

A Hybrid Framework Integrating Novel Segmentation and Hybrid Feature Engineering with M-SVM for the Detection and Classification of Wheat Rusts and Septoria

Ranjeet Singh¹, Dayashankar Singh²

Abstract: This study introduces a hybrid framework for the precise detection and classification of wheat rusts (Leaf Rust, Stem Rust, and Stripe Rust) and Wheat Leaf Septoria, aiming to enhance diagnostic accuracy and support effective disease management. The framework comprises key stages: lesion segmentation, Hybrid Feature Engineering and disease classification. Advanced machine learning and image enhancement methods are combined to effectively segment lesions using a multilevel hybrid segmentation model, followed by feature extraction involving shape, color, texture, and edge properties refined through PCA, entropy and skewness. To address challenges related to limited datasets and model overfitting, robust handcrafted features are utilized instead of deep feature dependency. Finally, a Multi-Class Support Vector Machine (M-SVM) is employed for classification.

Evaluated on four wheat diseases—septoria, stem rust, leaf rust, and stripe rust—the framework achieved a segmentation performance index of 98.21% on 40 images per disease and a classification accuracy of 99.1% using 10-fold cross-validation on 1,000 images. With an F1-score of 0.989, AUC of 0.991, and minimal error rates, the system outperforms existing methods, averaging 5.8 seconds per image, demonstrating its robustness and efficiency.

Keywords: Wheat Rusts and Septoria, Adaptive K-Means Clustering, Gaussian Mixture Model (GMM), PCA, Feature Extraction, Multi-Class Support Vector Machine (M-SVM).

¹Department of Information Technology & Computer Application, Madan Mohan Malaviya University of Technology, Gorakhpur, India, 273010 (ranjeetsinghcse@gmail.com)

²Department of Information Technology & Computer Application, Madan Mohan Malaviya University of Technology, Gorakhpur, India, 273010 (dssitca@mmmut.ac.in)

1. Introduction

Wheat ranks among the most extensively cultivated and economically significant crops worldwide, serving as a staple food for millions. However, wheat production faces significant challenges from various diseases that reduce 15–20% annual yield losses and also adversely affect grain quality [43]. Among these, wheat rusts (Leaf Rust, Stem Rust, and Stripe Rust) and Wheat Leaf Septoria are particularly destructive, causing substantial losses to farmers worldwide. Detecting these diseases promptly and accurately is essential for their effective control and reduction [12]. Traditional disease detection methods, such as visual inspection and manual laboratory tests, typically involve lengthy procedures and require advanced expertise, making them impractical for large-scale agricultural monitoring. Therefore, developing automated, dependable, and effective tools is essential to support the timely identification of wheat diseases.

Identifying and categorizing wheat rust and septoria spots using the proposed framework typically involves four main stages: preprocessing, segmentation, feature extraction, and classification. In this study [53], the authors proposed FERSPNET-50, a hybrid panoptic segmentation approach for real-time WRD (Wheat Rust Diseases) detection using images collected in varied weather conditions. The GNet model classifies weather, and Faster Region-Based Convolutional Network (FRCNN) detects wheat patches. A semi-automatic annotation method and Pyramid Scene Parsing Network (PSPNet) improve rust detection accuracy. The proposed method achieved 0.97 precision, outperforming RetinaNet (0.82) and YOLOV4 (0.88). However, this method's dependence on high-quality labeled data is a drawback.

Traditional approaches and machine learning (ML)-based methods for detecting wheat diseases continue to encounter several unresolved issues. Among the most notable are the inconsistencies in data quality, limited preprocessing efficiency, and challenges in choosing appropriate learning models. Although numerous researchers have attempted to mitigate these problems, the progress achieved so far remains partial and leaves room for further improvement.

For example, [33] designed a machine-learning framework that focused on wheat disease detection by gathering diverse field data under different environmental conditions. Their procedure involved segmenting and resizing the images to improve the preprocessing stage. Using this refined dataset, the model was trained and achieved an accuracy of 98.7%, which demonstrated a notable enhancement over prior methods.

Similarly, [53] highlighted that wheat—one of the most vital cereal crops globally—contributes nearly one by fifth of the daily caloric and protein intake for humans. Unfortunately, rust diseases can result in yield reductions of up to 30%, severely affecting food security. To counter this, accurate segmentation and localization of diseased regions are essential. Recent studies have adopted advanced segmentation strategies, including GrabCut, Watershed, and U2-Net, to analyze stripe rust. Among these, the U2-Net model, built upon the ResNet-18 backbone, achieved a classification accuracy of 96.19%. This result underscores how precise segmentation directly strengthens disease classification performance and supports the selection of effective computational approaches.

Wheat diseases also heavily influence crop productivity, as observed by [55]. Current deep learning frequently causes problems with variations in image aspects and orientations, leading to suboptimal performance. To overcome this, a convolutional neural network (CNN)-based method was developed, incorporating an adaptive labeling mechanism and a

localization assessment framework. This approach outperformed conventional disease localization methods with a mean average precision and a mean recall of 60.8% and 73.8% respectively, when used on the WDF2023 dataset. Furthermore, [13] highlighted the importance of early disease detection for controlling disease spread. CNN-based methods have proven particularly effective in this domain. Their research proposed the Framework for Integrated Deep Learning, which combines ensemble models, pre-trained neural networks, hybrid techniques, and image improvement strategies like CLAHE. This framework achieved an accuracy of 99.01%, outperforming existing models and offering practical, robust solutions for disease detection.

Progress in machine learning and image analysis has opened new possibilities for automating disease recognition. However, current approaches encounter challenges such as limited accuracy, significant computational requirements, and difficulties in managing complex real-world datasets. These issues point to the need for more reliable, precise, and adaptable methods. The current study aims to overcome these issues by designing a novel hybrid framework that integrates advanced hybrid color-based segmentation methods, refined feature extraction processes, and a multi-class SVM classifier to accurately detect and classify over small dataset of Wheat Rusts and Septoria.

The primary intent of the proposed research is to create a hybrid approach for the identification and classification of wheat rust and septoria diseases, precisely. This approach leverages cutting-edge image processing techniques, enhanced feature extraction, and sophisticated machine learning algorithms to overcome the limitations of current detection and classification systems.

The proposed hybrid framework begins with a segmentation technique designed to isolate disease-affected areas in wheat leaf images. By integrating machine learning and advanced color manipulation techniques, the framework extracts key attributes such as shape, color, texture, and edge features. These attributes are crucial for identifying subtle variations in leaf coloration, structural characteristics, and patterns associated with various wheat diseases.

In the feature extraction phase, the framework systematically computes shape features (roundness, perimeter, and eccentricity), color features (Lab*, RGB, HIS, and HSV spaces), texture features (using Gray Level Co-occurrence Matrices at multiple angles), and edge features (density, contrast, sharpness, and entropy). This multi-modal feature extraction ensures a comprehensive representation of the diseased areas, significantly improving the classification process.

To improve performance, the collected features are integrated into a single feature vector and refined by applying dimensionality reduction methods, such as Principal Component Analysis. This step reduces redundancy, generates uncorrelated components, and retains the most significant information, enhancing computational efficiency without compromising accuracy. Additionally, mutual information is used to select the top 40 features with low skewness and high entropy, further ensuring that the classifier focuses on the most informative and discriminative features.

Finally, the refined feature vectors are passed to a classifier M-SVM for accurate categorization of the diseases. The M-SVM efficiently separates the data into distinct disease categories by leveraging the high-quality input features. The main aim is to build a solution that ensures strength, accuracy and computationally efficient system for detecting and classifying wheat diseases in real-world agricultural scenarios.

A novel modular strategy is put forth in this work to boost both precision and effectiveness in wheat disease identification and classification. By addressing the limitations of traditional methods, it integrates advanced segmentation, feature extraction, and classification techniques. The framework's scalable design allows adaptation to various agricultural contexts and facilitates the detection of other crop diseases, broadening its applicability in precision agriculture. Leveraging advanced computational methods, this approach supports timely decision-making and reliable crop health monitoring, empowering farmers to manage diseases effectively.

Within the scope of present knowledge, this study marks an early effort to integrate these techniques into a unified framework for the effective identification and classification of wheat diseases. The proposed approach seeks to address the shortcomings of current methods, which frequently depend on single-modal features or involve high computational costs. Specifically, the color-based segmentation facilitates the identification of key areas within the images, allowing the M-SVM classifier to concentrate on crucial regions, which significantly enhances accuracy and performance of classification.

This research makes two significant contributions. First, it introduces a novel hybrid framework that enhances the detection and classification of wheat rusts and Septoria. Second, the framework is thoroughly evaluated using a diverse wheat image dataset, demonstrating superior performance compared to conventional approaches and studies conducted over the past four years. Additionally, the framework achieves higher classification accuracy and ensures efficient processing, making it ideal for real-time agricultural applications.

The paper's layout is set up as follows: A review of recent studies is provided in Section 2 regarding the detection of wheat diseases and classification techniques. The creation of the hybrid framework, which combines segmentation and classification techniques, as described in Section 3, offering a comprehensive explanation of the segmentation strategies, feature selection processes, and classification techniques employed. Section 4 provides a description of the experiments, highlighting the dataset used, evaluation standards, and outcomes. Section 5 evaluates the performance of the framework by discussing both its strengths and limitations. while Section 6 summarizes the study and highlights possible avenues for future work.

2. Literature Review

In recent years, various techniques in machine learning and image analysis have been designed to identify and categorize various plant diseases. These methods incorporate digital

image processing, feature extraction approaches, and machine learning algorithms to improve the precision of disease diagnosis.

An improved K-means algorithm for tomato leaf image segmentation is proposed by [51] to solve the shortcomings of the traditional method's need for a pre-defined cluster number by automatically determining the optimal clusters. The suggested algorithm, evaluated against DBSCAN, Mean Shift, and ExG-ExR, achieved the highest F1 score (0.982) and lowest entropy (0.118) on 20 images, while ExG-ExR outperformed the others by scoring F1 score and entropy of 0.765 and 0.643, respectively, on diseased leaf images.

[43] tackles wheat yellow rust disease in Pakistan, proposing a machine learning framework for classifying healthy, resistant, and susceptible infection types using mobile images. Six GLCM and four LBP texture features were extracted, and five models (DT, RF, LightGBM, XGBoost, and CatBoost) were trained. CatBoost scored the most excellent accuracy of 92.30% with GLCM features. The study suggests improvements through dataset expansion and deep learning models, offering a useful instrument for the early yellow rust detection and timely intervention to reduce yield losses. [47] presents a method for wheat disease classification in which reducing false negatives (misclassifying infected plants as healthy) and false positives (misclassifying healthy plants as diseased) is essential for improving model reliability. [55] performed an in-depth analysis of digital image processing methods for detecting leaf diseases, exploring techniques like color segmentation and texture analysis to identify conditions such as, powdery mildew and leaf spot. Their study highlighted the potential benefits of incorporating ML models to enhance the effectiveness of these techniques in Crop-related applications. Building on this, [13] highlighted the role of ML models in Crop-related disease detection, particularly their effectiveness in pattern recognition for plant disease classification. Their research focused on predicting strawberry fruit quality attributes, showcasing how color-based features and machine learning could advance plant health monitoring. [51] given an innovative method integrating transformer models with blended learning for wheat leaf disease identification. Their study addressed challenges in distributed datasets and achieved outstanding results with up to 99% accuracy, Showcasing the capabilities of deep learning models for resource-limited settings. Similarly, [47] investigated the application of UAV multispectral imagery paired with spectral and textural analysis for monitoring wheat *Fusarium* head blight, further showing how advanced imaging techniques can enhance disease detection.

The author [11] utilized the WFD2020 dataset to identify five wheat fungal diseases, achieving an accuracy of 94.2% with the EfficientNet CNN. The addition of data augmentation and style transfer techniques, along with a real-time mobile application, showcased the practical implementation of disease detection for early wheat disease management. Likewise, [15] applied a combined textural and deep feature approach for yellow rust severity assessment, achieving 92.4% accuracy with their CNN-SVM model, underscoring the benefits of combining traditional texture analysis with deep learning.

The research conducted by [16] made significant progress in understanding wheat diseases detection through a CNN model trained on a validated Kaggle dataset, achieving 98.43% accuracy with various performance metrics. This approach emphasized image augmentation and CIELAB color segmentation for more robust and efficient disease detection in wheat. In a similar vein, [60] utilized Lab* color space and SVM classification for wheat leaf disease detection, achieving reliable results in distinguishing diseased and healthy leaves.

[17] showcased a hybrid approach to early detection of breast cancer that integrated advanced image segmentation with machine learning. Their findings motivate the exploration of similar hybrid segmentation approaches for plant diseases, particularly those involving colorful images with diverse leaf disease patterns. Further building on hybrid methods, [54] introduced a fully convolutional neural network (FCNN) for automatic background subtraction in leaf images, demonstrating how efficient segmentation techniques can enhance disease detection in complex field conditions.

[33] developed a CBIR system for soybean foliage diseases by combining shape, color, and texture features. Their approach achieved high retrieval efficiencies for various

diseases, highlighting the importance of database size and image quality in improving disease detection performance.

[49] explore image processing for classifying olive leaf diseases, focusing on Neofabrea and Peacock leaf spots. They found k-means segmentation more accurate than histogram thresholding. Texture analysis revealed correlations between disease infection and features like energy, homogeneity, and entropy. This study aids in olive disease detection, crucial for olive oil production amid climate change. It underscores the importance of texture-based features, which can also be applied to wheat disease detection to differentiate between various infections like rusts and septoria.

Table 1. Comparative Analysis of Recent Approaches in Crop Disease Detection and Segmentation, Highlighting Goals and Future Perspectives.

