

Machine Learning Applications for Precise Nutrient Deficiency Detection in Paddy Farming

¹Hitesh Kalita, ¹Abhijit Bora, ^{1,2}Uzzal Sharma

¹Department of Computer Applications, Assam Don Bosco University, Guwahati, India

²Department of Computer Science, Birangana Sati Sadhani Rajyik Vishwavidyalaya, Golaghat, India

Abstract:

This study tackles the critical issue of nutrient deficiencies in paddy crops by leveraging machine learning for accurate detection, while also considering the economic implications for farmers. Since agriculture is fundamental to global food security—and rice is a key staple crop—sustainable cultivation depends on effective nutrient management. Deficiencies in paddy crops can significantly reduce yield and quality, posing a major challenge for farmers. Traditional detection methods, which rely on subjective visual inspection, are labor-intensive and often delay corrective measures. However, advances in machine learning and data-driven approaches present a promising solution. The research focuses on using machine learning to identify nutrient deficiencies in paddy leaves, employing the K-means clustering algorithm for precise and efficient detection. It explores practical applications in agriculture, emphasizing robust feature selection and model validation. The proposed method integrates image processing techniques, including image capture, RGB-to-HSI conversion, and segmentation, building upon prior work in nutrient deficiency detection. Performance is evaluated using key metrics such as accuracy, precision, recall, and F1 score. The model is trained and tested on datasets from the International Rice Research Institute (IRRI), targeting deficiencies in nitrogen (N), potassium (K), phosphorus (P), manganese (Mn), and zinc. By combining machine learning with image analysis, this approach offers a faster, more reliable alternative to traditional methods, ultimately supporting better crop management and food security.

Keyword: Paddy, Nutrient Deficiency, IRRI, Machine Learning, K-means, SVM

Introduction

Agriculture serves as a cornerstone of global food security, with rice (paddy) standing as a vital staple crop that sustains billions worldwide. Sustainable rice cultivation requires diligent monitoring of plant health and proper nutrient management, as deficiencies can significantly compromise both crop yield and quality - a pressing challenge for farming communities worldwide. Timely and accurate detection of nutritional deficiencies in paddy leaves is crucial for enhancing farming practices and maintaining stable food supplies. Conventional detection methods, which primarily depend on manual visual inspection, prove not only laborious but also subjective in nature. Such limitations frequently result in delayed corrective actions and suboptimal crop outcomes.

However, new technical advancements, especially the combination of data-driven methods and machine learning, present a viable answer to this long-standing issue. With a focus on identifying nutrient shortages in paddy leaves, this research study aims to explore the application of machine learning techniques in agriculture.

It is difficult to diagnose these deficiencies accurately since it involves interpreting and analyzing data on the health of the plants and their nutrient status. With its ability to process large datasets, identify patterns, and generate predictions based on data, machine learning has become an effective tool for tackling this problem. Our goal is to develop a reliable and effective approach for identifying nutrient shortages in paddy leaves by utilizing machine learning techniques.

Even in India, paddy yield declines as predicted since there is no reliable way to identify nutritional deficiencies in the crop. For farmers, it becomes a major worry. The challenge facing economists and policymakers in India is to come up with a strategy to increase yield levels from the current state of stagnation and broaden the declining profit margins. Paddy farming has stagnated even in places like Punjab, which benefited greatly from the green revolution, as a result of heavy resource extraction, a sluggish adoption of production technologies, and poor policy formation [11]. The cost of cultivation has increased due to the new economic policies that advocated cutting subsidies on essential farm inputs like fertilizer, which could lead to wash-off profits from paddy production. The weak margins enjoyed by farmers are a result of a combination of factors such as high input needs, rising input costs, and a slow increase in guaranteed prices. In a technology system like this, farmers' only remaining control is how well they use input to maximize their potential output.

Crop production efficiency studies enable us to comprehend the potential yield levels of the current production system and thereby enhance the actual yield; for example, they show us how to improve productivities without increasing input application [12,13,14] or how to more effectively utilize existing technology and institutional reforms to allow for innovations and investments in rural infrastructure to boost production growth [14]. Policymakers can use the results of a nationwide research comparing the production efficiency of different pre-classified farm categories to determine how best to deploy resources in order to maximize productivity. To fully realize the promise of modern technologies and implement productivity changes, these studies are crucial [12]. Kalirajan et al. [16] state that limited resources and a slowdown in technical advancement are the two main issues facing developing nations. In this case, by accurately identifying nutritional deficiencies in paddy crops, this method will be extremely important in increasing rice yield and farmers' revenue.

