Received: December 29, 2022. Revised: February 2, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An examination of Groundnut Crop Leaves to Identify the Chlorophyll Deficiency Before Visible Symptoms

Janani M[1]* **Jebakumar R[1]**

[1]*SRM Institute of Science and Technology, India*
* Corresponding author's Email: jr7791@srmist.edu.in

Abstract: Nitrogen nutrient concentration is predominant in crop field monitoring. Noticing the deficiency pattern, providing sufficient amounts of nutrients will help to recover the crop. Sometimes it may not be possible to reverse the crop in cases of deep deficiency. It will affect the quality and quantity of the crop product in yield time. Hence, timely identification is important. Therefore, detection of chlorophyll deficiency in groundnut leaves (DCDGL) system is proposed here for nutrient deficit recognition, before a notable pattern emerges. In first phase, groundnut leaf images are collected and clustered based on chlorophyll measurement. Then, different image processing techniques like image compression, resizing, and noise reduction filters are applied and analyzed on leaf images to enhance the images and secure the predominant features. The median filter is fixed because it provides high quality output with a low noise percentage. The quality of an image is measured in decibels with the help of the peak signal to noise ratio (PSNR). Then, in the next phase, colour features and grey level co-occurrence matrix texture features are extracted and given to developed intensified multinomial classification (IMC) model to detect nutrient deficiencies in groundnut leaves. The obtained average accuracy of IMC model is 97%. The DCDGL system's performance is tested with state-of-the-art techniques like K-nearest neighbour and support vector machine. The developed model performance is superior to state-of-the-art techniques regarding accuracy. The appropriate image quality analysis and pre-processing approach helped achieve less processing time to predict nutrient deficiency before tangible symptoms.

Keywords: Groundnut leaves**,** Multinomial classification, Chlorophyll deficiency, Machine learning.

1. Introduction

Bioreceptor-based methods are widely available for directly detecting plant health status [1]. However, those methods are time-consuming, require certain training, and human intervention. Considering these requirements as an alternative, a simple visual observation by an expert is used to detect diseases [2]. But based on the experience of the expert, there is a chance of unfairness in the result. Machine vision is an attractive and emerging technique in today's world. It has been widely used in several applications. In recent decades, machine vision has been used in agriculture applications to support precision farming [1]. This enables the agricultural system to provide more efficient and accurate solutions than performing agricultural activities manually, which holds considerable advantages when compared with the bioreceptor method and expert method [2]. It reduces human involvement, continuous monitoring can be made available, it automatically identifies defects in crop fields, it protects farmers from hazardous environments, etc. With the support of a machine vision system several agriculture applications developed. Among them, monitoring and analyzing growth conditions and the health status of plants through their leaves are important aspects [3, 4].

Machine vision systems includes digital image processing and machine learning techniques. The most commonly used image processing techniques in agriculture are image enhancement, feature extraction, and feature selection. Classification comes under the machine learning technique. A combination of image processing and machine learning techniques can be used to determine the health of plant leaves [5].

To create an efficient machine vision-based

system to detect crop field problems, proper data acquisition is important. There are many different types of leaf image acquisition resources available for researchers' choice. During data collection, light sources must be emphasized. In indoor cases, the lightning conditions are natural light, fluorescent, incandescent, etc., and in outdoor cases, the lightning conditions include sunny, cloudy, overcast, evening, and so on [6]. The acquired image features' noise level, and quality, vary with different lighting conditions, it will have an impact on the plant's health status result. Therefore, researchers use noise reduction techniques, also known as image quality enhancement techniques, to eliminate different types of noise that occurred during the leaf image acquisition. Some researchers acquired leaf image data under controlled illumination and lightning conditions [7-9]. But it requires high-cost resources for development, and sometimes the implemented setups are not scalable.

Machine learning techniques are vital in providing real-time field insights. The majority of machine learning classification techniques learn using numerical features. So in order to detect defects automatically from leaf images using machine learning techniques, they need to be converted into numerical features. Image feature extraction techniques help to extract the numerical features from the leaf images. If the researcher feels that the specific features are important, then the feature selection technique can be used. In cases of dependency between leaf features, feature reduction techniques can be applied.

Unlike the bio-receptor-based direct detection method, the image processing methods are not specific to particular stress; each time they need to be analyzed with different types of techniques to identify types of stress or disease. However, these limitations, can still be overcome with the amalgamation of image processing and machine learning classification techniques [10].

The health of the leaves is an important factor in determining plant growth conditions, including deficiencies and diseases. The pigment composition of leaves changes due to nutrient deficiencies and environmental stress. The stress and deficiency lead to a decrease in chlorophyll content. Lower chlorophyll content exhibits reduced greenness, which is a reliable indicator of senescent and unhealthy conditions. Based on visible symptoms, it is easy to know the leaf's status. However, the recovery process is complex, and anyway, it is going to affect the quality of the yield. Predicting nutrient deficiency in advance can facilitate stopping the deepening of deficiency or stress and preventing irreversible conditions in plants. Therefore, detection of chlorophyll deficiency in groundnut leaves (DCDGL) system is proposed to identify deficiency before tangible symptoms. This system developed by integrating image processing, feature extraction and machine learning techniques. The proposed work aims to (i) obtain quality and quickly processable images through pre-processing methods without the use of high-cost resources and controlled illumination and lighting setups. These pre-processed images help to acquire prominent features during the feature extraction process and reduce processing and execution time in the classification process, (ii) obtain enhanced nutrient prediction classification accuracy than the start of art techniques without the use of optimization techniques.

Section 2 discusses various meters and different types of image processing and machine learning techniques developed or utilized by researchers to identify leaves' nutrient status.