Author	Year	Goals	Future Perspective
V. Singh and Misra	2017	Employs color-based image segmentation to improve the detection of plant diseases and quantify their severity.	Combine segmentation approaches with machine learning models to achieve more reliable disease evaluation.
Tian et al.	2017	Proposes an enhanced PCA-driven multichannel framework for segmenting wheat leaf lesions.	Explore the deployment of realtime PMCV for continuous plant-disease surveillance.
Sharif et al.	2018	Presents refined segmentation methods to isolate leaf lesions, supporting precise severity analysis.	Advance weighted segmentation strategies to achieve more accurate lesion delineation.
S. Zhang et al.	2019	Develops a hybrid clustering approach for fast and accurate identification of leaf lesions.	Enhancing segmentation accuracy by integrating deep learning with hybrid clustering techniques.
K. Tian et al.	2019	Utilizes an enhanced K-means algorithm to achieve precise segmentation of tomato leaf lesions.	PMCV integration to enable automated monitoring of plant diseases.
Goncalves et al.	2021	Utilizes deep learning techniques to segment leaf lesions, contributing to improved severity estimation.	Optimize deep learning frameworks for handling diverse lesion categories.
Mukhopadhyay et al.	2021	Adopts dynamic programming-based segmentation to accurately delineate citrus disease lesions.	Combine dynamic programming and automated phenotyping for real-time assessment of plant diseases.
Gong et al.	2022	Proposes a lightweight segmentation model for powdery mildew, enabling efficient estimation of diseased regions.	Extend it of field-based disease management.
Balafas et al.	2023	Promotes precision agriculture through the integration of machine learning to strengthen disease detection and classification.	Investigate multimodal data fusion for more resilient detection across varied crops.
K Badri et al.	2024	Monitors canker disease and develop an advanced computer vision technique to reduce agricultural losses.	Improve model reliability, scalability, and suitability for real-world farming conditions.

[53] includes preprocessing steps such as resizing images to 256×256 pixels and applying histogram equalization, after which feature extraction is performed. The method uses K-means clustering for segmentation, contour tracing, and textural analysis with

DWT, PCA, and GLCM. Tested on a village dataset (685 samples), the model achieved 99.% accuracy, performing exceptionally well in identifying conditions such as Mosaic virus and leaf mold. Evaluation metrics included F1 score, precision and recall validated further with subsamples under specified system requirements.

[55] used image processing and ML to identify wheat stripe and leaf rust. Images were preprocessed and 140 features extracted, with ReliefF, 1R, correlation-based selection, and PCA refining features. SVM, RF and BPNN models trained on multi-variety datasets achieved 82.05–98.87% accuracy, significantly outperforming single-variety models. Their findings underscore the importance of multi-variety datasets for robust disease detection, directly informing our work on wheat disease classification using color-based segmentation and machine learning.

[58] presents a hybrid PSO-K-means algorithm for segmenting agricultural target regions from images in the YCbCr color space, improving performance under low illumination and shadow conditions. By combining PSO's global search with K-means' speed, the method achieves better stability and accuracy than traditional PSO. It effectively segments products in complex environments and addresses RGB limitations in super-green images using gray-scale processing. This approach can enhance segmentation and improve disease classification accuracy.

In [59] a Global-Pooling-Dilated-Convolutional-Neural-Network was proposed to identify cucumber foliage lesion by integrating global pooling methods with dilated convolutions. The model offers advantages such as larger receptive fields, enhanced spatial resolution, and reduced parameters. Tested on a cucumber leaf disease dataset, GPDCNN outperformed AlexNet and traditional methods like GLSVD and IPT, achieving higher accuracy and faster training with minimal preprocessing. This approach can inform wheat leaf disease detection by capturing complex patterns in rusts and septoria.

The introduced classification framework is evaluated against several well-known methods, such as those reported by [11], [43], [16] and [51] focusing on four specific Wheat diseases: wheat rust of leaf, stripe and stem, and wheat septoria.

The studies summarized in Table 1 highlight significant advancements in machine learning and image processing approaches to plant lesion identification and classification. Approaches like hybrid clustering [59], enhanced K-means clustering [51], and machine learning models like SVMs and CNNs [53] have demonstrated effectiveness in detecting various plant diseases, including rust and leaf mold. Advanced methods incorporating texture analysis [15], deep learning architectures [55], and lightweight models [47] have significantly improved segmentation performance and computational efficiency. The importance of multi-variety datasets, optimized preprocessing, and robust feature extraction techniques is underscored in these studies, facilitating accurate disease detection and severity quantification.

In particular, these methods have demonstrated their potential for scalable, real-time agricultural monitoring systems by being effectively applied to a variety of plant diseases. Future studies should concentrate on improving machine learning models, integrating multimodal data, and guaranteeing real-time Relevance in various agricultural circumstances. Building on these advancements, this study proposes a novel hybrid framework that integrates Adaptive K-Mean, Gaussian Mixture Model and advanced-color-manipulation techniques for precise segmentation, after segmenting the lesions, key features are extracted effectively. The process then involves feature selection and the use of M-SVM techniques to ensure accurate detection and classification. This approach offers

reliable results compared to deep learning methods, over a small dataset of wheat rusts and septoria.

3. Proposed Method

The proposed methodology involves four main stages: **(i)** Data acquisition and image processing, **(ii)** Lesion segmentation to detection of infected regions, **(iii)** Feature extraction and code-book construction, and **(iv)** Feature selection and classification. Each of these stages is further broken down into a series of steps, as illustrated in Fig. 1. Below, this paper provides a detailed explanation of each stage.

In the initial stage, image datasets are collected from Wheat Leaf dataset and CIMMYT Dataset. Subsequently, image processing techniques are employed to convert the RGB images of the wheat leaves into the Lab* color space. This conversion is necessary for the extraordinary result of the proposed hybrid segmentation method, which is designed to accurately segment the lesion region. This step is crucial because the proposed segmentation algorithm operates more effectively in the Lab* color space, ensuring that the subsequent processes yield accurate results. In second stage involves lesion segmentation, where the infected regions of the wheat leaves are precisely identified. This process isolates the areas affected by diseases, allowing for focused analysis. Next, in the feature extraction and codebook construction stage, various features from the segmented lesions are extracted. These features include shape, color, texture, and edge characteristics in the form of a vector matrix, which are then compiled into a codebook. The codebook provides a detailed representation of distinctive features linked to wheat diseases, like leaf rust, stripe rust, septoria, and stem rust. The key features are taken from the combined codebook during the feature selection and classification stage. PCA is then applied to minimize the dimensionality of the feature set while preserving the majority of the data's variance. This process changes the initial features into a number of independent components. Subsequently, classification algorithms are utilized to analyze these features, enabling precise identification and differentiation of various wheat leaf diseases.

Each of these stages is meticulously structured to ensure a systematic and efficient approach to wheat leaf disease classification. The following sections provide an in-depth look at the processes involved in each stage.

3.1. Data acquisition

The proposed methodology's first step is data acquisition, which entails gathering crisp photos of wheat leaves for further examination. This stage is crucial because the quality and representativeness of the data gathered determine the process's overall accuracy and dependability.

Healthy and diseased wheat leaves are collected from Wheat Leaf dataset and CIMMYT Dataset. This ensures a diverse and comprehensive dataset. The samples represent different stages of disease progression and various environmental conditions.

3.2. Image Processing

Numerous color space models exist, and the optimal choice depends on the specific problem. A thorough analysis of multiple color spaces was carried out to enhance the efficacy of the suggested segmentation technique. The color spaces considered include CMYK, RGB, YCbCr, XYZ, HSV/HSL, YUV, and Lab* [12]. A detailed comparison of these color spaces is given in Table 2, ranking them from least favorable to most favorable, thereby establishing Lab* as the most suitable choice due to its superior ability to separate color information for the proposed approach presented in this paper.

The Lab* color model, mathematically defined by Eqs. (1), (2), (3), and (4), demonstrates its suitability through the results presented in the "Final Segmented Output" (Fig. 1). This model comprises three components: lightness (L), representing the intensity, the green-red chromatic axis (a), and the blue-yellow chromatic axis (b). Although Lab* offers robust color separation capabilities, the results of its segmentation can be affected by the inherent color variability of wheat leaves and changes in lighting conditions. These challenges, which may lead to uneven background representation in the processed images, are effectively addressed by Algorithm-3, specifically designed to mitigate such issues and enhance segmentation accuracy.

To facilitate the implementation of the Lab* color model, OpenCV provides a straightforward method for converting RGB images to the Lab* color space. This transformation involves a sophisticated mathematical framework, including intermediate conversions from RGB to XYZ and then from XYZ to Lab*. OpenCV simplifies this complex process by offering pre-defined mathematical expressions for each step, ensuring efficient and accurate transformation for image segmentation tasks.

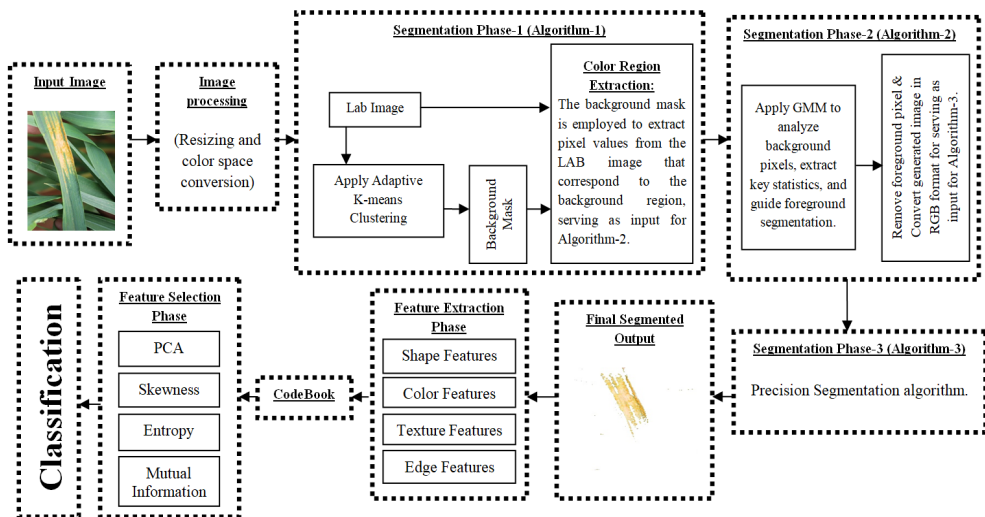


Figure 1. A Framework for the Detection and Classification of Wheat Rust and Septoria.

(1) Conversion from RGB to XYZ

Normalize the RGB values: Normalize the red values, green values, and blue values of the input RGB colour to the range [0, 1] by dividing each channel value by 255 as given [59] in Eq. (1).

$$\left. \begin{aligned} R' &= \frac{R}{255} \\ G' &= \frac{G}{255} \\ B' &= \frac{B}{255} \end{aligned} \right\} \quad (1)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M * \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (2)$$

Where, M is a 3×3 matrix:

$$M = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix}$$

To calculate the XYZ colour space, substitute the value of M into Eq. (2).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} * \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

Table 2. Comparing Color Spaces Ranked from Least to Most Favorable [12], [60].

Colour Space	Applications	Perceptual Uniformity	Separation of Components	Device-Dependent?
CMYK	Printing, Inks	No	Yes	Yes
RGB	Displays, Graphics	No	No	Yes
YCbCr	Video, Compression	Yes	Yes	Yes
XYZ	Colour Matching	Yes	No	No
HSV/HSL	Colour Selection	No	Yes	No
YUV	Video Encoding	Yes	Yes	Yes
Lab*	Colour Analysis	Yes	Yes	No

(2) Conversion from XYZ to Lab*

The XYZ matrix values define how the Lab* color space relates to the components [57] of lightness (L^*), green_red opponent channel (a^*) and blue_yellow opponent channel (b^*) in Eq (3).

$$\left. \begin{aligned} L^* &= 116 * f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500 * \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ b^* &= 200 * \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \end{aligned} \right\} \quad (3)$$

Where $f(t)$ is a nonlinear function that depends on t and is defined in the following manner [54]:

$$f(t) = \begin{cases} t^{\left(\frac{1}{3}\right)} & \text{if } t > 0.008856 \\ 16 * t & \text{if } t \leq 0.008856 \end{cases} \quad (4)$$

3.3. Lesion segmentation

The automatic identification of lesion region in infected images depends on image segmentation, which requires advanced techniques to ensure high precision. To address this, a color-based hybrid segmentation approach is proposed, integrating Adaptive K-Means Clustering optimized using the Silhouette Index, Gaussian Mixture Modeling (GMM) for probabilistic segmentation, and a Precision Segmentation Algorithm for enhanced pixel-level color manipulation and accuracy.

The lesion segmentation process initiates with the generation of a background mask utilizing adaptive K-means clustering within the Lab* color space, facilitating accurate separation of background and foreground regions. Subsequently, distinct color regions are extracted from the Lab* color space of the input image, with background pixels being labeled using the generated mask. The third step creates a refined foreground mask by carefully designing and training a GMM with the background pixel data. This allows the GMM to predict labels for every pixel in the reshaped Lab* image. In the fourth step, an inventive method is used to isolate the lesion areas by removing the foreground pixels from the RGB input image and giving them a white color. In order to achieve the definitive segmented leaf lesion, the segmentation is finally refined by changing pixels where the green component exceeds both the red and blue components to white.

3.3.1. Algorithm-1 (Adaptive K-Mean Clustering)

The optimal count of clusters is dynamically computed by Adaptive K-Means Clustering for image segmentation, making it useful in applications like object recognition, medical imaging and in agriculture. Unlike traditional K-Means, which requires a fixed number of clusters, the proposed Adaptive K-Means approach uses the silhouette score to automatically determine the optimal number of clusters based on data characteristics. This approach helps create effective background masks in image processing [46], [51].

Clustering divides data into subsets, and while some applications allow manual selection of the cluster number 'k,' it's often necessary to determine 'k' from the data to avoid over- or under-clustering[46]. This study evaluates various validity indices (DB, CH, and Silhouette) on a dataset of 50 complex background images for computing the optimal count of clusters. The Silhouette index, despite a slightly longer runtime, provides the highest accuracy and is used to determine the most suitable cluster count.[51]. The Silhouette indices are commonly represented by Eq. (5) as follows:

$$S(i) = \frac{\text{Min}(b_{i,k}) - (a_i)}{\text{Max}\{(a_i), \text{Min}(b_{i,k})\}} \quad (5)$$

In the Silhouette index, which ranges from -1 to +1, " a_i " denotes the average distance of the i th instance from the remaining instances of its own cluster, while " $b_{(i,k)}$ " refers to

the separation between the i th sample and the center point of the k -th cluster. " $\text{Min}(b_{i,k})$ " is the shortest distance across all other clusters [51].