This study explores machine-learning approaches and strategies for identifying nutrient deficiencies in rice leaves. It looks at how machine learning algorithms should be applied in the real world, how to choose the right features, and how to validate models to make sure they work as intended in agricultural environments.

Previous Work:

Finding nutritional deficiencies in rice fields is not a recent endeavor. Numerous researchers are pursuing various avenues in their efforts to identify nutritional deficiencies in rice. In particular, machine learning based on image processing has made a substantial contribution and been successful in identifying the nutritional deficit in paddy. In terms of illness identification, A.K. Ghorai et al. have achieved a noteworthy outcome after employing a certain methodology [1]. Combining automated farming with nutrition deficiency monitoring is another topic that has already seen a substantial amount of research. A noteworthy contribution was made by T. Rajasekar and associates [2]. The researchers now have an exceptional perspective thanks to the application of AI and image processing combined with ML. Any plant's health can be tracked with image processing methods. Vignesh Dhandapani [5] has

worked in a similar field and produced excellent results. The leaves of any tree will show you exactly where the plant is lacking in nutrients. In actuality, the symptoms on the leaves make the nutrient shortage quite evident. Amirtha T[4], et al.'s research has demonstrated the value of image processing techniques in the rapid and precise identification of nutrient deficiencies. In an alternative method, the researcher has attempted to combine IoT, ML, and AI for quick and precise nutrient deficit identification. Gaganjot Kaur [3] has conducted a comparative analysis of the methods employed by other researchers in other regions of the world.

The field of applied economics can greatly benefit from the addition of machine learning (ML) to its toolkit. Big data and machine learning have the potential to enhance farm management [17] and economic analysis in general [18], according to recent overview studies. Economic analysis is starting to use machine learning techniques [19]. A significant aspect of agricultural economics is smart farming. In order to address the issues of food security, productivity, environmental impact, and sustainability in agriculture, smart farming is crucial [20]. Given that the world's population has been growing steadily [22], there needs to be a significant increase in food production in order to preserve high nutritional quality and availability while also safeguarding natural ecosystems through the use of sustainable farming practices. Images make up a sizable portion of the vast amount of data gathered by remote sensing. Images can address a range of issues and often provide a comprehensive image of agricultural landscapes [21]. Therefore, imaging analysis is a crucial field of study in the agricultural sector, and a variety of agricultural applications use intelligent data analysis techniques for anomaly detection, picture identification/classification, and other purposes [23]. However, prior research has not concentrated on the financial side of farming by identifying nutritional deficiencies in paddy. In this sense, the use of machine learning (ML) to agricultural growth can greatly benefit farmers by improving productivity, production, and revenue through precise identification of nutrient deficiencies in rice fields.

The Recommended Approach:

The K-means clustering technique is employed in this study to identify nutrient deficiencies in paddy fields.

The proposed model is as shown in the Figure1.

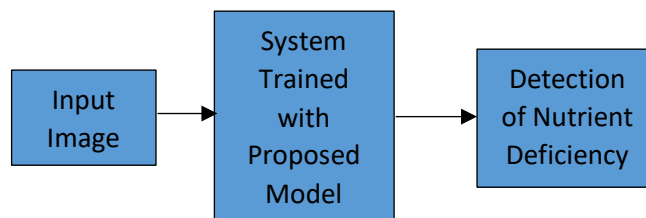


Figure1: Proposed model for detection of nutrient deficiency

The leaf photograph captured using a cell phone will be accepted by the suggested model. All necessary features will be extracted following the conclusion of the picture capture procedure and the pre-processing step. The pre-trained model will take into account every aspect of the obtained image. K-means clustering will next be applied in order to extract the most relevant features. We are taking into consideration 11 most important parameters for this investigation. The model will display the outcome

if the plant under consideration has any nutritional deficiencies based on these criteria. The many phases of the complete procedure are depicted in Figure 2.

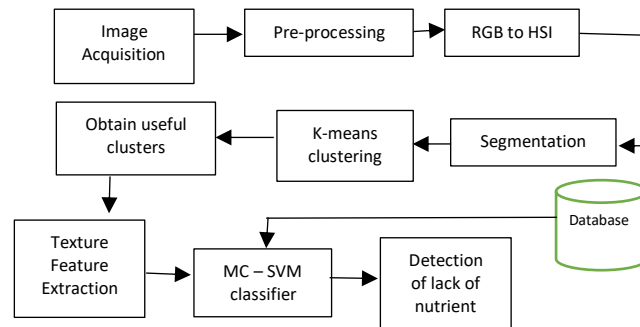


Figure2: Different stages of the process

Let's now shed some light on the various phases of the procedure:

3.1 Image Capture: Several photographs of the sample must be taken into account when building the model. The study's images were taken using a DSLR or a mobile device. The suggested approach is conceived of as starting with image acquisition. The process of getting an unprocessed image from a source is how it is defined. The test images were obtained from the Philippine IRRI database. The input image's RGB (red, green, and blue) format is employed. After that, the RGB leaf image is converted to the suitable colour space for our purposes.