2. Literature survey

Leaf reflectance information can be used as an indicator to predict physiological stress, nutrient deficiency, and plant health [11]. To measure the leaf spectral information rapidly, various methods and metres are used. Those are hyperspectral imaging and thermal imaging [12-14], spectroscopy [15], a fluorescence-based chlorophyll meter, and a transmittance chlorophyll metre [16, 17]. Specifically, to measure the leaf chlorophyll concentration, two approaches are used: destructive [18] and non-destructive [19]. In destructive methods, leaves are collected from the field, and the foliar chlorophyll concentration of leaves is measured by organic extraction. In a non-destructive way, the leaf spectral reflectance information is observed to analyze some features such as chlorophyll content, leaf water content, canopy temperature, and other leaf information. In general, the chlorophyll concentration of a leaf is closely associated with its nitrogen concentration. This relationship clearly shows that leaf chlorophyll concentration can be used to assess plant health. However, chlorophyll metres are expensive, and as a result, marginal farmers cannot afford them. So the computer vision based digital image analysis method with a machine learning algorithm could be a site-specific nutrient management (SSNM) system for groundnut leaves.

The spinach plant health status and chlorophyll concentration were estimated by Agarwal et al., using leaf colour features. Data collection were done on the 35th day after plantation. Thirty-four different colour indices were extracted and analyzed to find out the

relationship with chlorophyll content. Their result showed that among those colour indices, eight colour indices exhibited higher and considerable coefficient of determination with chlorophyll content. Principal component analysis and agglomerative hierarchical clustering analysis (AHCA) were used to segregate the range of chlorophyll content from the colour space information [20]. Analysis of spinach leaf images in different growth stage may provide intensified result. The experiment was done on controlled environment, since, the result may vary with outdoor spinach leaves.

Park et al., proposed minimum redundancy and maximum relevance (mRMR) technique to select essential raw bands directly from hyper-spectral images of apple leaves. To increase the amount of training data and avoid overfitting problems in the classification process multiple patches were taken from each hyperspectral image. Six leaf conditions were classified using a deep neural network, which aids in the identification of disease, nutritional status, and plant growth stage [21]. The performance of the selected waveband depends on several factors, including the number of samples taken at a location and the type of spectral data used, which determine the success or failure of the method. Some other drawbacks are high computational complexity, requiring huge computational resources, and being less popular and more expensive than RGB cameras. Acquiring a raw image dataset from the field cannot be used to achieve better accuracy. Therefore, different image pre-processing techniques can be applied to obtain high-quality images.

Chlorophyll fluorescence measured in the leaf images of radish, tomato, maize for predicting sulphur and nitrogen nutrient deficiency before the defective or noticeable symptoms. The integrated photosynthetic machinery response measurements were collected from those leaf images in large amounts using the OJIP test. Then the collected test parameter data with chlorophyll measurement applied in PCA to analyze the relationship between them to predict sulphur and nitrogen deficiency. They observed that integrative test parameters help to identify sulphur deficiency in those crops [22, 23].

HirotoYamashita et al., estimated chlorophyll content in tea tree leaves under a hydroponic environment. They collected hyperspectral leaf reflectance information on tea leaves under different lighting conditions. The reflectance data was analyzed to find the correlation with the help of a kernel-based extreme learning machine (KELM) regression algorithm to find out the relationship between spectral data and chlorophyll content and the obtained significant coefficient of determination. [24]. However, the findings of the work will differ in the outdoor field because of outdoor environmental factors. Haixia Qi et al., detected the chlorophyll content of peanut leaves by using the backpropagation neural network (BPNN) model. Input features from multispectral images were extracted and pre-processed with the de-trending (DA) technique. The root mean square error metric evaluated the model performance and compared it with random forest algorithm (RF). The obtained RMSE of BPNN is less than RF [25]. Generally, a neural network model requires many samples to provide the best performance. Otherwise, the model might suffer from overfitting or under fitting issue. The obtained result may be unreliable with this kind of issue.

Minu Eliz Pothen and Maya L Pai detected rice leaf disease using the otsu method for segmentation, histogram of oriented gradients (HOG) for feature extraction, and SVM for classification. The data were collected from the UC Irvine machine learning repository to perform this experiment. They analyzed and compared the HOG method with local binary patterns (LBP) feature extraction method and found that with HOG features, SVM can provide the best result than LBP [26].

Haiqing Wang, Shuqi Shang et al., developed a lightweight YOLOv5 model to detect peanut plant disease using the plant village website and self-made dataset. The collected image dataset was augmented using the augmentation technique to introduce new features. Because neural network requires a large number of samples to provide the best result. Their model performance was improved by introducing submodules and replacing existing structures of YOLOv5 model. The developed light weight model yielded best performance than YOLOv5 regarding accuracy and execution time [27]. However, it requires graphical Processing unit access to handle dataset and model structure.

From the survey, it is observed that previous research has not combined chlorophyll concentration measurements with HSV colour space features and leaf texture features to assess plant health status using a digitized groundnut leaf image. There is no existing model to analyze the various levels of chlorophyll concentration in different growth phases with continuous monitoring for a particular crop, which is important to know the day wise status of the crop. As a result, it is critical to detect leaf stress before visible or tangible symptoms appear in various growth stages at a low cost. The objective of the proposed work is to determine the chlorophyll concentration of groundnut leaves in order to determine nitrogen

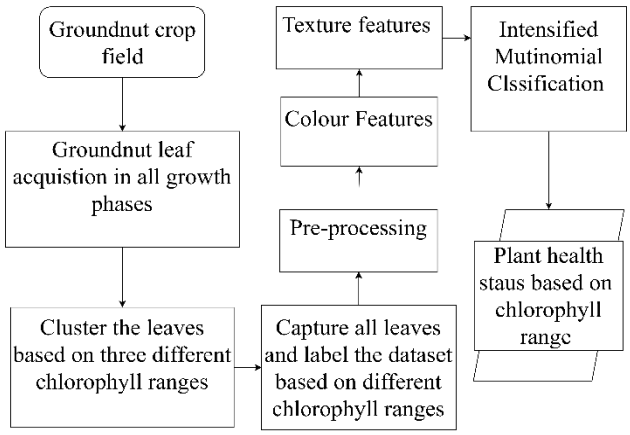


Figure. 1 Flow diagram of DCDGL system

nutrient status at various levels, such as deficient initial stage (DIS), deficient (D), and conventional(C) using RGB images.

