International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. IV, No. 08, pp. 13-25, August

Multi-Scale Vision Transformer Architecture for High Resolution Maize Leaf

T Senthil Kumar[1], P. Senthil Kumar[2], Jegan Chellakannu[3]

[1]Assistant Professor, Department of Electrical and Electronics Engineering, Roever Engineering College, Perambalur - 621 220, Tamil Nadu, India. Email: tsenthilkumar92ge25@gmai.com
[2]Assistant Professor, Department of Electrical and Electronics Engineering, Roever Engineering College, Perambalur - 621 220, Tamil Nadu, India. Email: 121212psk@gmail.com
[3]Assistant Professor, Department of Electrical and Electronics Engineering, M.A.M College of Engineering, Trichy - 621105, Tamilnadu, India. Email: jegan4me4pe@gmail.com

Corresponding Author E-mail: [1]tsenthilkumar92ge25@gmai.com

ABSTRACT: Early and accurate identification of Maize Leaf Diseases (MLD) is critical for crop health and achieving maximum production. Automated diagnosis systems are becoming more valuable across today's agricultural sector. These systems assist farmers with early tolerance response treatment of MLD while protecting maize yield and quality. Existing methods struggle to reliably identify MLD for early infections appear as tiny dots and asymmetrical shapes in natural scene images. In this paper, a Multi-Scale Vision Transformer (MS-ViT) is proposed to quickly recognise MLD. Using Maize leaf dataset the spatial proximity and pixel intensity technique removes noise and preserve edges. Then smoothen the maize leaf image using Joint Bilateral Filtering (JBF) to develop high quality images. Subsequently, the processed image is given as an input to the Grabcut segmentation that segments the leaf area by graph-cut algorithm and Gaussian Mixture Model (GMM). After that MLD image is featured by Gray Level Co-occurrence Matrix (GLCM) find the energy, contrast, homogeneity, and correlation values for further analysis. An MS-ViT framework is proposed to enhance the classification of MLD image using maize leaf dataset. Using the proposed framework implemented python software, the approach achieves a higher accuracy of 95% compared to other techniques.

Keywords: Multi-scale Vision Transformer, Joint Bilateral Filtering, Grab cut segmentation, Gray Level Co-occurrence Matrix , Maize Leaf Disease, Multi-Scale Vision Transformer.

1. Introduction

Maize is an essential source of raw materials for many industrial products in worldwide [1]. It is one of the most significant crops, since it is the primary source of energy in human meals and produces high production throughout India. However, plant leaf diseases have an impact on maize crop yield production [2]. Consequently, the Gramineae family, which includes maize, ranks third in terms of overall production and cultivated area, behind rice and wheat. Maize is a great feed for animals as well as for human use [3]. Furthermore, the main disaster affecting maize production is disease, which causes 6–10% of the annual loss. Statistics show that there are over 80 maize diseases in the world [4]. Whereas, sheath blight, rust, northern leaf blight, curcuma leaf spot, stem base rot, head smut, and other diseases are common diseases in maize leaf and have harmful effects [5]. Moreover, lesions caused by diseases such as sheath blight, rust, and northern leaf blight are clearly visible on maize leaves. Rapid and precise diagnosis is essential for many illnesses in order to increase yields, which aid in crop monitoring and prompt disease treatment [6].

MLD image is processed using the pre-processing technique such as, Discrete Wavelet Transform (DWT) to transform the image pixel from the MLD image. DWT makes it possible to analyse MLD at different resolutions and concentrate on different levels. Nevertheless, this technique computationally hard, particularly for high-resolution analysis of complex images [7]. Consequently, Gabor Filtering utilized by edge detection for MLD. It detects the edges of maize leaf with high accuracy and captures the texture analysis. Nonetheless, it is computationally intensive, and its performance heavily depends on complex parameter tuning [8]. Although, to smooth the maize leaf, Gaussian filter is utilized with high accuracy. It reduces noise educes high-frequency noise and effectively suppresses noise. However, this filter require more computation time [9]. Thus, the proposed JBF based pre-processing technique is utilized for removes the noise and preserve edges from MLD image for further Processing.

K-means clustering based segmentation method to cluster the MLD image. It works well with large datasets and have a low computational cost. However, it sensitivity to unusual centroid location and failure to handle irregularly sized clusters [10]. Consequently, Otsu thresholding based segmentation method segment the MLD with high performance. Using Otsu thresholding the diseased maize leaf is segmented suitable for bimodal histogram. Nevertheless, it difficult to deal with MLD images that have several peaks or uneven class variances [11]. Whereas, a semantic segmentation-based approach is employed to segment the MLD image, present accurate boundary detection in maize leaf structures. Nevertheless, it is struggle to large datasets and the possibility of losing area in complex situations [12]. Thus, Grabcut based segmentation approach is proposed in this research that effectively segments the leaf area by graph-cut algorithm and Gaussian mixture model.

From the segmented MLD images, the prominent features of MLD are extracted by using techniques such as, Scale Invariant Feature Extraction (SIFT). SIFT find the key points from the MLD image with high performance. Nonetheless, high computational complexity, possible patents problems, and the requirement for a fixed-size vector for encoding affect the performance of the method [13]. Although, Histogram of Oriented Gradient (HOG) based feature extraction to extracts the MLD image, it demonstrates superior computational efficiency and flexibility to differences in clarification. Nevertheless, it difficult in handling hard areas and complex structures [14]. Moreover, Handcrafted feature based feature extraction is designed and extracts a MLD image, offer better interpretability and efficiency, especially when working with smaller datasets. However, Handcrafted features often fail to capture complex patterns or high-level abstractions in the data 15]. Therefore, the proposed feature extraction technique is utilized GLCM to finds the value energy, homogeneity, correlation, and contrast MLD image.