3.2 Conversion of HIS to RGB: Using the Hue Saturation Intensity (HSI) colour model, the RGB pictures of the leaves are displayed. The HSI colour space representation is well-known due to its connection to human vision. The Hue component of the image's colour corresponds to the primary colour[8]. Saturation gives any image the right amount of purity. It is the amount that has been added to the colour value of white light. The colour space in this proposed work is changed from RGB to HSI representation since intensity is a measure of the amount of light present. The H component is taken into consideration for additional research after the transformation [9]. It is the amount that has been added to the colour value of white light. The colour space in this proposed work is changed from RGB to HSI representation since intensity is a measure of the amount of light present. After conversion, the H component is taken into consideration for additional research.

3.3 Image segmentation: The most crucial step in any investigation of image processing is picture segmentation. The following describes the many phases of the segmentation process:

3.3.1 Lookup for Green pixels: First, the pixels with a high percentage of green are found. Next, a threshold value is determined so that pixels are masked if the green portion of the pixel intensity is less than the threshold value. Red, green, and blue components of pixels are all set to zero. The green pixel has been hidden, negating the need to extract it for additional examination.

3.3.2 Useful Segment Extract: Next, the affected leaf section is divided into equal-sized patches, with the patch size selected so that critical information is retained. None of the portions provide any information that would be helpful. Patches containing more than 80% meaningful information are therefore taken into consideration for additional examination.

The morphology of leaf tips, the intervening and interspersed regions, sports, and other critical segment components are all segmented using the statistical region merging algorithm.

The primary characteristics that are extracted are the segment's colour and texture.

Colour separation: The colours of the photos are represented using the HSI (Hue, Saturation, and Intensity) colour model. The colour histogram shows how colour is distributed throughout an image. This stage involves computing the colour intensity of the leaves and displaying it as a histogram. One of the key parameters of the dataset under design is the colour intensity, which is also taken into account and kept [10]. The colour depicts the various vitamin deficiencies as seen in Figure 3.



Figure3: Nutrient deficiency symptoms on paddy leaves (a) N, (b) P, (c) K, (d) Mn, (e) Zn

Texture: To extract the texture of leaves, statistical region merging is utilized. These could have features like curling, reddish-purple tips, tip and margin deaths, and so forth. The testing database retains these attributes.

3.4 Creating a Testing Database: The attributes of the leaves extracted in the previous phase are stored in the testing database. The proposed method finds traits for each nutrient found in the literature review, such as different leaf types, appropriate colour placement, and texture changes. So that the experimental leaves may be compared to them, they are added to the testing database.

K-Means clustering, one of the most basic unsupervised machine learning algorithms, takes a very simple technique to grouping a given data set into several clusters. Only one cluster, denoted by its centroid or centre point, may contain all of the data points in the set; these clusters should be spaced significantly apart from one another. Subsequently, every point is taken out of the provided data set and allocated to the nearest centre [11].

Steps in the algorithm for K-Means clustering:

- a) Identify 'c' number of clusters at random.
- b) Determine the separation between the cluster centres and data points.
- c) Set every item in the group with the nearest centroid.
- d) After assigning a cluster to each point in the data set, the new cluster must be recalculated using

$$V_j = \left(\frac{1}{c_i}\right) \sum_{j=1}^{c_i} x_i \quad (1)$$

- a) Next, determine the separation between every data point and the obtained cluster centers.
- b) If no data point was reallocated, then end the process; if not, repeat steps (c) and (d) until the centroid stops moving.

The different features were extracted using K-means clustering. The colour feature and mean value of each plane (the R, G, and B planes) are calculated. Among the shape features that are computed are area, perimeter, breadth, height eccentricity, convex area, area/perimeter, and area / convex area [7]. The geometric moment value of the R plane is also calculated.