3. Detection of chlorophyll deficiency in groundnut leaves (DCDGL) system

Regarding the requirement of identifying nitrogen nutrient deficiencies in groundnut leaves using a machine vision-based approach, in first stage groundnut leaves without notable symptoms were randomly selected and acquired from the field in two phases. The number of leaves acquired is 304 and 316 in the pre-flowering phase and flowering phase, respectively. The leaves were picked during the day between 9 a.m. and 3 p.m. from June to August 2020. The chlorophyll content of the leaves was measured instantly after picking the leaves with the use of atLEAF CHL BLUE chlorophyll metre device (FT Green LLC, US).

The mean value of each leaf was measured while omitting the midrib and margins. Afterwards, the adaxial surfaces were captured using a mobile phone camera with a white background. Next, the images are put through different image processing techniques, such as compression, resizing, and filtering, which help reduce the size of the dataset, make processing easier, and remove noise to enhance the image quality. In Fig. 1 shown the flow diagram of DCDGL system.

Then in the next stage, colour space information-hue, saturation, and value (HSV) is acquired as a colour feature because of its superior performance over red, green, and blue colour space information, and it provides varying illumination ranges.

The GLCM technique is used for extracting different types of texture features. Eventually, the extracted features are given as input to the intensified multinomial classification technique to detect and classify the chlorophyll concentration range of groundnut leaf images.

4. Methods

4.1 Pre-processing and feature extraction

Image compression is a type of data compression applied to digital images. The intention of image compression is to obtain feature space from the high dimensional input space where the maximal variance is displayed. However, the size of the image factor alone cannot be considered for obtaining the objective. Another important factor that needs to be considered is image noise. In most cases, during data collection, image corruption will occur due to several reasons such as atmospheric conditions, fluctuation in the signal, a defective camera's sensor, electronic data transfer, etc. One more important reason is that while doing image compression, the noise level of the image will also increase. This kind of corruption is represented by different types of noise. It should be eliminated in order to improve the image quality even more. Through different noise reduction filters, these noises can be removed. Here, the noise reduction filters used to reduce the noise in the image are the mean, median, and Gaussian. [28] The mathematical representation of noise elimination filters such as mean, Gaussian and median are represented in Eqs. (1), (2), and (3) respectively.

$$f'(x,y) = \frac{1}{mn}\sum_{(s,t)\in Sxy} g(s,t) \quad (1)$$

S_{xy} represents set of coordinates in a sub image with window size $m \times n$ and $g(x,y)$ represents the average of corrupted image in the area defined by $S_{xy.}$

$$p(z) = \frac{1}{\sqrt{(2\pi)\sigma}} e^{\frac{-(z-\mu)^2}{2\pi^2}} \quad (2)$$

Where z represents grey level, μ represents mean of average value of z, σ represents standard deviation of z and σ^2 represents variance of z.

$$f'(x,y) = median_{(s,t)\in Sxy}\ \{g(s,t)\} \quad (3)$$

Where g (s, t) represents the original image and S_{xy} denotes the rectangular sub-image with window size $m \times n$ which is centred at x, y.

The noise level of the image was measured using PSNR (peak signal noise ratio) and expressed the results in decibel [24].

$$MSE = \frac{1}{mn}\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}\prod k(X,Y) - k'(x,y)\prod^2 \quad (4)$$

$$PSNR = 20\log_{10}\left(\frac{max(k(x,y))}{\sqrt{MSE}}\right) \quad (5)$$

PSNR is defined by MSE which is denoted in Eqs. 4 and 5. Where m and n represent the image dimensions, $k(X,Y)$ represents the captured image (noise image) and $k'(x,y)$ represents the quality enhanced image (denoised image).

The labelled groundnut leaf dataset is used to extract colour and texture features. The colour features are extracted based on colour histograms like red, green, and blue. The equation for procuring HSV values from the RGB values is as follows [29].

$$If\ max[R,G,B] = R \rightarrow 60\{(G - B)/ (max[R,G,B] - min[R,G,B]$$
$$If\ max[R,G,B] = G \rightarrow 60\{2\ (B - R)/ (max[R,G,B] - min[R,G,B])\}$$
$$If\ max[R,G,B] = B \rightarrow 60\{4\ (R - G)/ (max[R,G,B] - min[R,G,B])\} \quad (6)$$

$$Saturation = \{max[R,G,B] - min[R,G,B]\}/ max[R,G,B] \quad (7)$$

$$Value = max[R,G,B] \quad (8)$$

Where R-red channel, G-green channel, B-blue channel values.

Eq. (6) represents the Hue colour space.

The goal of GLCM is to characterize texture information from an image that cannot be distinguished visually. It measures the relationship between two pixels in the image through mathematical calculation via grey degree distribution to obtain the image's co-occurrence matrices. The first step of the GLCM calculation is to find the average of the red, green, and blue channels in order to achieve a reasonable grayscale approximation of the image. Then the luminosity method (weighted method) is used to find the greyscale value. The mathematical expression is given in Eq. (9).

$$Grey\ level = 0.299 \times red\ channel + 0.587 \times green\ channel + 0.114 \times blue\ channel \quad (9)$$

0.2999, 0.587, 0.114 are the of red, green blue channel contribution percentage respectively.

Then the second step is the computation of the co-occurrence matrix, which is computed with the help of a general-form grey-tone spatial dependence matrix with its gray-tone values. In the third step, that co-occurrence matrix is added along with a transposed copy of the co-occurrence matrix to generate the symmetric GLCM matrix. This symmetric GLCM has to be normalized. Therefore, in the next step is normalization of the symmetric GLCM matrix obtained by dividing each symmetric GLCM element with the sum of all elements.