To classify the MLD image, various classifiers used in previous studies. In existing methods, SqueezeNet is perfectly identified and precisely diagnosed for improved yields prediction. Squeezenet have faster training, efficient processing, it have possible by its compact size and low computing cost. Nevertheless, it decreased accuracy and a limited capacity to learn extreme complex features [16]. Whereas, to identify and classify the MLD image MobileNetV3 model is utilized. The MobileNetV3 model replaced a bias loss function for the cross-entropy loss function to increase the accuracy. Nonetheless, it have the condition for optimization to achieve peak performance with extremely low specifications [17]. To overcome the limitations of existing classification, a proposed MS-ViT is implemented for transforming the multi-scale vision to get high accuracy and efficiency of MLD prediction.

2. Related Work

He et al [18] (2023) have proposed a Faster Region-CNN (Faster R-CNN) for the classification of MLD image. It classifies the maize leaf region with high accuracy and performance. The Region Proposal Networks (RPNs) use common convolutional features to decrease computation and generate region proposals more quickly. Nevertheless, sensitivity to hyper parameters affects consistent performance across different image datasets.

Hu et al [19] (2023) have developed an improved You Only Look Once version 5 (YOLOv5) to identify the maize leaf. It is a lightweight detection algorithm to improve the detection accuracy without increasing the number of model parameters. It is appropriate for real-time applications including its high speed, efficiency, and simplicity. Nevertheless, YOLOv5 struggle with detecting small or densely packed objects, particularly in complex environments.

Li et al [20] (2022) have presented a CNN with Multi-scale feature fusion for MLD detection. CNN automatically and adaptively captures the image from the maize leaf. Multi-scale feature fusions have been shown to be highly valuable for gradient propagation. However, there is a possibility of image quality loss, computationally intensive and their performance heavily depend on the complexity of parameter tuning.

Haque et al [21] (2023) have proposed a Deep Convolutional Neural Network (DCNN) model for recognizing the diseases of maize crop. DCNN identified the maize leaf with higher accuracy and good performance. DCNN is extremely better at recognizing images of maize diseases and outperforms the widely used pre-trained models. However, it takes more time to process the data when large datasets are used.

Waheed et al [23] (2020) have presented a DenseNet to identify MLD image. The normal and diseased maize leaf is determined by extracting the MLD image. DenseNet is used to segregate and classify the numerous types of MLD images with higher accuracy. Every layer in DenseNet receives input from all previous layers, promoting reuse of learned features and improving representation. However, due to dense connections and memory overhead, training is slower.

The contribution of this work is follows

- JBF-based image pre-processing is introduced for MLD images to effectively remove unwanted noise and preserve edges, higher-quality input for analysis.
- Segmentation using the GrabCut algorithm accurately extracts the maize leaf region by applying graph-cut techniques combined with a GMM, preparing the image for precise feature extraction.
- GLCM is performed to compute energy, homogeneity, correlation, and contrast, which serve as significant indicators for identifying and analysing disease patterns in MLD.
- A MS-ViT is proposed to improve the classification and detection of MLD images, achieving high accuracy through efficient attention-based feature representation.

3. Proposed Work

The proposed block diagram in given figure 1 for MLD prediction using DL technique with Maize leaf datasets. The MLD image is randomly chosen as an input maize leaf image dataset, initially pre-processing by JBF that enhance the edges and reduce the unwanted noise for better quality image. Segmentation based on the GrabCut algorithm is used to extract the maize leaf area by generating a bounding box around the region of interest. Additionally, the GMM segments this region into foreground and background components. Following segmentation, the extracted MLD image undergoes feature extraction using GLCM. Key texture features such as energy, homogeneity, contrast, and correlation are computed to facilitate accurate classification in the subsequent stages.

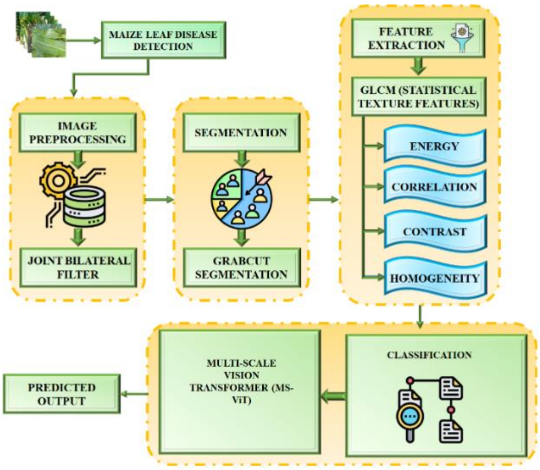

Figure 1: *Block diagram of proposed work*

The featured values are given as an input to classification technique which is based on MS-ViT. The MS-ViT classifier classifies the MLD classes as common rust, gray leaf spot, healthy, northern leaf spot and not maize leaf. The MS-ViT process transforms the multiscale vision to get great accuracy and efficiency for MLD image.

3.1 Pre-processing for MLD by Joint Bilateral Filtering

In the pre-processing stage of MLD identification, the JBF is applied to remove unwanted noise from the input images. This noise often arises from unprocessed or low-quality data within the maize leaf dataset. Therefore, it is essential to reduce and correct for such noise to enhance the quality of the data. The JBF is a frequently employed method and filtering is to reduce noise and enhance the image's edges. JBF functions by limiting diffusion in high-gradient regions and employing diffusion exclusively in low-gradient regions to selectively smooth portions of an image while preserving edges and boundaries.

