this percentage, the model offers a comprehensive crop yield assessment, factoring in the various health statuses of different field parts, which is essential for accurate agricultural planning and management.

$$Total\,Yield = \frac{H+NH}{Total\,Paddy\,Field\,Area} * 100 \tag{7}$$

## 4 Experimental result

### 4.1 Database

The proposed system employs images captured by various standard cameras to execute operations across five databases. Each database contains unique photographs in at least one characteristic, setting them apart from images in the other databases. This diversity in the image data is crucial as it allows the system to handle various scenarios and variations encountered in actual field conditions.

The system leverages image analysis tools to enhance its performance, particularly in identifying defects or anomalies in field images. By analyzing these diverse sets of images, the system can be fine-tuned to detect various issues, such as disease infestations, nutrient deficiencies, or other factors affecting crop health. The ability to process and analyze images from different databases ensures the system is robust and versatile, adapting to different image data types and environmental conditions.

This approach of using multiple databases with distinct characteristics for each set of images allows for a more comprehensive analysis. It increases the accuracy and reliability of the system in real-world agricultural applications. Standard cameras make the system more accessible and feasible for practical use, as it doesn't require specialized or expensive imaging equipment.

### 4.2 Experimental Setup

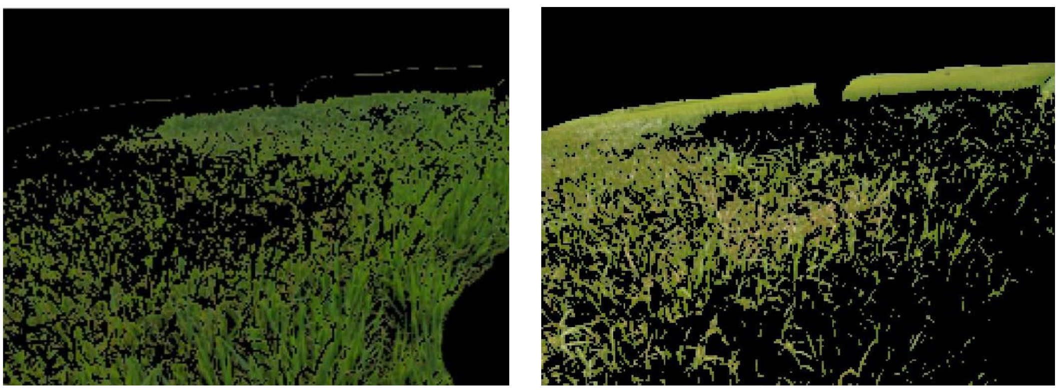
The proposed model was developed using the MATLAB Tool, facilitating distinct training and testing phases in the experimental setup. In the testing phase, real-time paddy field images are considered as input. These input images undergo a series of processes, including preprocessing, clustering, and segmentation, to prepare them for analysis. Following these preparatory steps, an SVM (Support Vector Machine) decision-based Classifier is employed to compare the input field image with the existing knowledge base. This comparison is enriched by feature extraction techniques, which help accurately assess and classify the input images based on the learned patterns and characteristics from the training phase.

Fig. 7: Input Paddy Field Image

The upcoming sections will delve into a detailed analysis of the results obtained at each stage of the proposed system. Initially, as depicted in Figure 7, an image of the paddy field is captured and subjected to pre-processing. This phase involves techniques like noise reduction and image resizing. Upon completing these pre-processing steps, the image is analyzed to identify Healthy and Unhealthy zones, as illustrated in Figures 8a and 8b. However, additional areas affected by nitrogen (N) deficiencies lead to inaccuracies in the initially identified N regions.

The next step involves passing the processed image through a hierarchical color segmentation block. This block employs threshold values to segment the image into various N-specific areas, including regions of poor health or lack of nitrogen. Subsequently, features derived from the Gray Level Co-occurrence Matrix (GLCM) are extracted and fed into the SVM classifier for validation.

The outcome of this classification, the validated segments VN, is showcased in Figure 9. This validation step is crucial for confirming the health status of different areas within the paddy field. Finally, as seen in Figure 10, the estimated yield based on this analysis is approximately 70.3799%. This percentage reflects the system's crop health and productivity assessment, considering the various health conditions across the paddy field.



(a) (b)

Fig. 8: (a) Healthy Region; (b) Unhealthy Region

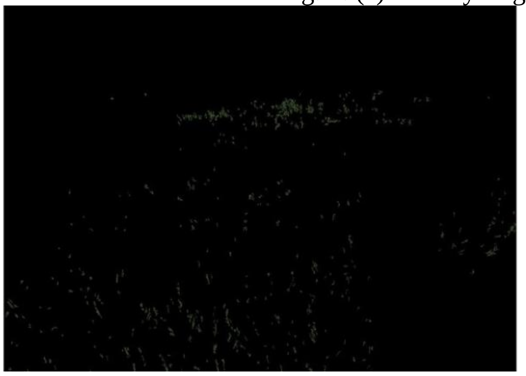


Fig. 9: VN Region


Fig. 10: Total Yield Estimated for the Input Image

Table 2 presents a comparative analysis between the current and proposed systems. This comparison highlights the advantages and improvements the proposed system offers over the existing ones, demonstrating its superiority in various aspects. In the comparative analysis of yield prediction accuracy between different systems, the data presented is as follows:

- Existing System 1 [5] has a yield prediction accuracy of 78%.
- Existing System 2 shows a lower accuracy at 47%.
- Existing System 3 varies between 78-84%, with a mid-value consideration leading to an accuracy of 81%.
- The Proposed System significantly outperforms these with a yield prediction accuracy of 92.22%.