Normalization equation:

$$G_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \quad (10)$$

where i and j are the row number and column number, respectively, in an image pixel. In the i and j image windows, V represents the value in the cell. $G_{i,j}$ denotes the probability value recorded for the cells i and j. N represents the number of rows and columns. From that normalized GLCM, fourteen different types of first order and second order statistics texture features [30] are extracted. To avoid multi collinearity and overfitting issue, the extracted features' relationships are analyzed with each other. The relationship analysis result implied that there is no high correlation between the extracted independent features. Hence, all the extracted features directly given to classification model without the use of feature selection technique.

4.2 Intensified multinomial classification technique

Supervised classification techniques are used in machine learning to perform classification on known input categorized data. If the data is linearly separable, then the model classifies the data using hyperplanes with marginal distance, which helps to find the right hyperplane with respect to support vectors. In binary classification, the model classifies the data into two classes. If the data is linearly inseparable (nonlinear data), then it uses the kernel trick to perform multi class classification. In proposed work, the leaf features are linearly inseparable.

Kernel is a mathematical function that transforms low dimensional input space into higher-dimensional space (in a linear equation). The multinomial classification model with its various kernels [31] is denoted as follows:

The training data set of leaf features has r cases represented by x_i

Where $i = 1, \ldots, r$

One vs rest approach employed on the x_i to separate the inseparable leaf features with the help of kernel function. The kernel functions transform the inseparable leaf features from low dimensional space to high dimensional space. Then makes marginal boundaries with the help of hyper parameters. This

marginal boundary facilitates to select best fit hyperplane in multi-dimensional space.

The optimum separation hyperplane (OSH) exists between the different classes in a way that maximizes the margin between them. A hyperplane is defined as

$$w^T \times x_i + b = 0 \tag{11}$$

Where x - data points, w - hyperplane, b - bias.

$x_i \in \{1, -1\}$ In a high dimensional space. Where either 1 or -1 denotes prominent features of specific class and other denotes the rest of the features. Prominent features of specific class is categorized from the rest of the features in high dimensional space. The classification process can be mathematically represented with constraints as showed in Eq. (12)

$$if\ w^T \times x_1 + b \geq -1\ then\ PFSC$$

$$if\ w^T \times x_2 + b \leq 1\ then\ RF \tag{12}$$

Where 1&-1 are constraints to categorize features appropriately.
x_1 and x_2 are marginal boundaries
PFSC- prominent features of specific class
RF- rest of the features

This process continues n times based on the number of classes in input layer. The simplified version of Eq. (12) shown in Eq. (13) as it satisfies both conditions.

$$y_i(w^T \times x_i + b_i) \geq 1 \tag{13}$$

Marginal distance between PFSC and RF are calculated. The obtained distance between marginal boundaries with magnitude is shown is Eq. (14).

$$w^T(x_2 - x_1) = 2. \tag{14}$$

To attain the optimization function and to remove the magnitude of w^T , norm function is utilized. Now the distance becomes

$$(x_2 - x_1) = \frac{2}{\|w\|} \tag{15}$$

The calculated distance based on the constraints needs to be maximized to get the optimization function. This will help the model to reach global minima with minimal error. A reciprocal function is applied to the calculated distance with optimization parameters to attain this, as shown in Eq. (16).

$$Min\ [\frac{\|w\|}{2} + C \sum_{i=1}^{r} \xi_i] \tag{16}$$

Where C- a penalty term
$\sum_{i=1}^{r} \xi_i$- Summation of error values

Penalty term is added to penalize solutions when the ξi vales is very large.

The different kernels analyzed are Polynomial, RBF, and sigmoid. The mathematical function of kernels is denoted as follows

$$f(x) = (\gamma x_1^T \times x_2 + 1)^d \tag{17}$$

$$f(x) = exp(-\gamma \parallel x_1 - x_{2'} \parallel^2) + c \tag{18}$$

$$f(x) = \left(tan\ h\ (\gamma x_1^T x_2 + 1)\right) \tag{19}$$

Where x_i and x_j - low dimensional features,
γ, c, d – hyper parameters.

These kernel functions f(x) parameter values need to be manually analyzed through trial and error methods. It is essential to analyze more than one parameter in order to obtain a suitable decision boundary. But it is a vague process to select a combination of parameters, which leads to time complexity. Hence, to address these issues, several optimization techniques were used during the classification process. In optimization techniques, the range of parameters needs to be initialized as input. It works with a large number of iterations based on the initialized input range to produce the tuned parameters. This process leads to computational complexity. In order to improve classification accuracy and deal with different kinds of complexity, the kernel parameter values are set based on the informatively labelled dataset with the goal of making good decision boundaries.

Replacement of default kernel parameter values (P_{vs}) is defined as follows

$$IHP = IP_{vs} \tag{20}$$

$$IP_{v1} = C_{bild} \quad IP_{v2} = g_{bild} \quad IP_{v3} = d_{bdc}$$

Where bild - informative labelled dataset
bdc - number of classes in dataset.

At the first step, different kernels are initialized, including sigmoid, linear, RBF, and polynomial, with their default parameter values such as c, gamma, and degree. Then specific kernel selection is done based on the procured highest accuracy (H_a). After that, the