By changing the photometric weight w_p in Equation (1) to use the guide image G instead of the original image E, the filtering process becomes more adaptive to the structural information provided by G, thereby enhancing edge-aware smoothing. To compute the Dual Bilateral Filter, an additional guide weight is incorporated into the JBF equation. This added weight enhances the filtering process by improving edge preservation and detail accuracy based on the guidance image

$$F(x) = \frac{\sum_{t\in s_m} w_s(\|t\| w_p\big(E(x)-E(x+t)\big)E(x+t)}{\sum_{t\in s_m} w_s(\|t\| w_p\big(E(x)-E(x+t)\big)} \quad (1)$$

The window s_m is usually a square $[-m,m] \times [-m,m]$. The function of $w_s$decreases symmetrically with the distance t from the center of s_m

$$\sum_{t\in s_m} w_s(\|t\| \varphi(F(x) - E(x+t))^2) \quad (2)$$

Thus the filter output value $F(x)$ attained at the lowest possible cost:

$$\sum_{t\in s_m} w_s(\|t\| w_g\big(G(x) - G(x+t)\big)\varphi((F(x) - E(x+t))^2) \quad (3)$$

$$q_t = w_s(\|t\| w_g\big(G(x) - G(x+t)\big) \quad (4)$$

Where, $q_t = w_g \;\; w_s$ denotes the spatial weight and w_g is the guide weight from the maize leaf image. The photometric noise model is characterized by φ, and the new similarity indicator is w_g. Whereas, JBF enhances image quality by preserving important details like edges and decreasing the impact of noise. JBF technique is used for improving the quality and diagnostic value of MLD images. The JBF enhance the edges and remove the noise for better quality MLD image. The processed image is fed to segmentation based on Grabcut algorithm.

3.2 Segmentation by Grabcut Algorithm

GrabCut is an iterative image segmentation method that combines Graph cuts and GMM to segment the MLD image. The system iteratively improves the segmentation, eventually producing accurate results, by rectangles designating the foreground and background. GrabCut, an efficient image segmentation algorithm, employs rectangular bounding boxes to effectively separate foreground objects from the background in input images. Pixels outside the bounding box are regarded as background for maize leaf. Using the GrabCut graph-cut technique, the Region of Interest (RoI) is identified by labelling a lasso or a bounding box on the MLD image. Because of two improvements iterative estimate and partial labelling grabcut method requires less user interference than the original graph cut method in figure 2. Also, the sequencer first estimates the color distribution of the background and foreground using the GMM. In this context, the GMM comprises two components: one for the background and one for the foreground. This dual-component structure allows the model to effectively differentiate between the regions of interest in the image based on their statistical properties.

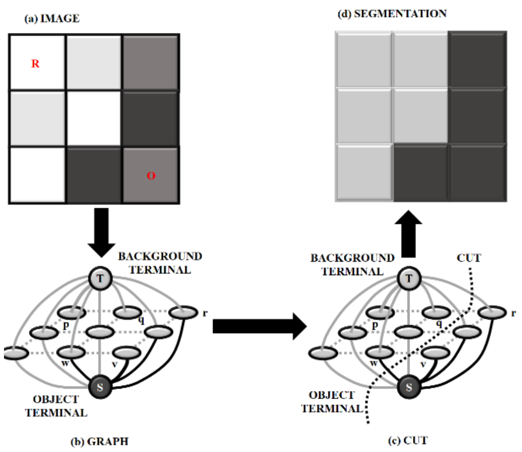

Figure 2: *Graph cut model*

Each Gaussian component needs its covariance matrix with k components, each of which represents a distinct cluster in the MLD image, allowing for flexible modeling of correlation between the features. To GMM computationally flexible, an additional k vector, $k = \{k_1, \ldots, k_n, \ldots k_N\}$, is added with $k_n \in \{1, \ldots K\}$. this gives each pixel in the MLD image a unique GMM element, one of which come from the foreground or background model depending on $\propto = 0$. This indicates that the pixel is assigned to the background model if $\propto = 0$ and to the foreground model if $\propto$ equals 1. T, And data item U are used to define Gibbs energy based on the GMM component variables k.

$$E(\propto, k, \theta, z) = U(\alpha, k, \theta, z) + T(\alpha, z) \quad (5)$$

The informational point U uses k, the Gaussian variable, to z the image data and gray histogram θ in order to compute the pixel label α.

$$(\alpha, k, \theta, z) = \sum_n D(\alpha_n, k_n, \theta, z_n) = -\sum_n (\log p(z_n | \alpha_n, k_{n,} \theta) + \log \pi(\alpha_n, k_n)| \quad (6)$$

In this case, p is the Gaussian probability distribution of pixel z_n to Gaussian pixels k_n, and θ, the current pixel α_n. The grabCut algorithm's main phases and general operation are as follows:

3.2.1 Initialization

The segmentation process begins by labelling the background and foreground regions. These labels, denoted as α, are used to initialize the GMMs for both regions.

3.2.2 GMM Model Processing

Once initialized, the GMMs are used to model the color distributions of the labelled regions. Each pixel in the image is assigned to the GMM component that minimizes the distance function, as defined by:

$$k_n = arg_{k_n}^{min} D_n(\alpha_n, k_n, \theta, z_n) \quad (7)$$

Subsequently, the GMM parameters θ are optimized by minimizing the energy function:

$$\theta = arg_{\theta}^{min} U(\alpha, k, \theta, z) \quad (8)$$

3.2.3 Graph Construction

An undirected graph is then constructed using the labelled foreground and background pixels. This graph incorporates both data terms representing well a pixel fits the GMM and smoothness terms.

3.2.4 Segmentation

The final segmentation is obtained by minimizing the total energy function, which combines data fidelity and smoothness. This is achieved using the minimum cut algorithm, which partitions the graph into foreground and background regions based on the optimal energy configuration

$$\min_{\{\alpha_n n \in T_{CP}\}} \min_{k} E(\alpha, k, \theta, z) \quad (9)$$

The GMM is a probabilistic model that represents the probability distribution of the maize leaf dataset as a weighted sum of multiple Gaussian distributions. Next, the segmented output is fed as an input to feature extraction technique for MLD image.

3.3 Feature extraction by Statistical Texture feature

A statistical method for texture feature extraction, the GLCM is employed to analyze MLD images. GLCM evaluates the spatial relationship between pixels, allowing for the extraction of meaningful texture features. Initially, the images are converted to grayscale to retain essential image information. From these grayscale images, texture indices are computed and selected based on their strong correlation with MLD characteristics. These texture features serve as important inputs for further analysis. This dimensionality reduction is necessary to balance between computational efficiency and texture representation accuracy. Figure 3 shows the GLCM feature extraction for MLD.