Analysis of the suggested model's performance:

We evaluated the performance of our proposed model using both quantitative and graphical methods. The qualitative evaluation is a popular method for comparing and evaluating visually the outcomes of picture classification [6]. To evaluate our suggested model, we used a confusion matrix and many metrics, which are described below: accuracy, precision, recall, and F1 score.

- i) Accuracy: An important factor in evaluating the model is accuracy. Out of all the data, it is the number of correctly predicted data out of the whole data and given by:

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} * 100 \quad (2)$$

- ii) Precision: Precision can be defined as the number of relevant instants that the model retrieves, as measured by

$$Precision = \frac{\sum TP}{\sum TP + \sum FP} * 100 \quad (3)$$

- iii) Recall: Recall is a metric that counts the number of correctly predicted positive outcomes out of all possible outcomes and is determined by

$$Recall = \frac{\sum TP}{\sum TP + \sum FN} \quad (4)$$

Where TN stands for true negative, FP for false positive, FN for false negative, and TP for true positive.

- iv) F1 score: It is the precision and recall harmonic mean. It is provided by and utilized to compare various models as follows:

$$F1\ score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) * 100 \quad (5)$$

6. Result and Discussion

The IRRRI provided the sample photos. The dataset has a size of 230 MB. The dataset is split into 75% and 20% for training, testing, and validation purposes, respectively, in order to conduct the training and testing processes. Following the K-Means feature extraction process, 11 different features were taken into consideration. Mean R, Mean G, Mean B, Area, Perimeter, Eccentricity, Convex area, Width, Height, Area / perimeter, and Area / convex are the features taken into consideration. When classifying rice plants that are nutrient-deficient, SVM classifier is employed.

6.1 Performance evaluation:

The mineral deficient leaves were classified into 5 classes, namely N, K, Mn, P and Zinc deficient leaves.

Table1: Accuracy in percentage

Slno	Mineral deficiency	Accuracy in Percentage
1	N	82
2	N	80
3	K	85
4	K	89
5	K	91
6	Mn	90
7	Mn	83
8	P	86
9	P	86
10	Zn	81
Average accuracy = 85.3		

Using a variety of characteristics that were retrieved using K-Means, SVM was able to accurately identify the samples of rice leaf images that were lacking, with an accuracy of just 85.3%. With a higher degree of effectiveness, the SVM model was able to classify the plants with varying nutritional deficiencies based on the futures that were derived from the images of rice plants using K-means. Additionally, assessed is the designed model's computational time complexity. Additionally, it is discovered that the suggested system is noticeably less.

Conclusion:

In summary, this research paper has explored the critical field of paddy crop nutrient deficiencies, highlighting the importance of precise detection for agricultural practice optimization and global food security. Conventional techniques that depend on visual observations have shown to be labour-intensive and subjective, which has caused crop performance to be compromised and interventions to be delayed. This long-standing issue may be resolved with the help of machine learning techniques, especially the suggested K-means clustering algorithm. In order to detect nutrient deficiencies in paddy, the research builds on earlier work in image processing and machine learning. It presents a thorough method that includes image acquisition, RGB to HSI conversion, image segmentation, and the extraction of texture and colour features. The suggested model finds nutrient deficiencies in paddy leaves by using K-means clustering to extract important parameters.

The model's performance analysis shows that it is effective in classifying mineral-deficient leaves into groups such as N, K, P, Mn, and Zinc deficiencies based on a range of metrics, including accuracy, precision, recall, and F1 score. The model's ability to correctly identify nutrient deficiencies based on the extracted features is attested by its average accuracy of 85.3%.

Moreover, the effectiveness of the suggested system is demonstrated by the computational time complexity assessment. It provides a strong and dependable approach to nutrient deficiency detection and exhibits computational efficiency, an essential component for practical implementation.

To sum up, this study makes a substantial contribution to the current initiatives aimed at improving agricultural practices by utilizing machine learning methods. In order to promote

sustainable rice production and enhance global food security, the suggested model tackles the critical problem of nutrient deficiency in paddy crops. This makes it an invaluable tool for farmers and agricultural community.