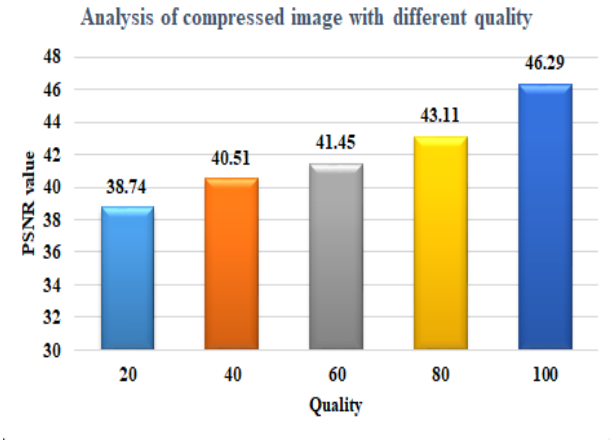


Figure. 2 Analysis of compressed image with different quality

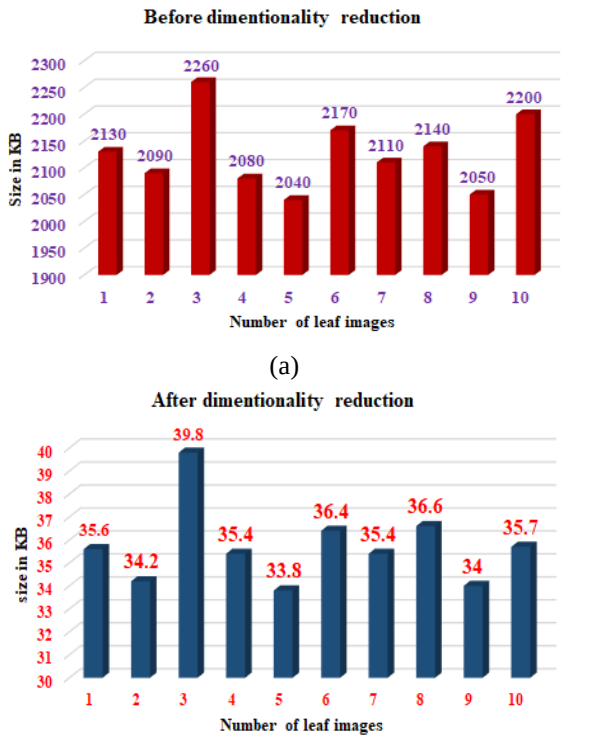
selected kernel parameters were updated with intended parameter values (IP_{Vs}) on the basis of an informative labelled dataset and a class of dataset. Those parameters facilitate the obtaining of an appropriate decision boundary with the highest accuracy (H_{a1}).

Confusion matrix is generated to visualize the result of the classification process. The result generated by the confusion matrix is substituted in performance metrics, namely precision, recall, and f1 score to evaluate the performance of the IMC model. This helps to know classification performance of each class. Precision result provides the percentage of positive predictions. Recall results quantifies the percentage of actual positive predictions. F1 score shows the harmonic mean value of precision and recall. Then the model's overall performance is predicted using averaging methods, namely accuracy, micro average and weighted average. Accuracy (ACC) is calculated by providing equal contribution to all class samples. Macro average (MA) is calculated by providing equal contribution to all classes. Weighted average (WA) is calculated by considering the size of each class samples.

5. Result and discussion

5.1 Image pre-processing

During the time of capture, the camera and imaging conditions remained the same. All the leaf images were stored in JPEG file format with a dimension of 2352×4160. In Fig. 3(a) Shown the sample leaf images' size in KB. Then to perform resizing and dimensionality reduction in a loss less way, image compression techniques were applied at different quality level. The intention of image compression is to obtain feature space from the high



(a)

(b)

Figure. 3 (a) Image compression before dimensionality reduction and (b) Image compression after dimensionality reduction

Table 1. Statistical measures of image with size and quality

Statistical measures	**Original image (Image size - 2.04 MB)**	**Image with 100% quality (Image size - 1.90 MB)**	**Image with 80% quality (Image size - 283 KB)**
Mean	166.646	166.656	166.607
Standard deviation	26.349	26.314	26.330

dimensional input space where the maximal variance is displayed. In Fig. 2 Showed the result of a compressed image of different quality. The compressed image's noise level was measured using PSNR (peak signal noise ratio), and the results were expressed in decibels.

While performing image compression, the size of the images got reduced from megabytes to kilobytes. After the image compression process, the images looked similar, and could not observe any visual differences between them. Histogram analysis was used to determine the significant differences.

In Fig. 4, displayed the green channel histogram

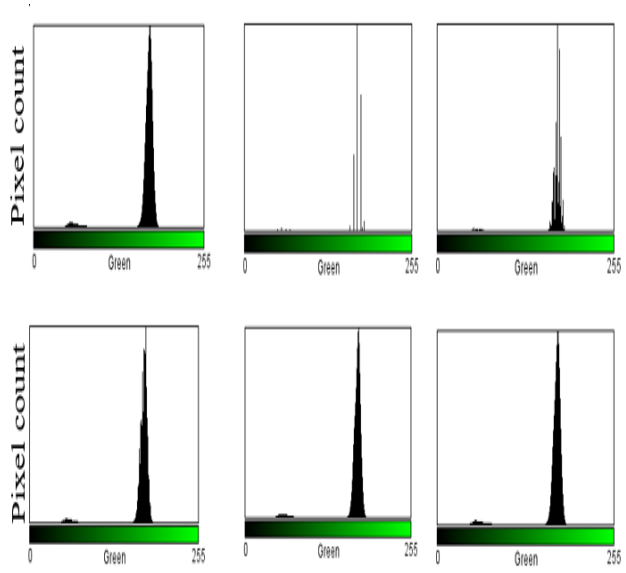


Figure. 4 Green channel histogram of captured images with different quality. First row from left original image, image with 20% quality, and Image with 40% quality, second row from left Image with 60% quality, image with 80% quality, and image with 100% quality

analysis of images with various compression parameters. As the histogram analysis result shows, the original image and the image compressed with 100% quality look similar. The image compressed with 80% quality shows minimal difference when compared with the original image. The histogram analysis of the remaining images reveals the clearly discernible differences.

To see the significant differences more clearly, statistical measures such as the mean and standard deviation were calculated. Table 1. shows statistical measures of the original image and an image with 80 and 100 percent quality, respectively. From that analysis, it is clearly seen that there is not much difference between the original image and the compressed image with 100% and 80% quality, respectively. But images compressed with 100% quality had not been taken for further processing. Instead, the images compressed with 80% quality were taken for further analysis. The reason was that when comparing the original image and the image with 100% quality, the size of the image varied very little: 2.04 MB and 1.90 MB, respectively. But the size of the image with 80% quality was only 283 KB. In general, reducing the size of the images will decrease the processing time, memory space, and complexity. To accomplish this, the images were fixed for further examination. The quality of the image loss percentage is 20. This is a small loss in quality, not even noticeable in histogram analysis or visual observation.