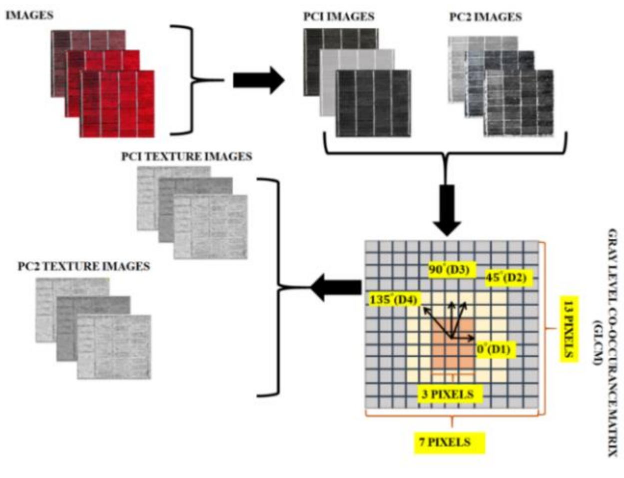

Figure 3: *Statistical texture feature*

Statistical features in GLCM experimentally determined four essential features: the correlation, entropy, contrasts, and Angular Second Moment (ASM) features. Prior to computation, the GLCM needs to be normalized. An equation expressed as follows

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

Contrast: A distinct moment of GLCM, the contrast quantifies the spatial frequency of a MLD image. It is the variation between the neighbouring set of pixels' highest and lowest values. The image's local variances are measured by the contrast texture. Low contrast images have low values and a GLCM concentration term around the major diagonal.

$$Contrast = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(|i-j|)^2 P_{\varphi,d}(i,j) \quad (11)$$

Angular Second Moment (ASM): Quantifies textural homogeneity, it identifies the abnormalities in the image textures. The ASM reach a maximum value of one. A constant periodic form of the gray level distribution results in higher values.

$$Asm = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{\varphi,d}^2(i,j) \quad (12)$$

$$Entropy = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_{\varphi,d}(i,j) log_2 P_{\varphi,d}(i,j) \quad (13)$$

$$Homogenity = \sum_{j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (14)$$

Correlation: A measure of the linear dependencies between the image's grayscale tones is called correlation

$$Correlation = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}[(i,j)P_{\varphi,d}(i,j)-\mu_x\mu_y]}{\sigma_x\sigma_y} \quad (15)$$

Where in equation,

$$\mu_x = \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} P_{\varphi,d}(i,j) \quad (16)$$

$$\mu_y = \sum_{i=0}^{N-1} j \sum_{j=0}^{N-1} P_{\varphi,d}(i,j) \quad (17)$$

$$\sigma_x^2 = \sum_{i=0}^{N-1}(i-\mu_x)^2 \sum_{j=0}^{N-1} P_{\varphi,d}(i,j) \quad (18)$$

$$\sigma_y^2 = \sum_{i=0}^{N-1}(i-\mu_y)^2 \sum_{j=0}^{N-1} P_{\varphi,d}(i,j) \quad (19)$$

Where, $\mathrm{P(i,j)}$ are the features i, j of the standardised regular GLCM, N denotes the sum of gray levels, μ is the mean of GLCM and σ^2 denotes the variance of the intensities. By using the GLCM values the features are extracted. The extracted MLD image is given to the proposed MS-ViT for further classification.

4. Classification by Multi-scale Vision Transformer

In the proposed work, MS-ViT is employed to reduce the computational complexity involved in processing MLD images by utilizing extracted feature values. By incorporating multi-scale attention, the model selectively focuses on important regions, significantly lowering the computational burden. A key innovation in this architecture is the integration of a Transformer model for classifying MLD images. The Transformer's self-attention mechanism allows it to simultaneously consider contextual relationships across the entire input sequence, marking a notable shift from earlier models that processed sequentially. In computer vision tasks, attention mechanisms are primarily applied to explore the relationships between different image regions or blocks. The basic architecture of MS-ViT, as illustrated in Figure 4, and demonstrates multi-scale attention strategy that is effectively implemented.

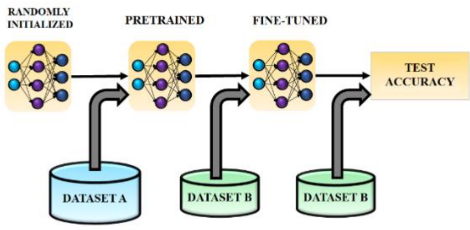

Figure 4: *Basic Multi-Scale Vision Transformer*

MS-ViT techniques enable the model to distinguish between relevant and irrelevant blocks, allowing it to focus on the most important regions for accurate object identification. As a result, the model improves its capacity to identify and process relevant features by learning to concentrate on important information. Three vectors are mainly used in the computation of attention: query, key, and value. A query and a collection of key-value pairs are mapped to an output, which is the weighted sum of values, during the whole computation process. Equation (20) defines the relevance between the current key and the query, which determines the weights.

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{C}}\right) V \quad (20)$$

Where the inputs are $Q \in RNq \times C, K \in RNk \times C, and\ V \in RNv \times C$. To address issues with weight concentration and gradient vanishing, √C is included. To improve the performance of self-attention layers in light of the evolution of attention mechanisms multi-head self-attention is introduced.