This study employs Silhouette indices, as depicted in Fig. 2, to ascertain the optimal count of clusters while effectively preserving the leaf's target area. The ideal cluster count varies across different images, a capability enabled by the proposed segmentation technique. By integrating Adaptive K-Means into the hybrid segmentation approach, the method achieves improved results, particularly for complex backgrounds, by accurately determining the cluster values. The process of generating a background mask using the Adaptive K-Means Clustering algorithm is outlined in Algorithm-1.

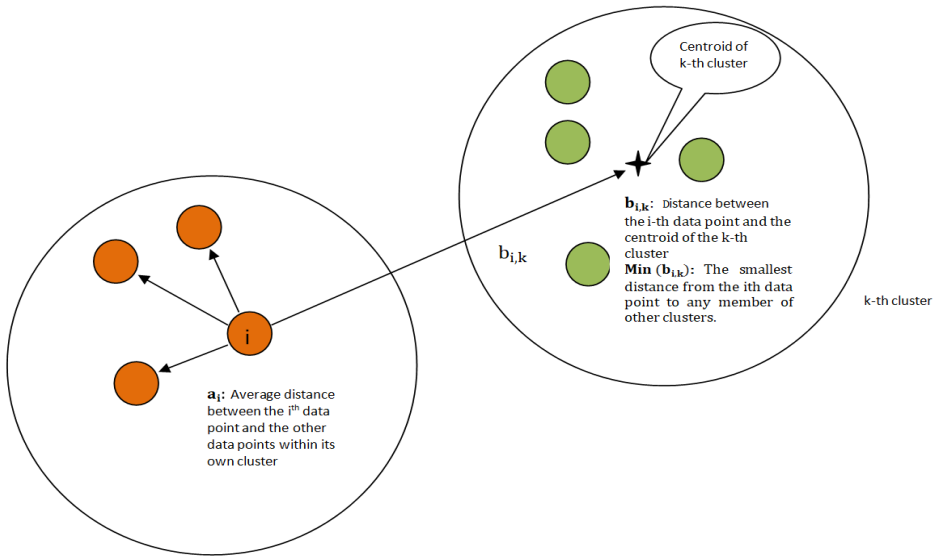


Figure 2. Representation of Silhouette Width.

ALGORITHM-1: ADAPTIVE K-MEANS CLUSTERING FOR BACKGROUND MASK

Input:

Image: Input colour image

n_clusters: Number of clusters for k-Means = 2

Output:

Background_mask: Binary mask indicating the background region

Start:

Step-1. Convert the input image to Lab* colour space.

Step-2. Convert the Lab* image into a 2D pixel array.

Step-3. for $n_clusters = 2$ to $max_clusters$

Step-3.1. Perform K-means clustering on the reshaped pixel data

Step-3.1.1. Create an instance of the clustering

```

| algorithm with 'n_clusters' as the number of
| clusters.
| Step-3.1.2. Fit the algorithm to the data and obtain
| cluster labels.
| Step-3.2. Calculate the silhouette score for the n_clusters
| labels
| Step-3.3. If the calculated silhouette score is higher than
| 'best_score'
|   Step-3.3.1. Update 'best_score' with the new score
|   Step-3.3.2. Update 'optimized_n_clusters' with the
|   current 'n_clusters'.
| End loop
| Step-4. Identify the background cluster by finding the cluster
| with the highest mean L value.
| Step-5. Create a binary mask for the background pixels.
| Return: Background_mask
| End

```

3.3.2. Algorithm-2 (Gaussian Mixture Model)

Gaussian Mixture Models (GMMs) play a crucial role in leaf lesion segmentation, which is a fundamental process in agricultural research and plant pathology. These models are utilized to capture the probability distribution of pixel intensities or colors in leaf images, particularly those affected by lesions or diseases. Mathematically, a GMM for leaf lesion segmentation can be represented as shown in Eq. (6) [17].

$$P(x) = \sum_{i=1}^N \pi_i N(x | \mu_i, \Sigma_i) \quad (6)$$

These models capture the probability distribution of pixel intensities or colors in leaf images, especially those affected by lesions or diseases. Each Gaussian component represents a unique region within the leaf, allowing the model to probabilistically assign pixels to these segments [17]. This capability is crucial for distinguishing between healthy leaf regions and those affected by lesions, contributing significantly to precise leaf lesion segmentation [54].

Algorithm-2 employs GMMs for foreground mask generation, serving as a technique for image foreground segmentation. We have customized all steps except for step 3 in this algorithm. This algorithm takes an input color image and a binary background mask generated by the Adaptive K-Means Clustering algorithm. At the next level, GMMs utilize the background pixels, resulting in the acquisition of crucial parameters such as means, covariance, and mixing coefficients for each Gaussian component. Algorithm-2 predicts labels for all pixels, facilitating the creation of a foreground mask based on these labels. The final step involves identifying the maximum-mean channel among the Gaussian components, reshaping the labels mask, and ultimately returning the foreground mask, effectively segregating the foreground from the background within the image.

ALGORITHM-2: GAUSSIAN MIXTURE MODELS FOR FOREGROUND MASK**Input:***Image:* Input colour image*Background_Mask:* Binary mask indicating the background region*N_Components:* Number of components for gaussian mixture models ($n_components \leftarrow 3$)**Output:***Foreground_Mask:* Binary mask indicating the foreground region (non-lesion pixel of input image).**Start:****Step-1.** Convert the input image to Lab* colour space.**Step-2.** Background pixels $\leftarrow$ Extract the pixel values of the background region from the Lab* image using the background mask.**Step-3.** Perform Gaussian Mixture Models on the extracted background pixels with the specified number of components.**Step-3.1.** Fit the Gaussian Mixture Models to the background pixels.**Step-3.2.** The GMM now holds learned parameters for each Gaussian component:

- **Mean:** The center of each component's distribution.
- **Covariance:** Describes the spread and correlation of the data.
- **Mixing coefficients:** Weights assigned to each component in the mixture.

Step-4. Predict the labels of all the pixels in the Lab* image using the trained GMM model.**Step-5.** Create a mask for the foreground pixels based on the labels predicted by GMM.**Step-6.** Determine the maximum-mean channel among the GMM components and compare its index with the label for each pixel.**Step-7.** Generate the foreground mask by reshaping the labels mask to the dimensions of the input image.**Return:** Foreground_mask**End**

Algorithm-2, Gaussian Mixture Models for Foreground Mask, separates foreground regions with precision. It transforms images to the Lab* color space, extracts background pixels, and employs GMM to predict labels for precise foreground segmentation.

3.3.3. Algorithm-3 (Precision Segmentation Algorithm)

Algorithm-3 has been customized to achieve accurate extraction of the leaf lesion region, playing a crucial role in enhancing image segmentation by refining the foreground and

ensuring color consistency. This algorithm takes an RGB image and a refined foreground mask as inputs, producing a finely-segmented image referred to as 'Output_img'.

ALGORITHM-3: PRECISION SEGMENTATION ALGORITHM

Input:

Image: Input RGB image

Foreground_Mask: Refined foreground mask

Output:

Output_img: Final output of Segmented image

Start:

Step-1. Initialize an empty image, **output_img**, with the same dimensions as the input image.

Step-2. Iterate over each pixel (x, y) in the input-image:

Step-2.1. Get the colour of the pixel (x,y) from input-image.

Step-2.2. Check if the corresponding pixel (x, y) in the foreground mask is labeled as foreground:

Step-2.2.1. If foreground mask for pixel[x, y] is true (indicating it's part of the foreground):

- Assign the colour white (255, 255, 255) to the pixel in output_img: output_img[x, y] = (255, 255, 255).

Step-2.2.2. Else (if it's not part of the foreground) Check the colour components:

- Get the R, G & B elements of pixel_colour.

Step-2.2.2.1. If $G > R$ and $G > B$ (green component is the largest):

- Assign the colour white (255, 255, 255) to the pixel in output_img: output_img[x, y] = (255, 255, 255).

Step-2.2.2.2. Else (green component not the largest):

- Assign pixel_colour to the pixel in output_img: output_img[x, y] = pixel_colour.

Return: Return output_img as the final segmented output image.

End

The process starts by creating a blank image that matches the dimensions of the input image. It then iterates through each pixel, determining its association with the foreground. If a pixel is part of the foreground, it is assigned the color white, resulting in clear separation. For non-foreground pixels, a meticulous check of color components is conducted. If green is the dominant color, the pixel is turned white to maintain image integrity. Otherwise, the original pixel color is retained. The algorithm ultimately returns 'Output_img' as the final segmented output image, representing an advanced approach to image segmentation.

3.3.4. Analysis and Comparison of Segmented Output

This section evaluates the proposed segmentation technique using 160 images of four wheat diseases affected by: septoria, leaf rust, stem rust and stripe rust. The images were sourced from the Wheat Leaf dataset and CIMMYT Dataset. The proposed approach employs the Adaptive K-Means clustering technique, which is capable of dynamically selecting the appropriate number of clusters for accurate segmentation. Following this, the subsequent phases of segmentation algorithm is implemented to produce the final segmented result, showcasing the effectiveness of the hybrid segmentation strategy.

Figure 3 illustrates how well the proposed segmentation approach works and underlines the role of Adaptive K-Means clustering in addressing wheat diseases—including septoria, leaf rust, stem rust, and stripe rust—that vary significantly in texture, shape, and color characteristics. The figure highlights that an inappropriate choice of cluster count can greatly affect the segmentation outcomes. Specifically, too few clusters may cause non-lesion regions to overlap with diseased areas, while too many clusters might result in the loss of critical regions. This highlights the necessity of determining the appropriate number of clusters for achieving precise segmentation.

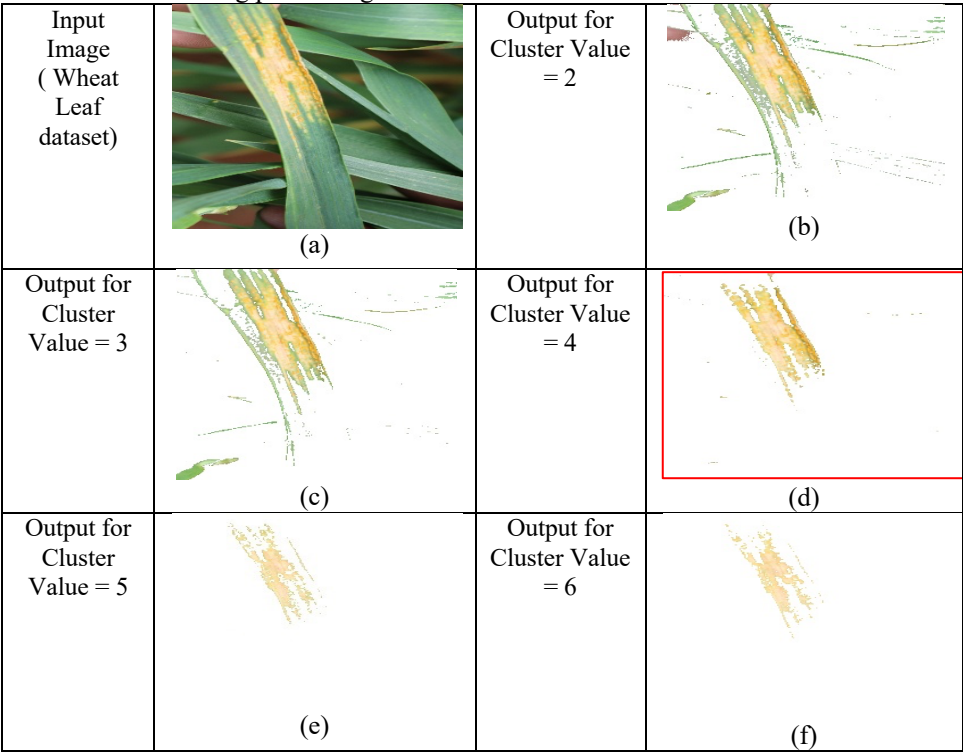


Figure 3. Segmentation outcomes using the proposed hybrid technique with K-Means clustering instead of Adaptive K-Means clustering, showcasing segmented outputs with different cluster numbers: (a) Input Image, (b) 2-Clusters, (c) 3-Clusters, (d) 4-Clusters, (e) 5-Clusters, and (f) 6-Clusters.

To address this, the proposed segmentation technique leverages Adaptive K-Means clustering to enhance segmentation precision. Fig. 3(d), highlighted in red, demonstrate accurate segmentations of leaf lesion target regions against complex backgrounds, achieved using the proposed method.

This comprehensive strategy consistently outperformed existing baseline methods (i.e., Texture Analysis with Edge Detection, Edge Detection with GMM, Edge Detection with FCN, and K-Means with Mask R-CNN) which were re-trained and fine-tuned using the same dataset to evaluate the proposed segmentation approach to ensure fair and optimal performance comparison.

Fig 4 illustrates the qualitative segmentation results for wheat stripe rust, while Table 3 provides a quantitative assessment for wheat leaf rust, stripe rust, septoria, and stem rust, comparing the performance of these baseline methods with the proposed technique. Notably, each baseline method integrates a foliage extraction approach with a corresponding lesion region extraction technique, highlighting the improvements achieved by the proposed strategy.

The proposed method, designed to accurately delineate diseased regions in wheat, effectively segments the leaf lesion area without leaving off any lesion pixels, as shown in Fig. 4(f). In contrast, Fig. 4(b), 4(c), and 4(d) display the inclusion of non-lesion pixels within the lesion region, leading to less precise segmentation. The K-Means + Mask R-CNN

method marginally truncates small lesion areas, as shown in Fig. 4(e). Nonetheless, the lesion region is extracted precisely and completely using the suggested technique in Fig. 4(f). These outcomes demonstrate the higher accuracy and efficiency of the suggested segmentation technique while maintaining reasonable latency, underscoring its practical utility in precision agriculture and crop management.