Through image resizing process, the original

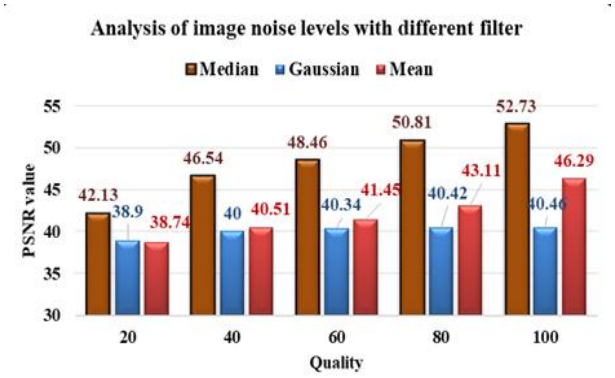


Figure. 5 Analysis of noise levels with median, Gaussian, and mean filter

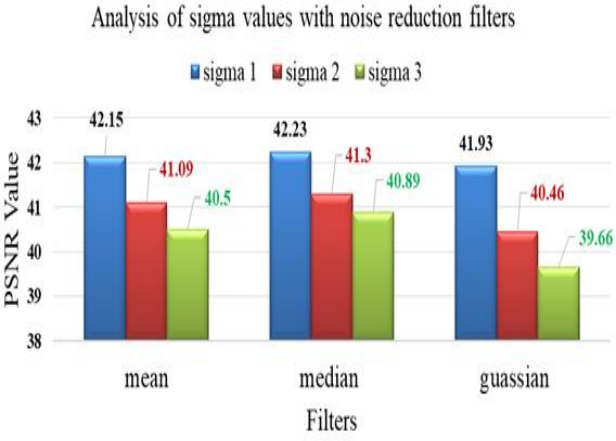


Figure. 6 Analysis of sigma value with noise reduction filters

images got reduced with a dimension of 235×416. In Fig. 3(b) Shown the sample leaf images' size in KB after dimensionality reduction. This would help to minimize the matrix representation of an image, reducing the computational complexity and running time of subsequent processes.

These three filters are applied to the image, and each filter is tried with a different sigma value. as shown in Fig 5. In order to find out how noisy an image of different quality is, it is measured using PSNR (peak signal-to-noise ratio) and given in decibels after noise-reduction filters have been applied.

The result implies that after applying median and Gaussian noise reduction filters, the noise level of the image decreased. But the mean filter does not perform any changes; the PSNR values of the compressed image result are the same as the mean filter result.

In Fig. 6, the result of the image processed with various filters and with various sigma values is shown. It implies that, when comparing the noise reduction filters with their sigma values, three of those filters reduce more noise in $\sigma = 1$. Which is founded out

based on the measured PSNR value of the image. It is noticed that, while increasing the sigma value, the noise level of the image also increased. Therefore, the sigma value is fixed as 1 to all filters for further analysis.

5.2 Performance analysis with machine learning algorithms

In order to fix the best selection boundary, classification algorithms are used with different kernels. The classification accuracy of the proposed IHP based multinomial classification technique is compared with four different learning classification algorithms. A comparison of classification accuracy is shown in Table 2(a), 2(b), and 2(c).

The classification accuracy is below 50% while using a sigmoid kernel. This implies that the sigmoid kernel could not process the unbalanced dataset. But the polynomial kernels performed well with that dataset and provided the accuracy of 95%. In order to increase the accuracy even more, the proposed IHP based multinomial classification is applied. As expected, intensified multinomial classification technique performs well and achieves the best classification accuracy of 97% for the groundnut leaf image dataset.

5.3 Comparative analysis of existing and proposed techniques

The results of various plant disease and stress diagnostic model developed by researchers with the help of statistical, machine learning and neural network approach is compared with the proposed approach as shown in table 3. The data collection was done in papers [15, 24, 25] by utilizing high-cost resources like hyperspectral cameras, multi spectral cameras and thermal cameras. Though the models' results are considerable, in the case of real-time testing, it is difficult for marginal farmers to adopt this kind of data acquisition due to uneconomical reason. The high-quality data collected by these kinds of resources takes a long time for processing during feature extraction. Hyperspectral data contains a large number of spectral bands. Selecting the appropriate band is a daunting task.

Further, the work [24, 27] used optimization techniques to enhance their model performance. In that hyper parameter values are tuned manually by trial and error method. But the proposed model does not require a tuning process. Instead, the intended hyperparameter values are fixed based on the informatively labelled dataset.

Table 2. (a) Comparison of classification accuracy based on groundnut leaf image dataset, (b) Comparison of classification accuracy based on groundnut leaf image dataset, and (c) Classification accuracy of IHP-multinomial classification with groundnut leaf image dataset

(a)

Groundnut leaf dataset	**KNN**			**SVM with SVC**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
D	0.77	0.76	0.77	0.77	0.77	0.77
DIS	0.70	0.64	0.70	0.70	0.64	0.70
C	0.64	0.70	0.67	0.64	0.70	0.67
ACC	0.70			0.71		
MA	0.72	0.70	0.71	0.70	0.70	0.70
WA	0.71	0.70	0.70	0.71	0.71	0.71

(b)

Groundnut leaf dataset	**SVM with sigmoid kernel**			**SVM with Polynomial kernel**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
D	0	0	0	1.00	1.00	1.00
DIS	0	0	0	1.00	0.86	0.92
C	0.45	1.00	0.62	0.90	1.00	0.95
ACC	0.45			0.95		
MA	0.15	0.33	0.21	0.97	0.95	0.96
WA	0.20	0.45	0.28	0.96	0.95	0.95

(c)

Groundnut leaf dataset	**Intensified multinomial classification**		
	Precision	**Recall**	**F1-Score**
D	1.00	1.00	1.00
DIS	1.00	0.90	0.93
C	0.95	1.00	0.95
ACC	0.97		
MA	0.98	0.96	0.96
WA	0.97	0.95	0.95

The authors [26, 27] captured data using a cost-effective RGB camera. They have utilized image pre-processing techniques for noise reduction, feature enhancement and extraction. But in the proposed work, the lossless compression technique was also applied. It helped to reduce dataset size, processing