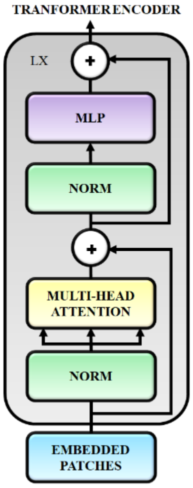

Figure 5: *Transformer encoder for multi-head attention*

This enables the model to concentrate on several informational facets, including the following equations.

$$Q_i = XW_i^Q, K_i = XW_i^k, V_i = XW_i^V, i = 1,2, \dots 8 \quad (21)$$

$$head_i = Attention(Q_i, K_i, V_i), i = 1, \dots, 8 \quad (22)$$

$$MultiHead(Q, K, V) = concact(head_1, \dots \dots head_8)W^0 \quad (23)$$

By splitting the input into multiple heads along the channel dimension, the model creates distinct representation subspaces, each corresponding to a separate attention head. Initially, multi-scale computations are performed to capture features at various resolutions. Building upon this foundation, the module then applies vision transformer operations to enhance contextual understanding. To further improve efficiency, pooled multi-head attention is employed, which facilitates multi-scale processing and reduces the sequence length of the output tokens. Figure 5 shows the transformer encoder for maize leaf. Here is a description of the multi-scale mechanism for the maize leaf. The channel capacity is increased hierarchically while decreasing the spatial resolution, begin with the input resolution and smaller channel sizes

$$S_Q = \left(L_Q, D\right) = (L_k, D), S_V = (L_V, D) \quad (24)$$

Where, L is equal to T × H × W. Consequently, the query's sequence length changed in order to decrease spatial resolution. Thus the proposed MS-ViT performs high classification accuracy and good performance for the MLD image.

6. Result and Discussion

In result and discussion the MLD images are predicted using the proposed MS-ViT. For identifying MLD the maize leaf dataset is taken from the kaggle.com implemented using Python software. There are five classes of maize leaf are common rust, gray leaf spot, healthy, northern leaf bright, not maize leaf. The total train images have the count of 7079, and the total test images have the count of 1773.

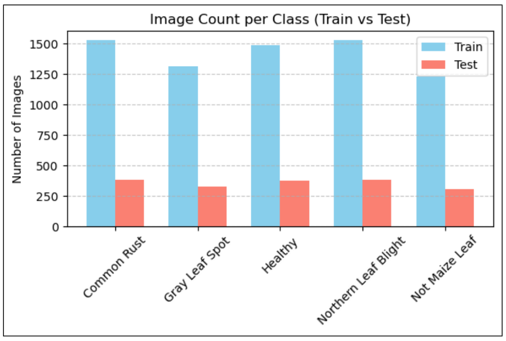

Figure 6: *Input test and train count*

Figure 6 shows the input test and train count per class for maize leaf. The total train images have the count of 7079, where the common rust have 1525 images, gray leaf spot have 1313 images, healthy have 1487 images, northern leaf blight have 1526 images and not maize leaf have 1228 images. And the total test images have the count of 1773 where the common rust have 382 images, gray leaf spot have 329 images, healthy have 372 images, northern leaf blight have 382 images and not maize leaf have 308 images.

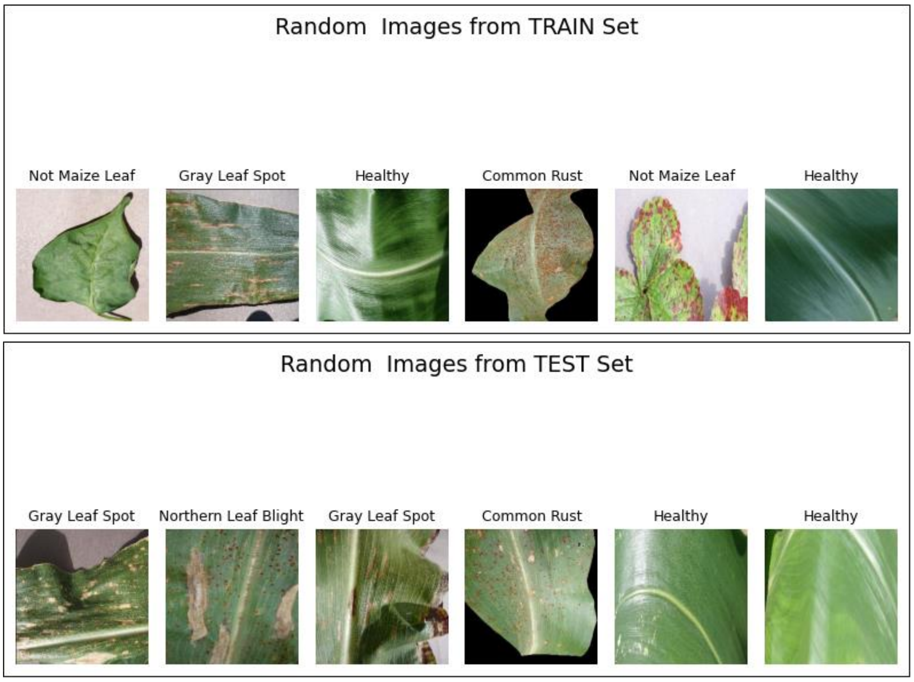

Figure 7: *Random image train and test maize leaf image*

Figure 7 shows random images in train and test dataset of maize leaf and MLD image. The random images are taken from the maize leaf dataset for further process. Here, the maize leaf classes of common rust, not maize leaf, northern leaf blight, gray leaf spot healthy, images are randomly chosen from maize leaf dataset to test and train for further analysis.

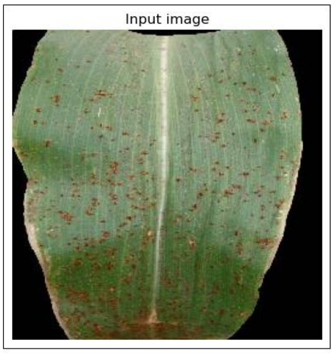

Figure 8: *Input maize leaf*

Figure 8 shows the input maize leaf image. The input image is from the maize leaf dataset randomly chosen test and train images. The MLD image is a raw image from maize leaf dataset from kaggle.com uses the following methods such as pre-processing, segmentation, feature extraction and classification for further processes.