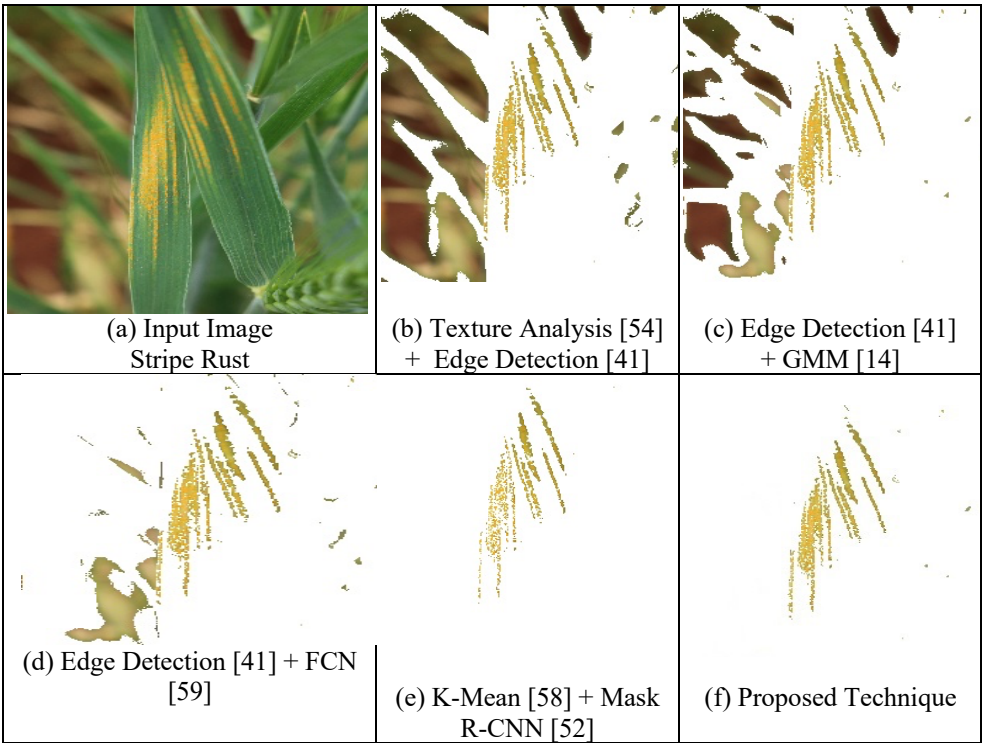


Figure 4. Results of wheat stripe rust segmentation using different baseline techniques and the proposed method for isolating the infected regions.

Fig. 5 provides a detailed evaluation of the proposed segmentation method across four types of wheat diseases septoria, stem rust, leaf rust and stripe rust using 40 images per disease (160 images in total) from the Wheat Leaf dataset and CIMMYT Dataset. Quantitative statistics for these 160 images are presented in Table 4. Fig. 5 illustrates the segmentation results for each disease alongside ground truth images, emphasizing the method's effectiveness. The proposed method attains a maximum segmentation performance index of 98.21% and an average segmentation performance index of 96.98%.

Additionally, similarity metrics are employed to evaluate the segmentation algorithm by comparing its outputs with ground truth images. Performance indicators, such as the Dice Coefficient, F1 Score, and Jaccard Index [41], further validate the effectiveness of the proposed method, as discussed in earlier sections.

Table 3. Comparison of segmentation method effectiveness based on average key metrics for 160 images.

Method	Average Dice Coefficient	Average F1-Score	Average Jaccard Index (IoU)
Texture Analysis + Edge Detection	0.711	0.6837	0.6932
Edge Detection + GMM	0.7317	0.7456	0.7211
Edge Detection + FCN	0.9216	0.8962	0.8626
K-Mean + Mask R-CNN	0.9467	0.939	0.8858
Proposed Method (Adaptive K-means + GMM + Precision Segmentation Algorithm)	0.9814	0.9619	0.9661

Dice Accuracy:

Dice Accuracy, sometimes referred to as the Dice coefficient, quantifies how well anticipated and ground truth segmentations in picture tasks match. It is computed as follows:

$$\text{Dice_Accuracy} = \frac{2 * |A \cap B|}{|A| + |B|} \quad (7)$$

Where:

- A and B are the pixel sets of ground truth and predicted masks.
- $|A \cap B|$ gives the common pixels, and $|A| + |B|$ shows the total pixels in both sets.
- Dice score varies between 0 (no overlap) and 1 (perfect match).

F1-Score:

The F1-Score serves as a balancing factor between precision and recall, where precision quantifies correctly segmented pixels out of all segmented pixels, and recall measures accurately segmented pixels relative to the total ground truth pixels [47], [54].

$$F1 = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (8)$$

Jaccard Index (IoU - Intersection over Union):

The Jaccard Index shows how closely the segmented regions match the ground truth by comparing the intersection of the two with their union [44].

$$IoU = \frac{\text{Intersection area}}{\text{Union area}} \quad (9)$$

Latency:

The paper assesses runtime efficiency by measuring inference time per sample image, using hardware including a Intel Core i5-12500H, 12th Gen processor, 16 GB RAM at 3.10 GHz, and Nvidia GeForce RTX 2060 GPU (4GB).

Table 4. Lesion segmentation performance assessment as compared to ground truth values

Image Index	Dice Coefficient	F1 Score	IoU (Jaccard Index)	Image Index	Dice Coefficient	F1 Score	IoU (Jaccard Index)
Img.1	0.98255	0.95453	0.92453	Img.81	0.98173	0.95955	0.94955
Img.2	0.97412	0.98722	0.97531	Img.82	0.98816	0.98021	0.97021
Img.3	0.93045	0.94889	0.93310	Img.83	0.98790	0.97586	0.96586
Img.4	0.98230	0.97988	0.98230	Img.84	0.98168	0.95584	0.94584
Img.5	0.98988	0.96264	0.97988	Img.85	0.98608	0.96849	0.97608
Img.6	0.97943	0.96143	0.95143	Img.86	0.98016	0.96176	0.98016
Img.7	0.98070	0.95915	0.97070	Img.87	0.98440	0.98299	0.97299
Img.8	0.98365	0.96969	0.95969	Img.88	0.98225	0.96940	0.95940
Img.9	0.98855	0.97017	0.96017	Img.89	0.98811	0.96049	0.95049
Img.10	0.98613	0.97531	0.96531	Img.90	0.98579	0.98174	0.97174
Img.11	0.98136	0.97264	0.96264	Img.91	0.98180	0.96717	0.95717
Img.12	0.98574	0.96774	0.97574	Img.92	0.98342	0.94850	0.97342
Img.13	0.97881	0.95915	0.96881	Img.93	0.98013	0.97328	0.96328
Img.14	0.98484	0.97242	0.96242	Img.94	0.98097	0.97174	0.97097
Img.15	0.97564	0.95378	0.96564	Img.95	0.97476	0.95706	0.96476
Img.16	0.98548	0.97454	0.96454	Img.96	0.97827	0.97089	0.96827
Img.17	0.98614	0.96166	0.97614	Img.97	0.97131	0.96862	0.96131
Img.18	0.98245	0.96915	0.95915	Img.98	0.98906	0.96476	0.97906
Img.19	0.98100	0.97433	0.96433	Img.99	0.98996	0.97247	0.96247
Img.20	0.98660	0.96266	0.95266	Img.100	0.97522	0.96522	0.96522
Img.21	0.98197	0.94502	0.96506	Img.101	0.98003	0.97067	0.98003
Img.22	0.98878	0.97774	0.96774	Img.102	0.95772	0.97533	0.94772
Img.23	0.97984	0.97350	0.96984	Img.103	0.98506	0.95850	0.94850
Img.24	0.97538	0.96378	0.95378	Img.104	0.98089	0.96411	0.97089
Img.25	0.98230	0.97350	0.98230	Img.105	0.98278	0.95947	0.94947
Img.26	0.96700	0.95666	0.94666	Img.106	0.98064	0.97083	0.98064
Img.27	0.98948	0.97166	0.96166	Img.107	0.98997	0.96706	0.95706
Img.28	0.98614	0.96933	0.95933	Img.108	0.97982	0.96373	0.95373
Img.29	0.98653	0.95772	0.97653	Img.109	0.98043	0.97968	0.98043
Img.30	0.98589	0.95818	0.97589	Img.110	0.98210	0.98127	0.98300
Img.31	0.97275	0.94414	0.96275	Img.111	0.98988	0.96517	0.97988
Img.32	0.98734	0.97498	0.96498	Img.112	0.98026	0.95147	0.97026
Img.33	0.98710	0.96232	0.95232	Img.113	0.96259	0.97686	0.95259
Img.34	0.96748	0.95139	0.94139	Img.114	0.98030	0.96002	0.98030
Img.35	0.98059	0.96772	0.95772	Img.115	0.96852	0.96208	0.95852
Img.36	0.98985	0.97276	0.96276	Img.116	0.98174	0.97922	0.96922
Img.37	0.98350	0.96275	0.97350	Img.117	0.98067	0.96415	0.97067
Img.38	0.98322	0.96276	0.97322	Img.118	0.98522	0.96965	0.95965
Img.39	0.97853	0.96733	0.95733	Img.119	0.97734	0.94059	0.96734
Img.40	0.98958	0.96818	0.95818	Img.120	0.98042	0.96204	0.97042
Img.41	0.97076	0.96071	0.95071	Img.121	0.98681	0.97496	0.96496
Img.42	0.96528	0.93138	0.95129	Img.122	0.97135	0.96021	0.95021
Img.43	0.97833	0.95414	0.94414	Img.123	0.98786	0.97418	0.96418
Img.44	0.98071	0.97653	0.98071	Img.124	0.98446	0.95417	0.97446
Img.45	0.98527	0.96610	0.95610	Img.125	0.98182	0.96425	0.95425
Img.46	0.99169	0.96682	0.95682	Img.126	0.98683	0.96970	0.95970
Img.47	0.98050	0.95867	0.94867	Img.127	0.98436	0.96685	0.95685
Img.48	0.97897	0.95733	0.96897	Img.128	0.98969	0.96942	0.97969
Img.49	0.98298	0.95506	0.94506	Img.129	0.98954	0.97792	0.96792
Img.50	0.98558	0.96907	0.95907	Img.130	0.98951	0.97527	0.96527
Img.51	0.98135	0.96225	0.95225	Img.131	0.98117	0.96350	0.95350

Img.52	0.98653	0.97507	0.96507	Img.132	0.97659	0.95433	0.96659
Img.53	0.98864	0.96609	0.95609	Img.133	0.98524	0.96446	0.95446
Img.54	0.95939	0.95939	0.94939	Img.134	0.98699	0.97152	0.96152
Img.55	0.97092	0.98386	0.96092	Img.135	0.98065	0.96484	0.95484
Img.56	0.98016	0.97548	0.96548	Img.136	0.97796	0.94134	0.96796
Img.57	0.98373	0.97187	0.96187	Img.137	0.98698	0.97697	0.96697
Img.58	0.98946	0.97422	0.96422	Img.138	0.97903	0.96912	0.95912
Img.59	0.98689	0.96486	0.95486	Img.139	0.98852	0.97516	0.96516
Img.60	0.98648	0.96757	0.95757	Img.140	0.98015	0.96027	0.95027
Img.61	0.98055	0.96448	0.95448	Img.141	0.96624	0.95629	0.94629
Img.62	0.98043	0.96760	0.95760	Img.142	0.98759	0.97559	0.96559
Img.63	0.98143	0.96307	0.95307	Img.143	0.98998	0.97287	0.96287
Img.64	0.96697	0.95394	0.94394	Img.144	0.97755	0.96559	0.96755
Img.65	0.98273	0.97297	0.96297	Img.145	0.98230	0.97436	0.98230
Img.66	0.98109	0.96718	0.95718	Img.146	0.98064	0.96796	0.97064
Img.67	0.98121	0.95490	0.94490	Img.147	0.97716	0.96207	0.95207
Img.68	0.98818	0.97583	0.96583	Img.148	0.98858	0.98237	0.97237
Img.69	0.98334	0.96213	0.95213	Img.149	0.97416	0.95804	0.94804
Img.70	0.98274	0.97172	0.96172	Img.150	0.98448	0.96658	0.95658
Img.71	0.98986	0.97849	0.96849	Img.151	0.98406	0.97052	0.96052
Img.72	0.98041	0.97249	0.96249	Img.152	0.98833	0.96659	0.97833
Img.73	0.98783	0.98259	0.97259	Img.153	0.97462	0.96375	0.95375
Img.74	0.97172	0.95885	0.94885	Img.154	0.97425	0.96294	0.95294
Img.75	0.97522	0.95213	0.96522	Img.155	0.98603	0.97022	0.96022
Img.76	0.98278	0.96435	0.95435	Img.156	0.98678	0.96457	0.95457
Img.77	0.98119	0.96249	0.95249	Img.157	0.98247	0.96697	0.97247
Img.78	0.98170	0.96541	0.95541	Img.158	0.97245	0.96040	0.95040
Img.79	0.98996	0.96808	0.95808	Img.159	0.98236	0.95973	0.94973
Img.80	0.98739	0.97176	0.96176	Img.160	0.98197	0.96789	0.95982

Table 4 summarizes key statistics for the proposed segmentation method: Average Dice Coefficient: 0.9814, Average F1 Score: 0.9619, Average IoU (Jaccard Index): 0.9661, Average Segmentation Performance Index: 96.98%, Maximum Dice Coefficient: 0.9916, Maximum F1 Score: 0.9872, Maximum IoU: 0.9830, and Maximum Segmentation Performance Index: 98.21%. These results demonstrate the high segmentation performance index of the proposed technique, as compared to existing methods in Table 3. The segmentation performance index for each of the 160 images shown in Fig. 5 is computed by averaging the Dice Coefficient, F1 Score, and IoU values for each image, then scaling the result into a percentage to yield a final accuracy measure for each image. This approach achieves a highest segmentation performance index of approximately 98.21%, out of 160 images.

The segmentation performance index for 160 input images using the suggested method are presented in Fig. 5 and Table 3. Comparisons with existing baseline methods—such as Texture Analysis with Edge Detection, Edge Detection with GMM, Edge Detection with FCN, and K-Means with Mask R-CNN—are shown in Fig. 4 and Table 2. The superior segmentation performance achieved by the proposed segmentation method significantly enhances the disease classification task discussed in subsequent sections, as accurately extracting the lesion region directly improves feature extraction. This enhanced feature extraction, in turn, bolsters the performance of the disease classification model, which effectiveness is entirely dependent on the proposed segmentation approach.