Table 3. Comparative analysis of recent plant disease and stress diagnostic techniques compared with proposed approach

Plant & diagnosis	**Dataset**	**techniques**	**Performance in %**
Peanut & diseases detection. [15]	Raman Spectroscopy features	PLSR	Accuracy=94
Tea leaves & chlorophyll estimation. [24]	hyperspectral leaf reflectance information	DA KELM	R^2 >60
Groundnut leaves & chlorophyll estimation [25]	Multispectral image features	ANN	RMSE = 0.814
Rice & disease detection [26]	Diseased RGB images	• Otsu • HOG • SVM	Accuracy = 94
Groundnut & disease detection [27]	Diseased RGB images	• Filtering • Augmentation • Optimized Lightweight YOLOv5	Accuracy = 92
Groundnut chlorophyll deficiency detection	RGB images of leaves without notable disease or deficiency symptoms	• Compression • Filtering • GLCM • IMC	Accuracy = 97

time during the feature extraction process, memory space, and computational complexity. Then for noise reduction, different filters are analyzed, and the best one is selected and applied based on the highest PSNR value to remove the noise from groundnut leaf images. The filters are applied directly on images without proper analysis in existing work. The appropriate image quality analysis and pre-processing approach helped achieve less processing time to predict nutrient deficiency before tangible symptoms.

The intensified multinomial classification model's accuracy performance is high than the other machine learning and neural network models mentioned in Table 3.

6. Conclusion

Three different noise reduction filters are applied and analyzed on the acquired groundnut leaf image dataset. The analysis results show that the median filter provides high quality images compared to the mean and Gaussian filters. Then HSV colour features and fourteen types of GLCM texture features are extracted from the filtered images. The retrieved features do not exhibit multicollinearity. So far, no selection technique features have been used. The multinomial classification model performed well with those seventeen types of groundnut leaf features and obtained the highest accuracy of 97%. The proposed DCDGL system could be adaptive for all farmers and support varying amounts of natural light because the model was evaluated with an outdoor leaf image dataset. Cultivars could easily adopt this system for relative chlorophyll concentration determination at low cost and with a fast response. Otherwise, they could photograph groundnut leaves, send them to researchers to determine the relative chlorophyll concentration. As future work, developing an app to analyze sufficient nutrients and find crop-related problems for multiple crops.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Janani Malaisamy and Jebakumar Rethnaraj; methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation Janani Malaisamy; review and editing, Janani Malaisamy and Jebakumar Rethnaraj;"

Reference

[1] Z. Alanna and J. Y. Yoon, "Proximal Methods for Plant Stress Detection Using Optical Sensors and Machine Learning", *Biosensors*, Vol. 10, No. 12, p.193, 2020.

[2] K. Prabhjot, S. Harnal, V. Gautam, M. P. Singh, and S. P. Singh, "An Approach for Characterization of Infected Area In Tomato Leaf Disease Based On Deep Learning And Object Detection Technique", *Engineering Applications of Artificial Intelligence*, Vol. 115, No. 1, p.105210, 2022.

[3] V. G. Dhanya, A. Subeesh, N. L. Kushwaha, D. K. Vishwakarma, T. N. Kumar, G. Ritika, and A. N. Singh, "Deep Learning Based Computer Vision Approaches For Smart Agricultural Applications", *Artificial Intelligence in*

Agriculture Issue, Vol. 6, No. 9, p. 211, 2022.

[4] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer Vision Technology in Agricultural Automation - A Review", *Information Processing in Agriculture*, Vol. 7, No. 1, pp. 1-19, 2020.

[5] M. Janani and R. Jebakumar, "Detection and classification of groundnut leaf nutrient level extraction in RGB images", *Advances in Engineering Software*, Vol. 175, No. 1, p. 103320, 2023.

[6] T. T. Nguyen, T. D. Hoang, M. T. Pham, T. T Vu, T. H Nguyen, Q. T. Huynh, and J. Jo, "Monitoring Agriculture Areas with Satellite Images and Deep Learning", *Applied Soft Computing*, Vol. 95, No. 1, p. 106565, 2020.

[7] K. S. Alves, M. Guimaraes, J. P. Ascari, M. F. Queiroz, R. F. Alfenas, E. S. Mizubuti, and E. M. D. Ponte, "RGB-Based Phenotyping of Foliar Disease Severity under Controlled Conditions", *Tropical Plant Pathology*, Vol. 47, No. 1, pp. 105-117, 2022.

[8] L. Fang, Z. Ma, Q. Wang, H. Nian, Q. Ma, Q. Huang, and Y. Mu, "Plant Growth And Photosynthetic Characteristics Of Soybean Seedlings Under Different Led Lighting Quality Conditions", *Journal of Plant Growth Regulation*, Vol. 40, No. 2, pp. 668-678, 2021.

[9] H. D. Nguyen, V. Pan, C. Pham, R. Valdez, K. Doan, and C. Nansen, "Night-Based Hyperspectral Imaging to Study Association of Horticultural Crop Leaf Reflectance and Nutrient Status", *Computers and Electronics in Agriculture*, Vol. 173, No. 1, p. 105458, 2020.

[10] A. V. Zubler and J. Y. Yoon, "Proximal Methods for Plant Stress Detection Using Optical Sensors and Machine Learning", *Biosensors*, Vol. 10, No. 12, p. 193, 2020.

[11] J. Sun, L. Yang, X. Yang, J. Wei, L. Li, E. Guo, and Y. Kong, "Using Spectral Reflectance to Estimate the Leaf Chlorophyll Content of Maize Inoculated With Arbuscular Mycorrhizal Fungi under Water Stress", *Frontiers in Plant Science*, Vol. 12, No. 5, p. 646173, 2021.

[12] Z. Xin, S. Jun, T. Yan, C. Quansheng, W. Xiaohong, and H. A. Yingying, "A Deep Learning Based Regression Method On Hyperspectral Data For Rapid Prediction Of Cadmium Residue In Lettuce Leaves", *Chemometrics and Intelligent Laboratory Systems*, Vol. 200, No. 5, p. 103996, 2020.