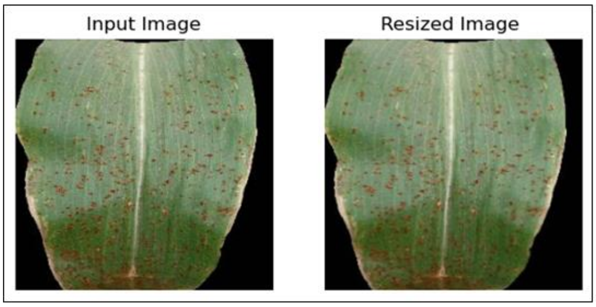

Figure 9: *Input image into resize image*

Figure 9 shows the input image to resize image. Here the input MLD image is resized using image resizing technique. The noises present in the original image are removed and develop the better quality image. Next, step the better quality MLD image is given as an input to grayscale conversion technique.

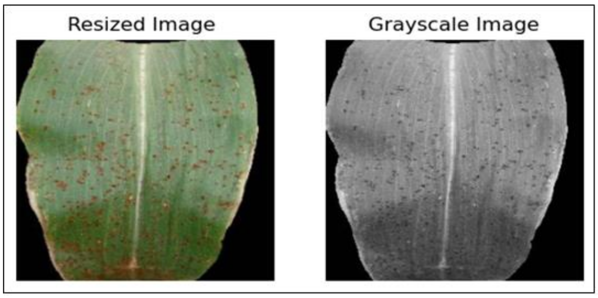

Figure 10: *Resize image into grayscale image*

Figure 10 shows the resized image into gray scale image for MLD approach. Here resized image is transformed into grayscale image for enhancing the MLD region. Resizing involves changing the dimensions of an image to a fixed width and height, which standardize input data. Converting an image to grayscale simplifies the data by reducing the number of color channels from three (red, green, and blue) to one.

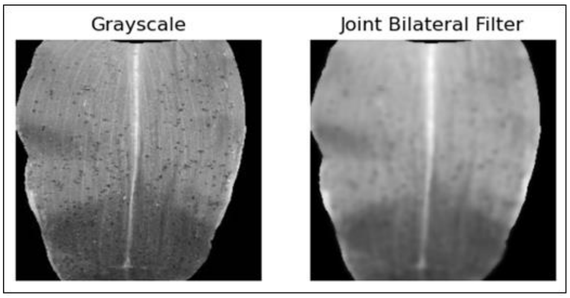

Figure 11: *Grayscale image into joint bilateral filtering*

Figure 11 shows the grayscale image into JBF for MLD image. Here the MLD image is enhanced and preserves the edges by JBF, which also detected and removed the noise from disease affected image. Then the filtered MLD image is given as an input to Grabcut algorithm for further segmentation.

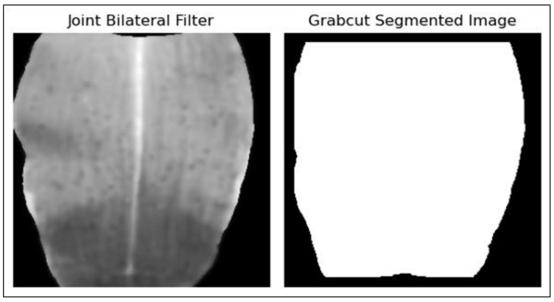

Figure 12: *JBF image into Grabcut segmentation*

Figure 12 shows the JBF image into grabcut segmentation for MLD image. Here the grabcut calculates the foreground and background using the graph cut and GMM. The graph cut calculates the boundary box and GMM calculates the foreground and background region. Then the segmented image is fed in to input to feature extraction here statistical texture feature is used.

Table 1: *Statistical Texture Feature for MLD Image*

Key features	Values
Energy	0.7886
Contrast	478.12
Homogeneity	0.9926
Correlation	0.9802

Table 1 shows the statistical texture feature for MLD image. The statistical texture feature shows the value for energy of 0.7886, contrast of 478.12, and homogeneity of 0.9926 and correlation of 0.9802 for MLD image.

6.1 Performance metrics

Performance metrics like accuracy, sensitivity and specificity which indicate the likelihood that a MLD be correctly detected and appropriately diagnosed as the probabilities calculated by the MS-ViT. The performance metrics in this study explained in table 2.

Table 2: *Performance Metrics*

Performance metrics	Formula
Accuracy	$\frac{(TN+TP)}{TS}$
Precision	$\frac{(TP)}{TP+FP}$
Recall	$\frac{(TP)}{(TP+FN)}$
$F1-score$	$\frac{(2 * \text{Precision} * \text{Recall})}{(Precision\ + Recall)}$

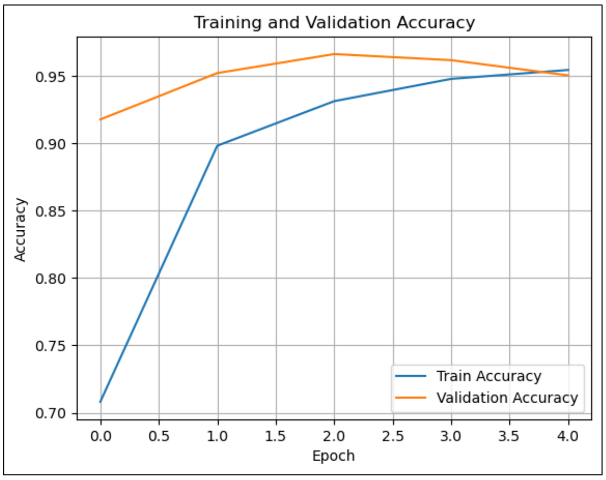

Figure 13: *Training and validation accuracy*

Figure 13 shows the training and validation accuracy. The X-axis denotes the epochs and Y-axis denotes accuracy for MLD image. The model accuracy graph illustrates increasing accuracy of 95%, with the model achieving high performance and minimal over fitting.