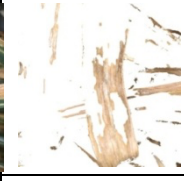
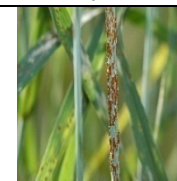
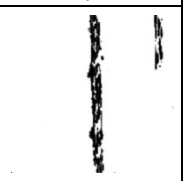
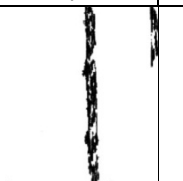
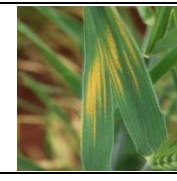
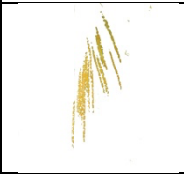
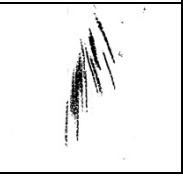
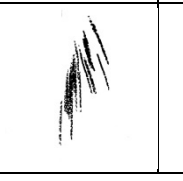
Sr. No.	(i) Input Image	(ii) Segmented image	(iii) Segmented Binary image	(iv) GT Image	(v) Segmentation Performance Index
(1)					95.38%
(2)					97.88%
(3)					93.74%
.	.		.	.	.
.	.		.	.	.
.	.		.	.	.
.	.		.	.	.
(110)					98.21%
.	.		.	.	.
.	.		.	.	.
.	.		.	.	.
.	.		.	.	.
(160)					96.98%

Figure 5. Showcases the segmentation performance index with respect to ground truth image: (i) Input image, (ii) Segmented image, (iii) Segmented binary image, (iv) Ground truth image, and (v) Accuracy.

3.4. Feature Extraction

Feature extraction is essential in computer vision for effectively representing images, with applications in farming, surveillance, and healthcare. The key challenge lies in efficient feature extraction and selection to identify targeted regions accurately. This study develops a codebook by extracting four types of features:

(a) Shape Features: Shape descriptors like Area, Perimeter, Eccentricity, Aspect Ratio, Compactness, Centroid, and Moments are analyzed. For wheat leaf diseases—stripe rust, leaf rust, septoria, and stem rust—**Roundness, Perimeter, and Eccentricity** are the most effective shape features [5].

(b) Color Features: Color features are extracted from RGB, Lab*, HIS, and HSV spaces using metrics like mean, standard deviation, skewness, and variance for each channel. These metrics capture color distribution and asymmetry, aiding in disease classification [52], [33].

(c) Texture Features: Texture analysis uses six Haralick features—contrast, entropy, homogeneity, energy, dissimilarity, and correlation—derived from the Gray-Level Co-occurrence Matrix (GLCM). The GLCM, computed at angles 0° , 45° , 90° , and 135° with distance $d=1$, captures lesion texture variations. [43].

(d) Edge Features: Edge features such as **Edge Density, Contrast, Sharpness, and Entropy** provide structural image information for tasks like segmentation and object recognition. Edge density, calculated as the ratio of edge pixels to total pixels, assesses image complexity [46], [58].

The features obtained are organized in a code-book, and the optimal ones for classification are identified through the suggested hybrid feature selection approach. This method employs PCA to identify principal components, computes skewness for the PCA-transformed features to evaluate distribution and detect outliers, and calculates entropy to analyze texture and complexity. Ultimately, mutual information is evaluated between skewness, entropy-augmented PCA features, and target labels to determine the most significant features. Fig. 6 presents the feature extraction process, showcasing the schematic representation of the proposed framework.

3.4.1. Shape features

This subsection briefly describes the methodology for extracting shape features from segmented wheat leaf lesions. The selected shape features—Roundness, Perimeter, and Eccentricity—are vital for understanding the geometric characteristics of lesions, which can aid in the identification and classification of various wheat diseases [54]. The effectiveness of shape features in relation to disease symptoms for our classification problem of wheat leaf diseases is listed in Table 5.

Each segmented lesion of a wheat leaf is quantified using the following shape features: roundness, perimeter, and eccentricity. These features offer important insights into the geometric properties of lesions, assisting in diagnosing and classifying wheat diseases. The shape features are as follows:

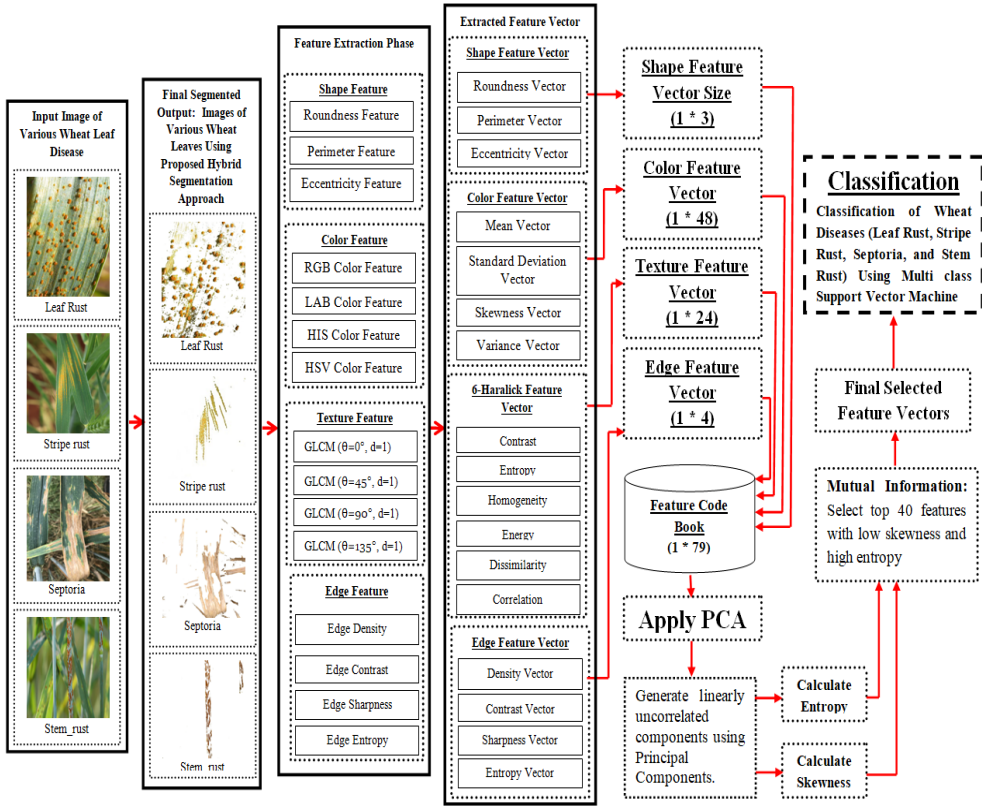


Figure 6. Flow diagram of the proposed classification framework, illustrating feature extraction, codebook construction, feature selection, and classification.

The **Roundness** (R) is also known as circularity, roundness quantifies how close the shape of the lesion is to a perfect circle [19]. A roundness value near 1 signifies a shape that is nearly circular, whereas a value closer to 0 suggests an elongated shape. Roundness is computed using Eq. (10):

$$\text{Roundness (R)} = \frac{4\pi A}{P^2} \quad (10)$$

where:

A: lesion area.

P: lesion perimeter.

The perimeter measures the boundary length of the lesion [55]. It can be computed using a contour tracing method. The perimeter is calculated using Eq. (11) as the total length of the boundary formed by all contour pixels of the lesion.

$$\text{Perimeter (P)} = \sum \text{Boundary Length} \quad (11)$$

Table 5. Shape Feature Analysis for Wheat Leaf Diseases

Sr. No.	Disease	Symptoms	Effective Shape Features	Ref.
1	Leaf Rust	Small, circular to oval, orange-brown pustules encircled by a pale halo.	Roundness: Differentiates circular pustules.	[19]
			Perimeter: Provides size distribution of lesions.	
2	Stripe Rust	Tiny, orange-yellow pustules in streaks along leaf veins.	Perimeter: Measures elongated lesion boundaries.	[55]
			Roundness: Identifies elongated lesions with low values.	
3	Septoria	Irregular, chlorotic lesions with black fruiting bodies (pycnidia).	Eccentricity: Measures lesion elongation.	[40]
			Perimeter: Measures boundary of irregular lesions.	
4	Stem Rust	Larger, reddish-brown, circular pustules on stems and leaves.	Roundness: Indicates circular pustules.	[54]
			Perimeter: Captures detailed lesion borders.	

The Eccentricity (E) describes the elongation of the lesion, indicating how much the shape deviates from being circular [40]. An eccentricity of 0 reflects a perfect circle, while an eccentricity closer to 1 reflects more elongated shapes. Eccentricity is computed using Eq. (12):

$$\text{Eccentricity (E)} = \sqrt{1 - \frac{b^2}{a^2}} \quad (12)$$

where:

Major and minor axes are represented by a and b, respectively.

3.4.2. Color features

First, as explained in Section 3.3, color features are taken out of the segmented images. This part next uses these segmented images to extract color features from lesions on wheat leaves in four different color spaces: RGB, Lab*, HIS, and HSV. In order to properly classify and analyze wheat diseases like leaf rust, septoria, stem rust, and stripe rust, the features mean, standard deviation (SD), skewness, and variance offer a thorough statistical representation of color distributions within the lesions [49].

To represent images in digital imaging, the color space of RGB combines the three primary color channels. Although luminance (L) and color components (A and B channels) are separated in the Lab* (CIELAB) color space, ensuring perceptual uniformity [60]. The HIS (Hue, Intensity, Saturation) color space defines colors based on hue, intensity, and

saturation, closely aligning with human visual perception. Similarly, the HSV color space explore colors through hue, saturation, and value, offering an intuitive approach to color representation [13].

By combining these features, a comprehensive feature vector is formed, providing more detailed information than any individual color space alone. Eq. (13-28) is used to determine the standard deviation, mean, skewness, and variance for every color space of the segmented leaf lesions, aiding in the extraction of color-related features essential for effective disease classification.

The extracted color characteristics from the chosen color spaces are shown in Fig. 6, which also shows how these spaces are used to extract features. These features are then fused into a single vector through concatenation, producing a 1×48 feature vector. The calculations for mean, standard deviation (SD), skewness, and variance across each channel in the four color spaces are conducted using the following mathematical expressions for each color space:

$$Mean_{RGB} = \frac{1}{N} \sum_{i=1}^N C_i \quad (13)$$

$$SD_{RGB} = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_i - Mean_{RGB})^2} \quad (14)$$

$$Skewness_{RGB} = \frac{1}{N} \sum_{i=1}^N \left(\frac{C_i - Mean_{RGB}}{SD_{RGB}} \right)^3 \quad (15)$$

$$Variance_{RGB} = \frac{1}{N} \sum_{i=1}^N (C_i - Mean_{RGB})^2 \quad (16)$$

Where C_i represents color components (R, G, B)

$$Mean_{LAB} = \frac{1}{N} \sum_{i=1}^N C_i \quad (17)$$

$$SD_{LAB} = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_i - Mean_{LAB})^2} \quad (18)$$

$$Skewness_{LAB} = \frac{1}{N} \sum_{i=1}^N \left(\frac{C_i - Mean_{LAB}}{SD_{LAB}} \right)^3 \quad (19)$$

$$Variance_{LAB} = \frac{1}{N} \sum_{i=1}^N (C_i - Mean_{LAB})^2 \quad (20)$$

Where C_i represents color components (L, A, B)

$$Mean_{HIS} = \frac{1}{N} \sum_{i=1}^N C_i \quad (21)$$

$$SD_{HIS} = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_i - Mean_{HIS})^2} \quad (22)$$

$$Skewness_{HIS} = \frac{1}{N} \sum_{i=1}^N \left(\frac{C_i - Mean_{HIS}}{SD_{HIS}} \right)^3 \quad (23)$$

$$Variance_{HIS} = \frac{1}{N} \sum_{i=1}^N (C_i - Mean_{HIS})^2 \quad (24)$$

Where C_i represents color components (H, I, S)

$$Mean_{HSV} = \frac{1}{N} \sum_{i=1}^N C_i \quad (25)$$

$$SD_{HSV} = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_i - Mean_{HSV})^2} \quad (26)$$

$$Skewness_{HSV} = \frac{1}{N} \sum_{i=1}^N \left(\frac{C_i - Mean_{HSV}}{SD_{HSV}} \right)^3 \quad (27)$$

$$Variance_{HSV} = \frac{1}{N} \sum_{i=1}^N (C_i - Mean_{HSV})^2 \quad (28)$$

Where C_i represents color components (H, S, V)

3.4.3. Texture features

Texture features play a crucial role in disease classification based on texture in the agricultural sector. Many researchers commonly utilize 14 Haralick features for texture analysis, while some have expanded this set to 18 features in an effort to improve classification accuracy. However, not all of these additional features prove beneficial across all types of problems, and they can often lead to overfitting, potentially degrading classification performance [49].

Table 6. Selected Haralick and GLCM Texture Features and Their Relevance for Wheat Leaf Disease Classification.

Sr. No.	Feature	Justification for Selection	Ref.
1	Contrast	Lesions and disease spots create significant local variations in intensity. Higher contrast values help in identifying areas with these variations.	[43]
2	Entropy	Disease affects the regularity of the leaf texture, increasing the complexity and randomness. Higher entropy values can indicate diseased regions.	[53]
3	Homogeneity-1	Healthy leaves generally have more uniform textures. Lower homogeneity values can indicate irregular textures associated with disease.	[15]
4	Energy	Uniform textures result in higher energy values. Diseased areas with irregular textures show lower energy.	[57]
5	Dissimilarity	Measures how different the pixel values are from each other, which is useful in identifying the irregularities caused by disease.	[47]
6	Correlation-1	Captures the linear dependency between pixel values, which can change due to disease-induced texture alterations.	[50]

In this study, the focus is on effective wheat disease classification, including stem rust, leaf rust, stripe rust, and septoria. The feature selection process reduces an initial set of 18 Haralick texture features (Energy, Contrast, Homogeneity-1, Variance, Correlation, Sum-Average, Sum-Variance, Sum-Entropy, Entropy, Difference-Variance, Difference-Entropy, Correlation-1, Correlation-2, Maximal Correlation Coefficient, Homogeneity-2, Cluster-Shade, Cluster-Prominence, and Haralick Correlation) together with an additional GLCM descriptor (Dissimilarity) to six features: Contrast, Entropy, Homogeneity-1, Energy, Correlation-1, and Dissimilarity. Although dissimilarity is not part of the original Haralick set, it is widely used to measure linear gray-level variations and complements contrast-

based difference statistics. The final features were selected based on their statistical significance, discriminative power, and effectiveness in capturing lesion texture characteristics, as detailed in Table 6.