[13] P. Vítek, B. Veselá, and K. Klem, "Spatial and Temporal Variability of Plant Leaf Responses Cascade after PSII Inhibition: Raman, Chlorophyll Fluorescence and Infrared Thermal Imaging", *Sensors*, Vol. 20, No. 4, p. 1015, 2020.

[14] Y. J. Wang, L. Q. Li, S. S. Shen, Y. Liu, J. M. Ning, and Z. Z Zhang, "Rapid Detection Of Quality Index Of Postharvest Fresh Tea Leaves Using Hyperspectral Imaging", *Journal of the Science of Food and Agriculture*, Vol. 100, No. 10, pp. 3803-381, 2020.

[15] J. A. Prananto, B. Minasny, and T. Weaver, "Near Infrared (NIR) Spectroscopy as a Rapid and Cost-Effective Method for Nutrient Analysis of Plant Leaf Tissues", *Advances in Agronomy*, Vol. 164, No. 1, pp. 1-49, 2020.

[16] J. M D. Silva, P. C. Fontes, C. D. Milagres, and E. Garcia Junior, "Application of Proximal Optical Sensors to Assess Nitrogen Status and Yield of Bell Pepper Grown In Slab", *Journal of Soil Science and Plant Nutrition*, Vol. 21, No. 1, pp. 229-237, 2021.

[17] M. D. R. Sereno, M. A. Ayuso, J. L. Pancorbo, J. L. Gabriel, C. Z. Camino, P. J. Z. Tejada, and M. Quemada, "Residual Effect and N Fertilizer Rate Detection by High-Resolution VNIR-SWIR Hyperspectral Imagery and Solar-Induced Chlorophyll Fluorescence in Wheat", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, No. 8, pp. 1-17, 2021.

[18] P. J. Morley, A. S. Jump, M. D. West, D. N. Donoghue. "Spectral Response of Chlorophyll Content during Leaf Senescence in European Beech Trees", *Environmental Research Communications*, Vol. 2, No. 7, pp. 071002, 2020.

[19] R. Sonobe, Y. Hirono, and A. Oi, "Non-Destructive Detection of Tea Leaf Chlorophyll Content Using Hyperspectral Reflectance and Machine Learning Algorithms", *Plants*, Vol. 9, No. 3, pp. 368, 2020.

[20] A. Agarwal, S. D. Gupta, "Assessment of Spinach Seedling Health Status and Chlorophyll Content by Multivariate Data Analysis and Multiple Linear Regression of Leaf Image Features", *Computers and Electronics in Agriculture*, Vol. 152, No. 1, pp. 281-289, 2018.

[21] K. Park, Y. K. Hong, G. H. Kim, and J. Lee, "Classification of Apple Leaf Conditions in Hyper-Spectral Images for Diagnosis of Marssonina Blotch Using Mrmr and Deep Neural Network", *Computers and Electronics in Agriculture*, Vol. 148, No. 5, pp. 178-187, 2018.

[22] I. A. Samborska, H. M. Kalaji, L. Sieczko, W. Borucki, R. Mazur, M. Kouzmanova, and V. Goltsev, "Can Just One-Second Measurement Of Chlorophyll A Fluorescence Be Used To Predict Sulphur Deficiency In Radish (Raphanus Sativus L. Sativus) Plants?", *Current Plant*

Biology, Vol. 19, No. 9, p. 100096, 2019.

[23] H. M. Kalaji, A. Oukarroum, V. Alexandrov, M. Kouzmanova, M. Brestic, M. Zivcak, I. A. Samborska, M. D. Cetner, S. I. Allakhverdiev, and V. Goltsev, "Identification Of Nutrient Deficiency In Maize And Tomato Plants By Invivo Chlorophyll A Fluorescence Measurements", *Plant Physiology and Biochemistry*, Vol. 81, No. 8, pp. 16-25, 2014.

[24] H. Yamashita, R. Sonobe, Y. Hirono, A. Morita, and T. Ikka, "Dissection Of Hyperspectral Reflectance To Estimate Nitrogen And Chlorophyll Contents In Tea Leaves Based On Machine Learning Algorithms", *Scientific Reports*, Vol. 10, No. 1, p. 1, 2020.

[25] H. Qi, Z. Wu, L. Zhang, J. Li, J. Zhou, Z. Jun, and B. Zhu, "Monitoring of peanut leaves chlorophyll content based on drone-based multispectral image feature extraction", *Computers and Electronics in Agriculture*, Vol. 187, No. 8, p. 106292, 2021.

[26] M. E. Pothen and M. L. Pai. "Detection of Rice Leaf Diseases Using Image Processing", In: *Proc. of Fourth International Conf. on Computing Methodologies and Communication*, Erode, India, pp. 424-430, 2020.

[27] H. Wang, S. Shang, D. Wang, X. He, K. Feng, and H. Zhu, "Plant Disease Detection and Classification Method Based On the Optimized Lightweight Yolov5 Model", *Agriculture*, Vol. 12, No. 7, p. 931, 2022.

[28] K. S. Archana and A. Sahayadhas, "Comparison of Various Filters for Noise Removal in Paddy Leaf Images", *International Journal of Engineering and Technology*, Vol. 7, No. 2.21, pp. 372-374, 2018.

[29] N. R. Deepa and N. Nagarajan, "Kuan Noise Filter With Hough Transformation Based Reweighted Linear Program Boost Classification For Plant Leaf Disease Detection", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 6, pp. 5979-92, 2021.

[30] R. M. Haralick, K. Shanmugam, and H. Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, pp. 610-621, 1973.

[31] P. Mathialagan and M. Chidambaranathan, "Computer Vision Techniques For Upper Aero-Digestive Tract Tumor Grading Classification – Addressing Pathological Challenges", *Pattern Recognition Letters*, Vol. 144, No. 4, pp. 42-53, 2021.