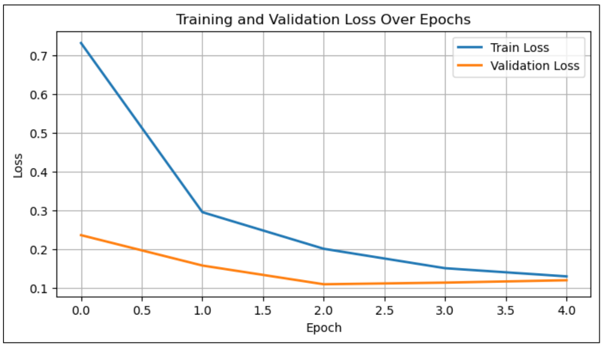

Figure 14: *Training and validation loss*

Figure 14 shows the training loss and validation loss for MLD image. The X-axis indicates the epochs and Y-axis indicates the loss. Training and validation loss is reliably declining in training and validation loss, representing improved learning and reduced error.

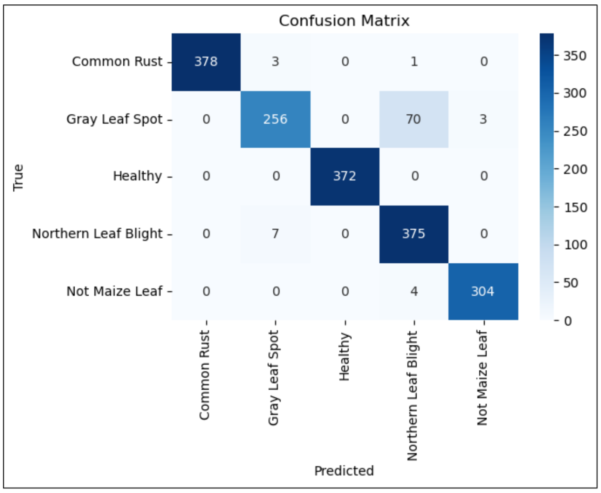

Figure 15: *Confusion matrix for MS-ViT*

Figure 15 shows the confusion matrix for MS-ViT. Proposed method evaluates the model's performance for five classes' common rust, gray leaf spot, healthy, northern leaf blight, and not maize leaf. Common rust have the value 378, gray leaf spot have the value of 256, healthy have the value of 372, northern leaf blight have the value of 375 and not maize leaf have the value of 304 respectively.

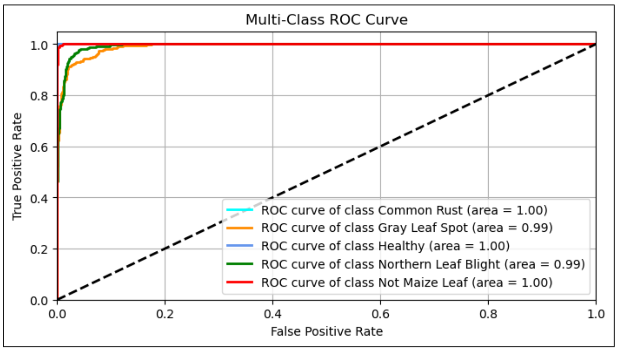

Figure 16: *Multi-class ROC curve*

Figure 16 shows the multi-class ROC curve for proposed MS-ViT continues to have the highest AUC value of 1 for common rust, healthy, not maize leaf and for gray leaf spot, and northern leaf blight the AUC value of 0.99 demonstrating the unique overview and flexibility of MLD approach.

6.2 Comparison for the proposed method

The comparison of the proposed method with existing method is presented in table 3. The Faster R-CNN have the classification accuracy of 88% [18] and DenseNet have the accuracy of 94% [22]. The proposed MS-ViT have the higher accuracy of 95% compared to the existing method in table 3. The proposed MS-ViT have better performance for MLD image.

Table 3: *Comparison for the proposed method*

Study	Method	Accuracy
He *et al.* [18]	Faster R-CNN	88%
Waheed *et al.*[22]	DenseNet	94%
Proposed approach	**MS-ViT**	95%

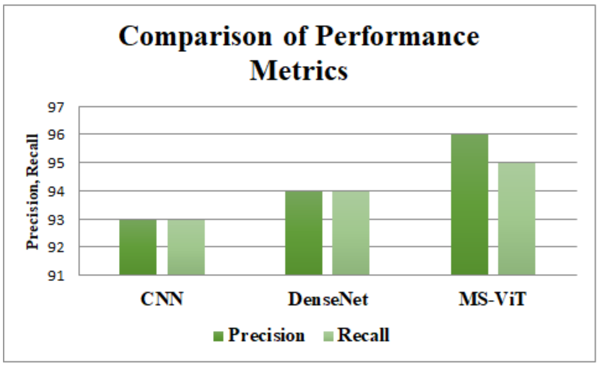

Figure 17: *Comparison of performance metrics*

Figure 17 shows the comparison of performance metrics for precision and recall value of proposed work and the existing work. The CNN have precision value of 93% and recall value of 93% [10], DenseNet have precision value of 94% and recall value of 94% [22] and proposed MS-ViT have the precision value of 96% and recall of 95% for identifying MLD image. The proposed precision and recall value is high compared to the existing method.

7. Conclusion

In this paper, MS-ViT is proposed for the detection and classification of MLD image. The model successfully classified five classes of common rust, healthy, northern leaf blight, not maize leaf and gray leaf spot. The preprocessing stage JBF effectively preserves the edges and removed noise. For GrabCut segmentation, to determine the bounding box and a GMM to distinguish between foreground and background regions. Statistical Texture feature effectively calculates the vales of energy, contrast, homogeneity and correlation from the MLD images. Finally, the proposed MS-ViT classifies the MLD image with better accuracy and improving early detection of MLD. The performance evaluation, conducted on the maize leaf dataset, demonstrates the superior capabilities of the MS-ViT method that achieves an improved accuracy of 95%, precision value of 96%, recall of 95% and performance in contrast to the surviving methods.

References

1. Vivek Sharma; Ashish Kumar Tripathi; Purva Daga; Himanshu Mittal, Year: 2024, "ClGanNet: A novel method for maize leaf disease identification using ClGan and deep CNN", Signal Processing: Image Communication, Vol: 120, pp. 117074.