In the analysis of wheat leaf lesions caused by diseases like septoria, leaf rust, stripe rust and stem rust, the GLCM (Gray-Level Co-occurrence Matrix) plays a pivotal role in detecting textural differences. Computed for each direction ($\theta = 0^\circ, 45^\circ, 90^\circ$, and 135°) with a distance ($d = 1$), the GLCM captures spatial relationships between pixel pairs at varied orientations [28]. From these GLCMs, five Haralick texture features and one additional GLCM descriptor (dissimilarity) are extracted using Eqs. (29–34) given below.

Contrast: Measures intensity contrast between pixels and their neighbors across the entire image, highlighting local GLCM variations. Higher values denote sharper texture transitions [28].

$$\text{Contrast} = \sum_{i,j} (i - j)^2 P(i, j) \quad (29)$$

Entropy: Quantifies randomness and complexity within the GLCM, reflecting image disorder. Elevated entropy values indicate a more intricate texture pattern [53].

$$\text{Entropy} = - \sum_{i,j} P(i, j) \log P(i, j) \quad (30)$$

Homogeneity: Assesses proximity of GLCM element distribution to its diagonal, signifying smoother textures with values concentrated near it [15].

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|} \quad (31)$$

Energy: Represents the textural uniformity measured by the sum of squared GLCM elements. More consistent texture patterns are indicated by higher energy values [28].

$$\text{Energy} = \sum_{i,j} P(i, j)^2 \quad (32)$$

Dissimilarity: Measures pixel pair differences, emphasizing texture variations. Greater values denote increased textural diversity [28].

$$\text{Dissimilarity} = \sum_{i,j} |i - j| P(i, j) \quad (33)$$

Correlation: Assesses pixel correlation across the image, indicating gray level linear dependency. Higher correlation values imply a more linear texture pattern [50].

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)P(i, j)}{\sigma_i \sigma_j} \quad (34)$$

Where:

i: The GLCM's row index, which indicates the reference pixel's intensity

j: The column index in the GLCM, representing the intensity value of the neighboring pixel.

$P(i, j)$: The probability that intensity values i and j will co-occur is represented by the normalized value in the GLCM at position (i, j).

μ_i : Using Equation (35) as a guide, the mean of the GLCM values for row i is:

$$\mu_i = \sum_j P(i, j) \quad (35)$$

μ_j : Using Eq. (36) the mean of the GLCM values for column j was determined as follows:

$$\mu_j = \sum_i P(i, j) \quad (36)$$

σ_i : The GLCM values standard deviation for row i , who shows the dispersion of values close to the mean μ_i , computed using Eq. (37):

$$\sigma_i = \sqrt{\sum_j (i - \mu_i)^2 \sum_j P(i, j)} \quad (37)$$

σ_j : The GLCM values standard deviation in column j , which indicates the dispersion of values around the mean μ_j , computed using Eq. (38) as:

$$\sigma_j = \sqrt{\sum_i (j - \mu_j)^2 \sum_i P(i, j)} \quad (38)$$

To enhance the accuracy of disease classification, these traits are crucial in capturing various textural elements in wheat leaf lesions affected by various diseases. They are crucial for precise disease detection and classification, as they utilize detailed texture information encoded in GLCM orientations. The fusion of GLCM and the selected six texture features, as illustrated in Fig. 6, results in a final texture vector with dimensions of 1×24 .

3.4.4. Edge features

In image processing and computer vision, edge features are essential because they provide important information about the spatial content and structural properties of digital images. For tasks like object recognition, image segmentation, and edge detection, these characteristics are crucial, where understanding the presence, orientation, and quality of edges is crucial for accurate analysis and interpretation. Various edge features can be used in classification problems, such as Edge Density, Edge Contrast, Edge Sharpness, and Edge Entropy [33].

Among all edge features, Edge Density is a key feature as it effectively captures the frequent and complex boundaries characteristic of these lesions. There are various methods for extracting edge density features, but based on a comparative analysis of famous techniques like Canny, Sobel, and Prewitt edge detection methods (as shown in Fig. 7), the Canny approach was chosen for this paper because it can improve the model's classification accuracy. Furthermore, a thorough grasp of the lesion patterns can be obtained by combining Edge Density with other edge-related characteristics like Edge Contrast, Edge Sharpness, and Edge Entropy. Better differentiation and classification of septoria, leaf rust, stem rust, and stripe rust in wheat leaves are made possible by this multi-feature approach.

The frequency of edges in an image is measured by edge density. It produces a single scalar value and is computed as the ratio of edge pixels to total number of pixels. This metric offers a rapid evaluation of edge spatial distribution and serves as a fundamental descriptor of image complexity [33]. It can be computed using the expression given in Eq. (39).

$$\text{Edge Density} = \frac{N_e}{N} \quad (39)$$

where:

- N_e denotes the count of edge pixels.
- N denotes the total count of pixels.

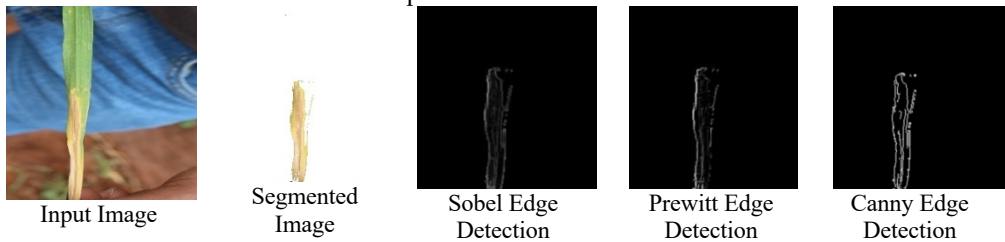


Figure 7. Comparative analysis of edge detection methods for effective Edge Density extraction.

According to Eq. (40), edge contrast quantifies the average variation in pixel intensities across detected edges. It provides a numerical assessment of the sharpness and clarity of the image's edge boundaries. While low contrast values imply smoother gradients, high contrast values show distinct transitions between neighboring pixels. Applications needing improved image visualization and quality evaluation will benefit from this feature [27].

$$\text{Edge Contrast} = \frac{1}{N_e} \sum_{(x,y) \in E} |I(x,y) - I_{avg}(x,y)| \quad (40)$$

where:

- The set of edge pixels is represented by E .
- The intensity at pixel (x,y) is denoted by $I(x,y)$.
- The average intensity of the area surrounding pixel (x,y) is represented by $I_{avg}(x,y)$.

The gradient magnitude at edge locations, which represents the rate at which pixel intensity changes across boundaries, is measured by edge sharpness. For the purpose of recognizing edge-related features in segmented images of leaf lesions, higher sharpness values signify clearer edge definitions [44]. The formula for calculating it is given in Eq. (41).

$$\text{Edge Sharpness} = \frac{1}{N_e} \sum_{(x,y) \in E} |\sqrt{G_x(x,y)^2 + G_y(x,y)^2}| \quad (41)$$

where:

- The gradients in the x- and y-directions are denoted by the symbols $G_x(x,y)$ and $G_y(x,y)$, respectively.
- E is the set of edge pixels.

Edge entropy quantifies the complexity or randomness of edge distributions in an image, based on the entropy of the edge map. A more complicated edge distribution is indicated by higher entropy, whereas homogeneity is suggested by lower entropy [56]. Eq. (42) is used to estimate edge entropy.

$$\text{Edge entropy (H)} = - \sum_{i=1}^n p_i \log(p_i) \quad (42)$$

where:

- p_i is the probability of edge orientation i occurring (obtained from the histogram).
- The number of orientation bins is n .

By creating an edge feature vector of size 4, each feature vector represents one of the previously mentioned edge features, which are essential for disease classification.

3.4.5. Codebook construction

Four feature categories are extracted: shape, color, texture, and edge. These features are then combined using a simple concatenation-based fusion method to create a codebook. This process sequentially combines individual feature sets, enhancing correlations within the extracted features, reducing variations between classes, and integrating diverse feature information into a comprehensive representation. The resulting unified and cohesive codebook facilitates more effective analysis and classification [42]. Subsequently, a feature codebook of size 1×79 is constructed and refined using a novel hybrid feature selection method.

3.5. Feature selection and SVM-based Classification

In machine learning and computer vision applications, high-dimensional feature sets increase computational cost, processing time, and memory usage, while redundant features may degrade classification performance. To address this issue, feature selection techniques are used to identify the most informative subset of features while eliminating irrelevant or redundant ones.

This study proposes a hybrid feature selection framework integrating Principal Component Analysis (PCA), skewness, entropy, and mutual information (MI) to improve wheat disease classification. From an initial set of 79 features, the method selects the 40 most relevant features, producing a compact and discriminative representation for classification.

First, PCA reduces dimensionality by transforming correlated features into principal components that retain maximum variance. Next, skewness analyzes feature distributions and detects potential outliers, while entropy measures the uncertainty within the data. Subsequently, mutual information (MI) evaluates the dependency between features and disease labels, enabling the selection of the most informative features. By integrating PCA-transformed features with skewness and entropy through MI analysis, the method identifies the top 40 significant features, as illustrated in Fig. 6. This strategy helps mitigate overfitting, enhance prediction performance, and reduce feature redundancy.

The optimized feature subset is then provided to a Multi class Support Vector Machine (M-SVM) classifier for disease classification. By retaining only the most discriminative features, the method forms a structured feature space where different disease classes become more separable. This improved separability enables the M-SVM to construct precise decision boundaries and maximize the margin between classes, leading to improved classification performance.

Since the framework targets multiple wheat diseases, including Leaf Rust, Stripe Rust, Septoria, and Stem Rust, a multi-class SVM strategy is employed to learn decision boundaries between disease categories. During training, the SVM determines optimal separating hyperplanes within the optimized feature space. As a result, the classifier can accurately predict the disease class of unseen wheat leaf samples, achieving improved accuracy, reduced overfitting, and better generalization. The complete process from feature selection to classification is explained in Steps 1–7.

Step 1: Dimensionality Reduction with PCA

Let the original feature set be:

$$X = \{X_1, X_2, X_3, \dots \dots \dots X_{79}\}$$

The i -th characteristic is denoted by X_i . These features are converted into principle components, which are linearly uncorrelated, using PCA:

$$Y = \{Y_1, Y_2, Y_3, \dots \dots \dots Y_k\}, \quad k \leq 79$$

where Y_i are the principal components. The number of components, k as given in Eq. (43), is determined by retaining those components that explain at least 90% of the total variance:

$$\text{Variance Explained by } k = \sum_{i=1}^k \lambda_i \text{ such that } \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^{79} \lambda_i} \geq 0.9 \quad (43)$$

Here, λ_i represents the eigenvalues of the covariance matrix.

Step 2: Skewness Calculation

For the PCA-transformed features, skewness is calculated to analyze data distribution and identify outliers. Features with low skewness are preferred as they reflect more typical data patterns. The skewness of the i -th principal component Y_i is given by Eq. (44):

$$\text{Skewness of } Y_i = \frac{E[(Y_i - \mu_i)^3]}{\sigma_i^3} \quad (44)$$

Where μ_i is mean and σ_i is standard deviation of Y_i .

Step 3: Entropy Calculation

Entropy is computed to measure uncertainty or randomness in each PCA-transformed feature Y_i . Higher entropy indicates more informative features. The entropy of Y_i is represented by Eq. (45):

$$(Y_i) = -\sum_j p_j \log(p_j) \quad (45)$$

The probability of the j -th bin in the Y_i histogram is denoted by p_j .

Step 4: Mutual Information Calculation

To further refine feature selection, mutual information (MI) is calculated between the PCA-transformed features augmented with skewness and entropy (Z) and the target labels (C). Let the augmented feature set be:

$$Z = \{Z_1, Z_2, Z_3, \dots \dots \dots Z_k\}$$

The MI between a feature Z_i and the target C is computed using Eq. (46) given as:

$$MI(Z_i, C) = \iint p(z_i, c) \log \left(\frac{p(z_i, c)}{p(z_i) p(c)} \right) dz_i dc \quad (46)$$

where the joint probability distribution of Z_i and C is denoted by $p(z_i, c)$.

Step 5: Feature Ranking and Selection

Each feature Z_i is ranked based on a composite score that integrates skewness, entropy, and MI computed using Eq. (47) given as:

$$\text{Score}(Z_i) = w_1 \cdot \frac{1}{\text{Skewness}(Z_i)} + w_2 \cdot H(Z_i) + w_3 \cdot MI(Z_i, C) \quad (47)$$

where w_1, w_2, w_3 are weights assigned to each metric.

Based on their scores, the top 40 features are chosen. These features exhibit low skewness, high entropy, and high mutual information, making them essential and diverse for accurately classifying wheat diseases such as leaf rust, septoria, stem rust, and stripe rust.

Step 6: Final Feature Set

The final selected feature set is:

$$F = \{Z_1, Z_2, Z_3, \dots \dots \dots Z_{40}\}$$

This hybrid feature selection method effectively reduces dimensionality, identifies informative features, and improves classification accuracy by addressing overfitting and redundancy challenges.

Step 7: Multi-class Support Vector Machine (M-SVM) Classifier

To perform the final disease classification, a **Multi-class Support Vector Machine (M-SVM)** classifier is employed. Support Vector Machine (SVM) is a supervised learning algorithm widely used for classification tasks due to its strong generalization capability and effectiveness in handling high-dimensional feature spaces. The primary objective of SVM is to determine an optimal separating hyperplane that maximizes the margin between samples belonging to different classes [41].