References:

- [1] A A.K. Ghorai, S. Mukhopadhyay, S. Kundu¹, S. N.Mandal, A. Roy Barman, M. De Roy, S. Jash and S. Dutt, “Image Processing Based Detection of Diseases and Nutrient Deficiencies in Plants” ISSN 0971-975X, January 2021.
- [2] T. Rajasekar, M. Arun Kumar, K. Mohamed Ismail, M. Sabarimuthu, “Automated Farming and Nutrition Deficiency Detection using Swarm Bots” (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020
- [3] GaganjotKaur,”Automated Nutrient Deficiency detection in plants:A review “ Palarch’s journal of Archaeology of Egypt/egyptology17(6).ISSN 1567-214x.
- [4] Amirtha T, Gokulalakshmi T, Umamaheswari P, T Rajasekar, “Machine Learning Based Nutrient Deficiency Detection in Crops” (IJRTE) ISSN:2277-3878, volume-8 ISSUE-6, March2020
- [5] Vignesh Dhandapni,S.Remya,T.Shanthi,R.Vidhy, “Plant Health monitoring using image Processing” (IJERCSE) vol 5,Issue 3,March 2018ities.
- [6] M.F. Taha, et al., Using deep convolutional neural network for image-baseddiagnosis of nutrient deficiencies in plants grown in aquaponics, Chemo sensors(10) (2022) 45.
- [7] P. Sanyal¹, U. Bhattacharya², S. K. Parui, S. K. Bandyopadhyayand S. Patel. Color Texture Analysis of Rice Leaves to Diagnose Deficiency in the Balance of Mineral Levels Towards Improvement of Crop Productivity.10th International Conference on Information Technology.
- [8] PrabiraSethy, BaishaleeNegi, NilamaniBhoi(2017). Detection of Healthy & Defected Diseased Leaf of Rice Crop Using K-Means Clustering Technique. International Journal of Computer Applications. January 2017, Volume 157- No.1, ISSN: 0975-8887.
- [9] Li ZL, Wu H, Duan SB, Zhao W, Ren H, Liu X, Leng P, Tang R, Ye X, Zhu J, Sun Y. Satellite remote sensing of global land surface temperature: definition, methods, products, and applications. Reviews of Geophysics. 2023, Mar;
- [10] Thorat T, Patle BK, Kashyap SK. Intelligent insecticide and fertilizer recommendation system based on TPF-CNN for smart farming. Smart Agricultural Technology. 2023 Feb
- [11] Singh, S. A study on technical efficiency of wheat cultivation in Haryana. *Agric. Econ. Res. Rev.* 2007, 20, 127–136. [[Google Scholar](#)]

- [12] Ali, M.; Chaudhry, M.A. Inter-Regional Farm Efficiency in Pakistan's Punjab: A Frontier Production Function Study. *J. Agric. Econ.* 1990, *41*, 62–74. [[Google Scholar](#)] [[CrossRef](#)]
- [13] Umesh, K.B.; Bisaliah, S. Efficiency of groundnut production in Karnataka: Frontier profit function approach. *Indian J. Agric. Econ.* 1991, *46*, 20–33. [[Google Scholar](#)]
- [14] Gaddi, G.M.; Mundinasmani, S.M.; Hiremath, G.K. Resource use efficiency in groundnut production in Karnataka—An economic analysis. *Agric. Situat. India* 2002, *58*, 517–522. [[Google Scholar](#)]
- [15] Kalirajan, K.; Shand, R.T. A generalized measure of technical efficiency. *Appl. Econ.* 1989, *21*, 25–34. [[Google Scholar](#)] [[CrossRef](#)]
- [16] Kalirajan, K.; Obwona, M.; Zhao, S. A Decomposition of Total Factor Productivity Growth: The Case of Chinese Agricultural Growth before and after Reforms. *Am. J. Agric. Econ.* 1996, *78*, 331–338. [[Google Scholar](#)] [[CrossRef](#)]
- [17] Raj, M. P., Swaminaraya, P. R., Saini, J. R. and Parmar, D. K. (2015). Applications of pattern recognition algorithms in agriculture: a review. *International Journal of Advanced Networking and Applications* 6: 2495.
- [18] Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America* 113: 7353–7360.
- [19] Athey, S. (2019). The impact of machine learning on economics. In: Agrawal, A., Gans, J., and Goldfarb, A. (eds.), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press (p. 507–547).
- [20] Gebbers, Robin, and Viacheslav I. Adamchuk. "Precision agriculture and food security." *Science* 327.5967 (2010): 828-831.
- [21] Liaghat, Shohreh, and Siva Kumar Balasundram. "A review: The role of remote sensing in precision agriculture." *American journal of agricultural and biological sciences* 5.1 (2010): 50-55.
- [22] Kitzes, Justin, et al. "Interpretation and application of the Ecological Footprint: A reply to Fiala (2008)." *Ecological Economics* 68.4 (2009): 929-930.
- [23] Singh, Arti, et al. "Machine learning for high-throughput stress phenotyping in plants." *Trends in plant science* 21.2 (2016): 110-124.