2. Yan Zhang; Shiyun Wa; Yutong Liu; Xiaoya Zhou; Pengshuo Sun; Qin Ma, Year:2021, "High-accuracy detection of maize leaf diseases CNN based on multi-pathway activation function module", Remote sensing, Vol: 13, No: 21, pp. 4218.

3. Lixiang Sun; Jie He; Lingtao Zhang, "CASF-MNet: Multi-scale network with cross attention mechanism and spatial dimension feature fusion for maize leaf disease detection", Crop Protection, Vol: 180, pp. 106667.

4. Li Ma; Qiwen Yu; Helong Yu; Jian Zhang, "Maize leaf disease identification based on yolov5n algorithm incorporating attention mechanism", Agronomy, Vol: 13, No: 2, pp. 521.

5. Mamillapally Nagaraju; Priyanka Chawla, Year: 2023, "Maize crop disease detection using NPNet-19 convolutional neural network", Neural Computing and Applications, Vol: 35, No: 4, pp.3075– 3099.

6. Hanmi Zhou; Yumin Su; Jiageng Chen; Jichen Li, Linshuang Ma; Xingyi Liu; Sibo Lu; Qi Wu, Year: 2024, "Maize leaf disease recognition based on improved convolutional neural network shufflenetv2:, Plants, Vol: 13, No: 12, pp. 1621.

7. Navin Rajpal; Jyotsna Yadav; Kalyan Kumar Mondal, Year: 2024, "Multi-resolution analysis and deep neural network architecture based hybrid feature extraction technique for plant disease identification and severity estimation", Evolutionary Intelligence, Vol: 17, No: 2, pp. 1163– 1183.

8. J. Dhakshayani; B. Surendiran, "GF-CNN: an enhanced deep learning model with gabor filters for maize disease classification", SN Computer Science, Vol: 4, No: 5, pp. 538.

9. Dikang Dai; Dikang; Peiwen Xia; Zeyang Zhu; Huilian Che, Year: 2023, "MTDL-EPDCLD: A multi-task deep-learning-based system for enhanced precision detection and diagnosis of corn leaf diseases", Plants, Vol: 12, No: 13, pp. 2433.

10. Nidhi Kundu; Geeta Rani; Vijaypal Singh Dhaka; Kalpit Gupta; Siddaiah Chandra Nayaka; Eugenio Vocaturo; Ester Zumpano, "Disease detection, severity prediction, and crop loss estimation in MaizeCrop using deep learning", Artificial intelligence in agriculture, Vol: 6, pp. 276– 291.

11. PP. Rajeena; Fathimathul; Aswathy SU; Mohamed A. Moustafa; Mona AS Ali, Year: 2023, "Detecting plant disease in corn leaf using efficientnet architecture—an analytical approach", Electronics, Vol: 12, No: 8, pp. 1938.

12. Shuo Chen; Kefei Zhang; Suqin Wu, Ziqian Tang, Yindi Zhao, Yaqin Sun; Zhongchao Shi, Year:2023, "A weakly supervised approach for disease segmentation of maize northern leaf blight from UAV images", Drones, Vol: 7, No: 3 pp. 173.

13. Siddharth Singh Chouhan; Uday Pratap Singh; Sanjeev Jain, Year: 2021, "Automated plant leaf disease detection and classification using fuzzy based function network", Wireless Personal Communications, Vol: 121, No: 3, pp. 1757– 1779.

14. Javaid Ahmad Wani; Sparsh Sharma; Malik Muzamil; Suhaib Ahmed; Surbhi Sharma; Saurabh Singh, Year: 20222, "Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: Methodologies, applications, and challenges", Archives of Computational methods in Engineering, Vol: 29, No: 1, pp. 641– 677.

15. Yashwant Kurmi; Suchi Gangwar, Year: 2022, "A leaf image localization based algorithm for different crops disease classification", Information processing in agriculture, Vol: 9, No: 3, pp. 456– 474.

16. Prasannavenkatesan Theerthagiri; A. Usha Ruby; J. George Chellin Chandran; Tanvir Habib Sardar; Ahamed Shafeeq BM, Year: 2024, "Deep SqueezeNet learning model for diagnosis and prediction of maize leaf diseases", Journal of Big Data, Vol: 11, No: 1, pp. 112.

17. Chunguang Bi; Suzhen Xu; Nan Hu; Shuo Zhang; Zhenyi Zhu; Helong Yu, Year: 2023, "Identification

method of corn leaf disease based on improved Mobilenetv3 model", Agronomy, Vol: 13, No: 2, pp. 300.

18. Jie He; Tao Liu; Liujun Li; Yahui Hu; Guoxiong Zhou, Year: 2023, "MFaster r-CNN for maize leaf diseases detection based on machine vision", Arabian Journal for Science and Engineering, Vol: 48, No: 2, pp. 1437– 1449.

19. Yanxin Hu; Gang Liu; Zhiyu Chen; Jiaqi Liu; Jianwei Guo, Year: 2023, "Lightweight one-stage maize leaf disease detection model with knowledge distillation", Agriculture, Vol: 13, No: 9, pp. 1664.

20. Ying Li; Shiyu Sun; Changshe Zhang; Guangsong Yang; Qiubo Ye, "One-stage disease detection method for maize leaf based on multi-scale feature fusion," Applied Sciences, Vol: 12, No: 16, pp. 7960.

21. Md Ashraful Haque; Sudeep Marwaha; Chandan Kumar Deb; Sapna Nigam; Alka Arora, Year: 2023, "Recognition of diseases of maize crop using deep learning models", Neural Computing and Applications, Vol: 35, No: 10, pp. 7407– 7421.

22. Abdul Waheed; Muskan Goyal; Deepak Gupta; Ashish Khanna; Aboul Ella Hassanien; Hari Mohan Pandey, Year: 2020, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf", Computers and Electronics in Agriculture, Vol: 175, pp. 105456.