Given a training dataset (x_i, y_i) , where x_i represents the extracted feature vector and $y_i \in \{1, 2, \dots, K\}$ denotes the corresponding disease class label, the decision function of SVM can be expressed as:

$$f(x) = w^T x + b \quad (48)$$

where w represents the weight vector and b denotes the bias term. The optimal hyperplane is obtained by solving the following optimization problem:

$$\min_{w, b, \xi} \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi_i \quad (49)$$

subject to:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (50)$$

where C is the regularization parameter and ξ_i are slack variables that allow soft margin classification.

Since the proposed framework involves multiple wheat disease classes, including Leaf Rust, Stripe Rust, Septoria, and Stem Rust, a multi-class extension of SVM is required. In proposed work, the multi-class classification problem is handled using the One-vs-One (OvO) strategy, where binary classifiers are constructed for every pair of disease classes. The final class label is then determined using a majority voting scheme based on the outputs of these pairwise classifiers.

Using the optimized feature subset obtained after feature selection, the M-SVM learns discriminative boundaries among disease classes. This enables the classifier to accurately categorize unseen wheat leaf samples, resulting in improved classification accuracy, robustness, and generalization performance.

4. Experimental results

A dataset of 1000 images, compiled from publicly accessible sources such as the CIMMYT Dataset and Wheat Leaf Dataset, is utilized to validate the proposed methodology. A 10-fold cross-validation approach is applied to evaluate the result. The framework includes two key processes: lesion segmentation and the classification of wheat lesions into four categories, namely septoria, and rust of stripe, stem and leaf. Detailed results for lesion segmentation are discussed in Section 3.3.4, while this section emphasizes the classification outcomes derived from the framework.

4.1. Dataset Details of Wheat Rust and Septoria

The specific technical details of the selected wheat diseases, including sample images, the number of images used for classification in the proposed framework, as well as the causes, symptoms, and environmental factors promoting their spread, are summarized in Table 7 below.

To perform the classification, features related to shape, color, texture, and edge properties are extracted to create a comprehensive feature representation. This feature set is further refined using a method of hybrid selection incorporating Principal Component

Analysis (PCA), skewness, entropy, and mutual information (MI), optimizing the classification accuracy of segmented lesion images. The primary classification method employed is a M-SVM, which is benchmarked against several established models like Gradient-Boosting Machines, Random-Forest, Ensemble-Boosted-Tree, and Weighted K-Nearest Neighbors (W-KNN).

Accuracy, Precision (PPV), Recall (TPR), False Positive Rate (FPR), F1-Score, False Discovery Rate (FDR), False Negative Rate (FNR), and Classification Time (for one image) are the eight evaluation metrics utilized to assess the efficacy of the suggested approach. The combination of accuracy, PPV, TPR, and F1-score evaluates overall and class-wise performance, while error-based metrics (e.g., FPR, FNR, and FDR) provide insights into

Table 7. Detailed characteristics of wheat diseases, including sample images, number of images used in the proposed classification framework, causes, symptoms, and environmental conditions favoring their spread.

Sr. No	Disease Type	Sample Image	No. of Images	Cause of Disease	Symptoms	Conditions Favoring Spread	Ref
1	Wheat Leaf Rust (Brown Rust)	 (Wheat Leaf dataset)	250 Images	Fungus <i>Puccinia triticina</i> .	Appearance of tiny, orange-brown pustules on the foliage and stems. Leaves may dry out earlier than usual.	Mild temperatures (15–22°C) with wet weather.	[35]
2	Wheat Stem Rust (Black Rust)	 (CIMMYT Dataset)	250 Images	Fungus <i>Puccinia graminis tritici</i> .	Brick-red pustules on stems and leaves. Weak stems.	Warm, humid conditions (25–30°C).	[48]
3	Wheat leaf Stripe Rust (Yellow Rust)	 (Wheat Leaf dataset)	250 Images	Fungus <i>Puccinia striiformis</i> .	Yellow pustules along leaf veins. Stunted growth.	Cool (10–15°C), moist nights.	[55]
4	Wheat Leaf Septoria	 (Wheat Leaf dataset)	250 Images	Fungus <i>Septoria tritici</i> .	Brown lesions with black pycnidia.	High humidity (>85%) and prolonged wetness.	[60]

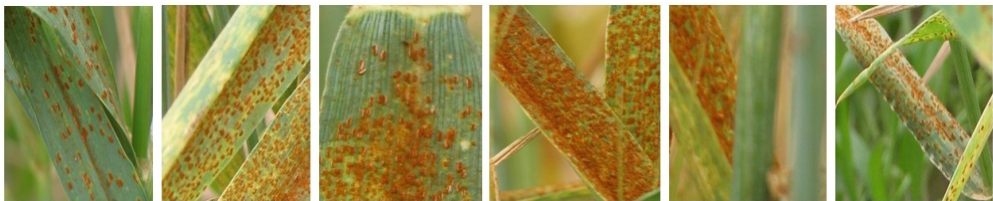
misclassification patterns. The inclusion of Classification time ensures that the method is practical for real-world applications.

For further information on classifier performance across different thresholds, particularly in the setting of multiclass classification issues, this publication also includes the Area Under the Curve.

All experiments were carried out using Python on a local machine featuring a Core i5 (12th Gen) processor, 16GB of RAM, and a 4GB RTX Nvidia Graphics Card.

4.2. Dataset and Evaluation of Selected Wheat Diseases

This study utilizes a comprehensive dataset comprising images of wheat diseases, specifically Septoria and rust disease of leaf, stem and stripe. The dataset contains 1000 high-quality images of diseased wheat plants, including leaves and stems. Each image has dimensions of 444×387 pixels with a resolution of 96 dpi, as shown in Fig. 8(a-d).



(a) Wheat Leaf Rust (Source: Wheat Leaf dataset)



(b) Wheat Stem Rust (Source: CIMMYT Dataset)



(c) Wheat leaf Stripe Rust (Source: Wheat Leaf dataset)



(d) Wheat Leaf Septoria (Source: Wheat Leaf dataset)

Figure 8. Sample images of selected diseases (a) Wheat Leaf Rust (b) Wheat Stem Rust (c) Wheat leaf Stripe Rust (d) Wheat Leaf Septoria.

For this research, four major wheat diseases were selected: Septoria and rust disease of leaf, stem and stripe. The selected diseases were initially evaluated against healthy wheat images to understand their unique characteristics and visual patterns. Subsequently, these diseases were classified using the proposed hybrid segmentation and M-SVM techniques. The thorough examination results demonstrate how well the model detects and categorizes the chosen diseases. The results of the detailed evaluation are then displayed.

4.2.1. Wheat Leaf Rust

Leaf Rust in wheat manifests as tiny pustules with an orange-brown hue, harboring fungal spores that accelerate premature leaf aging and markedly impair photosynthetic performance. For the experimental evaluation, a dataset comprising 250 images of Wheat Leaf Rust-affected samples was employed for both training and testing.

Table 8. Performance measures on wheat leaf rust.

Method	Accuracy	PPV	TPR	F1-Score	FPR	FNR	FDR	Classification Time
Proposed Framework	98.4	0.976	0.98	0.977	0.16	0.19	0.024	5.9
RF	96.2	0.958	0.96	0.958	0.94	0.38	0.042	9.7
EBT	94.3	0.936	0.94	0.937	2.16	1.98	0.064	10.2
GBM	93.5	0.939	0.94	0.939	2.96	3.56	0.061	11.4
W-KNN	92.8	0.925	0.93	0.927	5.39	6.46	0.075	7.6

Table 8 presents an analysis that compares the proposed framework against four advanced classifiers. The experimental findings reveal that the proposed method surpasses other approaches, achieving a peak accuracy of 98.4% with a minimum False Positive Rate (FPR) of 0.16. Furthermore, Table 9's confusion matrix provides additional evidence of the framework's efficacy.

Table 9. Wheat leaf rust classification confusion matrix.

Leaf Types Image	Unhealthy (%)	Healthy (%)
Unhealthy wheat leaf rust	98.57%	1.42%
Healthy wheat leaf	0%	100%

4.2.2. Wheat Stem Rust

Stem rust in wheat is characterized by elongated, reddish-brown pustules containing fungal spores on the stems, leaves, and leaf sheaths. This causes significant physiological stress in the plant and a notable reduction in grain yield. For experimental evaluation, a dataset of 250 images of Wheat Stem Rust-affected samples was utilized for both training and testing.

Table 10. Performance measures on wheat stem rust.

Method	Accuracy	PPV	TPR	F1-Score	FPR	FNR	FDR	Classificati on Time
Proposed Framework	97.9	0.977	0.98	0.978	0.12	0.31	0.023	6.1
RF	95.6	0.952	0.96	0.955	0.84	0.46	0.048	10.1
EBT	95.2	0.949	0.95	0.949	3.56	2.28	0.051	10.8
GBM	93.1	0.93	0.93	0.932	3.94	4.76	0.07	11.9
W-KNN	91.3	0.909	0.91	0.909	5.68	5.96	0.091	8.3

Table 10 provides a detailed evaluation of the developed framework against four advanced classification techniques. The method demonstrated outstanding results, achieving a peak accuracy of 97.9% and a minimum FPR of 0.12, outperforming existing approaches. The reliability of the framework is further confirmed using the confusion matrix presented in Table 11.

Table 11. Confusion matrix for stem rust in wheat disease.

Image	Unhealthy wheat stem rust (%)	Healthy wheat leaf (%)
Unhealthy wheat stem rust	97.5%	2.5%
Healthy wheat leaf	0%	100%

4.2.3. Wheat Leaf Stripe Rust

Stripe Rust of Wheat Leaf is distinguished by the formation of yellow-orange, elongated stripe-like pustules of fungal spores on wheat leaves, primarily along the veins, leading to reduced photosynthetic activity and significant yield losses. For the experimental evaluation, a dataset containing 250 images of Wheat Leaf Stripe Rust-infected samples was employed for both training and testing.

Table 12. Performance measures on wheat leaf stripe rust.

Method	Accuracy	PPV	TPR	F1-Score	FPR	FNR	FDR	Classification Time
Proposed Framework	99.1	0.989	0.99	0.989	0.03	0.01	0.011	5.8
RF	97.4	0.972	0.97	0.970	0.76	0.64	0.028	9.9
EBT	94.8	0.947	0.95	0.948	2.96	3.15	0.053	10.5
GBM	93.7	0.935	0.94	0.937	4.13	3.94	0.065	11.3
W-KNN	92.1	0.918	0.92	0.918	6.17	5.43	0.082	8.1

Table 12 provides a comprehensive evaluation of the designed framework against four advanced classification models. The method achieved remarkable results, attaining a peak accuracy of 99.1% and a minimum FPR of 0.03, surpassing current techniques. The framework's dependability is further illustrated in Table 13 with the help of confusion matrix.

Table 13. The confusion matrix of stripe rust in wheat disease.

Leaf Types Image	Unhealthy wheat leaf rust (%)	Healthy wheat leaf (%)
Unhealthy wheat leaf rust	99.09%	00.91%
Healthy wheat leaf	00.90%	99.01%

4.2.4. Wheat Leaf Septoria

Wheat Leaf Septoria is characterized by the presence of irregular, necrotic lesions on wheat leaves, often surrounded by yellowish halos, leading to reduced photosynthetic efficiency and premature leaf senescence. The origin of the disease is fungus *Septoria tritici*, can significantly impact wheat yield under favorable conditions such as prolonged leaf wetness and high humidity. For the experimental analysis, a dataset of 250 images of Wheat Leaf Septoria-infected samples was utilized for model training and testing.

A thorough comparison of the suggested framework with four cutting-edge classifiers is given in Table 14. The developed method demonstrated outstanding results, achieving a top accuracy of 98.9% and recording a minimum FPR of 0.27, surpassing other existing methods. The confusion matrix shown in Table 15 provides additional validation of the framework's robustness and dependability.

Table 14. Performance measures on wheat leaf septoria.

Method	Accuracy	PPV	TPR	F1-Score	FPR	FNR	FDR	Classification Time
Proposed Framework	98.9	0.987	0.989	0.9878	0.27	0.15	0.013	6.0
RF	97.1	0.97	0.971	0.971	1.26	2.72	0.03	10.0
EBT	94.5	0.945	0.946	0.946	3.59	2.96	0.055	10.7
GBM	93.3	0.933	0.935	0.934	5.27	4.74	0.067	11.5
W-KNN	91.8	0.916	0.918	0.918	8.96	6.74	0.084	8.2

Table 15. Confusion matrix of wheat leaf septoria.

Leaf Types Image	Unhealthy wheat leaf septoria (%)	Healthy wheat leaf (%)
Unhealthy wheat leaf septoria	98.76%	01.24%
Healthy wheat leaf	01.11%	98.89%

Table 16. Consolidated performance metrics for wheat disease classifiers: Accuracy, F1-Score, FPR, FNR, PPV, TPR, FDR, and Classification Time.