TABLE 2: COMPARISON TABLE FOR PROPOSED AND EXISTING SYSTEMS

| Paper | Yield Prediction Accuracy |
|---|---|
| (Existing System 1) [5]. | 78% |
| (Existing System 2) | 47% |
| (Existing System 3) | 78- 84% (considering mid-value i.e. 81%) |
| Proposed system | 92.22% |

The confusion matrix is a tool that presents both actual and predicted values, allowing for detailed analysis of datasets with more than four criteria. It includes different types of predictions: the true positive, which correctly predicts a 'yes' outcome; the true negative, accurately predicting 'no'; the false positive, incorrectly predicting 'yes' when the actual outcome is 'no'; and the false negative, wrongly predicting 'no' when the outcome is actually 'yes.' Additionally, the confusion matrix provides the total counts for each row and column. Errors in the matrix are categorized as Type 1 (false positives) and Type 2 (false negatives) errors, representing inaccuracies in prediction outcomes.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

$$Specificity = \frac{TN}{FP + TN} \tag{11}$$

Figure 11 presents a confusion matrix that illustrates the performance of the constructed system, providing a comprehensive assessment through various criteria. The confusion matrix is instrumental in evaluating key performance metrics such as precision, recall (sensitivity), accuracy, and specificity. These metrics are crucial in determining how effectively the system performs its tasks.

The specific performance parameters of the proposed system are detailed in Table 3, where they are categorized under the headings of Precision, Recall or Sensitivity, and Specificity. To calculate these parameters, mathematical formulas are employed, starting from Equation (8) and extending through Equation (11). Each equation plays a role in quantifying different aspects of the system's performance.

Additionally, the performance of the proposed system is visually represented in a graph, as shown in Figure 12. This graphical representation provides an intuitive understanding of how well the system performs, highlighting its strengths and areas for improvement in a clear and accessible manner.

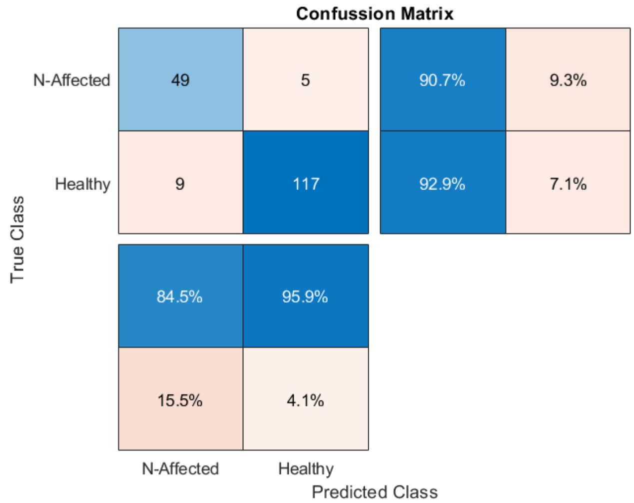



Fig. 11 Confusion Matrix of Proposed Method

TABLE 3: PERFORMANCE ANALYSIS TABLE FOR THE DIFFERENT DATA SET

| Database | Number of Images worked | Precision | Recall or Sensitivity | Specificity |
|---|---|---|---|---|
| Total Number of Images = 180 | | 95.90 | 90.74 | 92.86 |

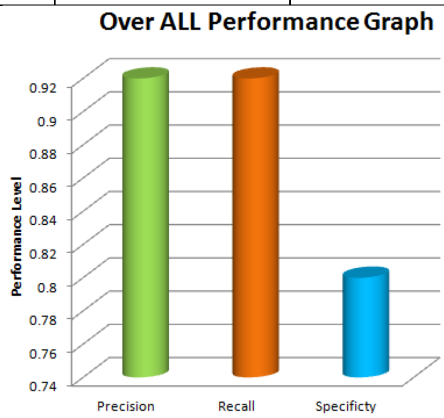



Fig. 12 All System Performance Graph

## 5 Conclusion

The results of this study introduce a method for segmenting the affected areas in paddy field images, with a specific focus on addressing nutrient deficiencies like nitrogen (N). This method incorporates sophisticated techniques such as selecting suspected regions based on hierarchical color segmentation and using an SVM classifier. These techniques contribute significantly to the accuracy of estimating the total yield in the input field image by taking into account the effects of nutrient deficiencies.

The methodology proposed in this study showcases a more precise and systematic approach to calculating the total yield, particularly by considering the severity of illnesses affecting the crop. This approach has been applied to various field images, yielding impressive results with an accuracy rate of 92.22%. The study suggests integrating an exact prediction mechanism within the Paddy Field Image system, enhancing the ability to compute the total yield with even greater precision. This advancement promises to significantly contribute to agricultural imaging and analysis, providing a more accurate and reliable tool for yield estimation in paddy fields.

**Reference**

[1] 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.
[2] 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.
[3] Manickam Gopperundevi 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.
[4] Mrs. K.R. Sri Preethaa M.E, “Crop Yield Prediction,” International Journal on Engineering Technology and Sciences, Vol. III, Issue III, 2016.
[5] 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, 2006.77
[6] Manickam Gopperundevi and Vijayaraghavan Kannan, “PaddyYield Estimation Using Remote Sensing and Geographical Information System,” Journal Of Modern Biotechnology, Vol. 1, Issue 1, pp. 26 – 30, 2012.
[7] 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.
[8] 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.
[9] 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.
[10] Aswini Kumar Mohanty, Swapnasikta Beberta 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] J. S. J. Wijesinghaa, N. L. Deshapriyab, and L. Samarakoonb, “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.