Method	Class	Accuracy	PPV	TPR	F1-Score	FPR	FNR	FDR	Classification Time
Proposed Framework	Leaf rust	98.4	0.976	0.981	0.977	0.16	0.19	0.024	5.9
	Stem rust	97.9	0.977	0.982	0.978	0.12	0.31	0.023	6.1
	Stripe rust	99.1	0.989	0.990	0.989	0.03	0.01	0.011	5.8
	Septoria	98.9	0.987	0.989	0.987	0.27	0.15	0.013	6
Random Forest (RF)	Leaf rust	96.2	0.958	0.961	0.958	0.94	0.38	0.042	9.7
	Stem rust	95.6	0.952	0.960	0.955	0.84	0.46	0.048	10.1
	Stripe rust	97.4	0.972	0.970	0.970	0.76	0.64	0.028	9.9
	Septoria	97.1	0.97	0.971	0.971	1.26	2.72	0.031	10
Ensemble Boosted Tree (EBT)	Leaf rust	94.3	0.936	0.941	0.937	2.16	1.98	0.064	10.2
	Stem rust	95.2	0.949	0.952	0.949	3.56	2.28	0.051	10.8
	Stripe rust	94.8	0.947	0.951	0.948	2.96	3.15	0.053	10.5
	Septoria	94.5	0.945	0.946	0.946	3.59	2.96	0.055	10.7
Gradient Boosting Machines (GBM)	Leaf rust	93.5	0.939	0.940	0.939	2.96	3.56	0.061	11.4
	Stem rust	93.1	0.930	0.931	0.932	3.94	4.76	0.072	11.9
	Stripe rust	93.7	0.935	0.940	0.937	4.13	3.94	0.065	11.3
	Septoria	93.3	0.933	0.935	0.934	5.27	4.74	0.067	11.5
Weighted K-Nearest Neighbor (W-KNN)	Leaf rust	92.8	0.925	0.931	0.927	5.39	6.46	0.075	7.6
	Stem rust	91.3	0.909	0.912	0.909	5.68	5.96	0.091	8.3
	Stripe rust	92.1	0.918	0.921	0.918	6.17	5.43	0.082	8.1
	Septoria	91.8	0.916	0.918	0.918	8.96	6.74	0.084	8.2

4.2.5. Comprehensive Performance Analysis Across Wheat Diseases

To evaluate the proposed classification framework for each disease individually, a thorough comparison was conducted across four wheat diseases: Wheat Rust of Stem, Leaf, and Stripe and Wheat Leaf Septoria. Accuracy, F1-Score, FPR, FNR, PPV, TPR and FDR are among the important performance metrics that are compiled in Table 16 along with the classification time for every classifier. The findings show that the proposed framework consistently exceeds other methods in accuracy, precision, and efficiency, achieving a classification accuracy of 99.1% with minimal FPR (0.03) and FNR (0.01) for Stripe Rust. This emphasizes the framework's robustness and reliability for real-world applications. Fig. 9 and Fig. 10 compare classification accuracy and time for W-KNN, GBM, EBT, RF, and the proposed framework across four wheat diseases. The proposed framework consistently achieves the highest accuracy, exceeding 98%, outperforming RF (96–97%), EBT (94–95%), and GBM and W-KNN (90–93%), making it highly effective for agricultural monitoring.

In terms of efficiency, the proposed framework also excels, with classification times under 6 seconds for all diseases, compared to GBM (over 11 seconds), EBT and RF (9–10 seconds), and W-KNN (7–8 seconds). This blend of superior accuracy and speed highlights its suitability for real-time agricultural applications and its practicality for precision farming systems.

4.2.6. Final Classification Results for All Four Diseases

To evaluate the proposed classification framework, a 10-fold cross-validation strategy was applied to a dataset containing 1,000 images, with 250 images representing each of the four wheat diseases. The dataset was partitioned into 10 equal folds using stratified sampling to

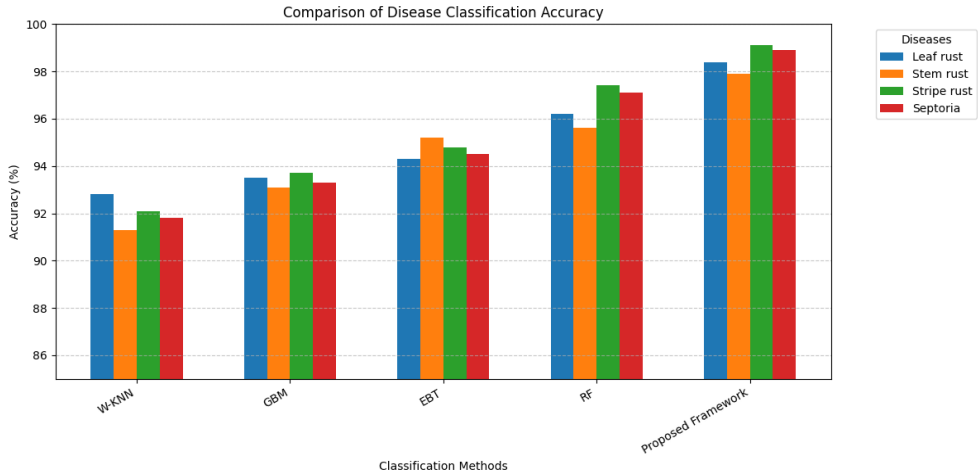


Figure 9. Classification accuracy comparison showing the proposed framework surpassing W-KNN, GBM, EBT, and RF with over 98% accuracy for four wheat diseases.

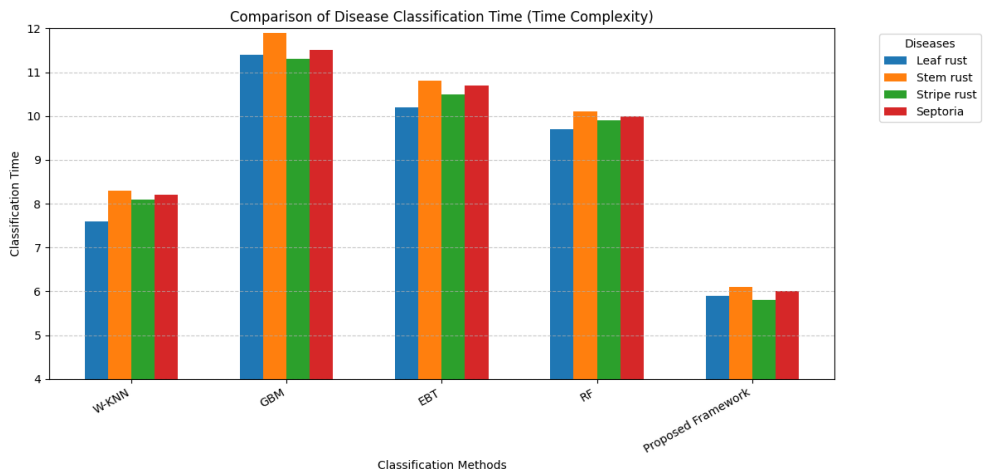


Figure 10. Classification time comparison showing the proposed framework achieving faster results (under 6 seconds) than W-KNN, GBM, EBT, and RF for all four wheat diseases.

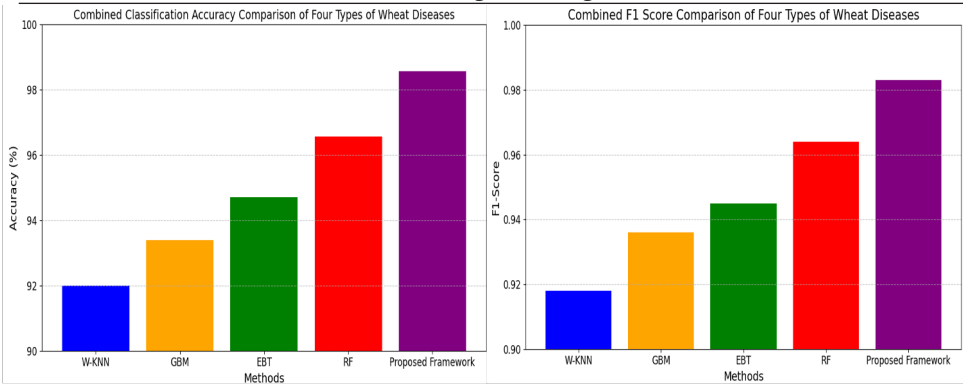


Figure 11-a

Figure 11-b

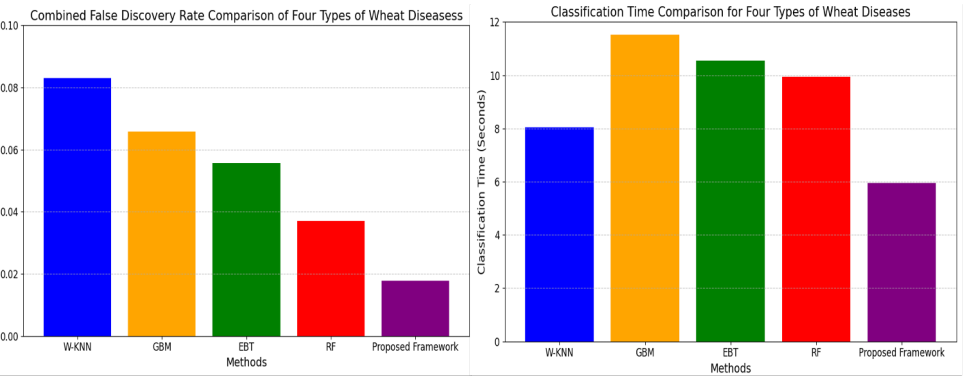
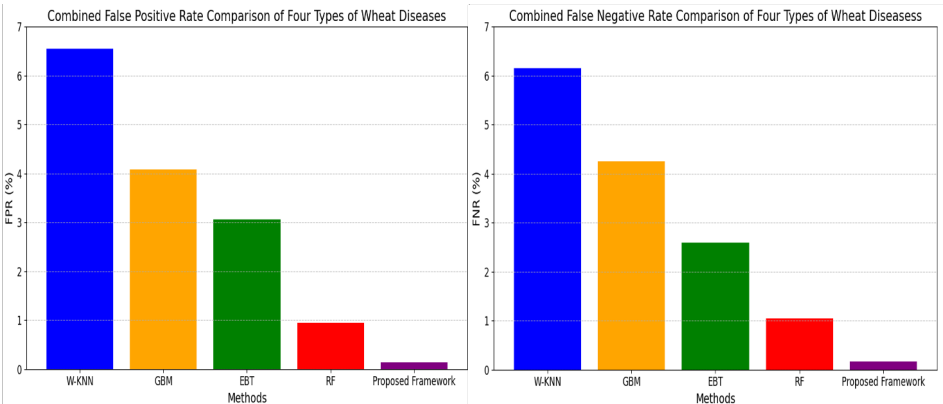


Figure 11-e

Figure 11-f

Figure 11. Performance comparison in Fig. 11 (a-e) for Accuracy, F1-Score, FPR, FNR, FDR, and Classification Time, showcasing the proposed framework's superior accuracy (98.58%), F1-Score (0.983), and efficiency (5.95 seconds) compared to W-KNN, GBM, EBT, and RF.

ensure balanced representation across all disease categories. During the process, the model underwent training and testing over 10 iterations, where nine-folds were allocated for training and one-fold for validation. This method ensured that each fold served as the validation set exactly once, while contributing to the training set in the other iterations.

Evaluation metrics, such as accuracy, F1-score, FPR, FNR, FDR, AUC (Area Under the ROC Curve), and computational time, were calculated exclusively on the validation set during each iteration. For an impartial assessment of the model's capability to generalize over unseen data, these metrics play a vital role. Upon completing all 10 iterations, the averaged performance metrics offered a dependable estimation of the framework's overall classification effectiveness.

Table 17. Final classification outcomes for all four types of wheat diseases.

Class	Accuracy	F1-Score	FPR	FNR	FDR	AUC	Classification Time
Proposed Framework	98.58	0.983	0.145	0.17	0.018	0.990	5.95
Random Forest (RF)	96.57	0.964	0.950	1.05	0.037	0.982	9.93
Ensemble Boosted Tree (EBT)	94.70	0.945	3.068	2.59	0.056	0.954	10.55
Gradient Boosting Machines (GBM)	93.40	0.936	4.075	4.25	0.066	0.941	11.53
Weighted K-Nearest Neighbor (W-KNN)	92.00	0.918	6.550	6.15	0.083	0.935	8.05

The proposed framework scores a peak classification performance of 98.58% and an AUC value of 0.99. In comparison, W-KNN attained an accuracy of 92%, with an FPR of 6.55 and an FNR of 6.15. The classification results for all four diseases, along with comparisons to alternative approaches, are detailed in Table 17. The confusion matrix in Table 18 is presented for additional verification of the classification outcomes.

A relative assessment of performance metrics for Several classifiers, including the Proposed Framework, RF, EBT, GBM, and W-KNN, is presented with the help of bar graphs in Fig. 11(a-f). The Proposed Framework excels in key performance metrics, achieving top accuracy (98.58%), F1-Score (0.983), and AUC (0.990), while also recording minimal rates for False Positives (FPR: 0.145), False Negatives (FNR: 0.17), and False Discoveries (FDR: 0.018).

Table 18. Confusion Matrix for All four Disease Types: R1 (Leaf.Rust), R2 (Stem.Rust), R3 (Stripe.Rust), S (Septoria).

Leaf Types Image	R1	R2	R3	S
R1	97.6	0.8%	1.2%	0.4%
R2	2.8	96.4%	0.8%	-
R3	-	-	100%	-
S	-	0.9%	-	99.1%

In terms of classification time, the Proposed Framework demonstrates significant efficiency, achieving an average time of 5.95 seconds, which is notably faster than RF

(9.93 seconds), EBT (10.55 seconds), GBM (11.53 seconds), and W-KNN (8.05 seconds). These outcomes emphasize the ability of the framework to balance accuracy, robustness, and computational speed, thereby positioning it as a viable approach for real-time wheat disease identification and classification within precision agriculture systems.

4.2.7. ROC Curve and AUC Comparison

The Receiver Operating Characteristic (ROC) curve presented in Figure 12 evaluates the effectiveness of the proposed work in comparison with other machine learning algorithms, including RF, EBT, GBM, and W-KNN. Plotting of the ROC curve is materialized with the help of TPR against FPR under several threshold conditions, based on the data from Table 19.

Table 19. Threshold-based TPR, FPR, and AUC comparison for Proposed Framework, RF, EBT, GBM, and W-KNN, highlighting superior performance of the proposed model.

Thresh old	TPR Proposed	FPR Proposed	TPR (RF)	FPR (RF)	TPR (EBT)	FPR (EBT)	TPR (GBM)	FPR (GBM)	TPR (W- KNN)	FPR (W- KNN)
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.9	0.80	0.005	0.78	0.015	0.76	0.027	0.75	0.033	0.70	0.041
0.8	0.90	0.008	0.88	0.021	0.85	0.040	0.83	0.048	0.80	0.061
0.7	0.95	0.009	0.94	0.025	0.92	0.047	0.91	0.055	0.88	0.073
0.6	0.97	0.010	0.96	0.028	0.94	0.051	0.93	0.061	0.90	0.078
0.5	0.989	0.011	0.971	0.029	0.946	0.054	0.935	0.065	0.918	0.082
0.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Final AUC using ROC curve	0.99		0.98		0.95		0.94		0.93	

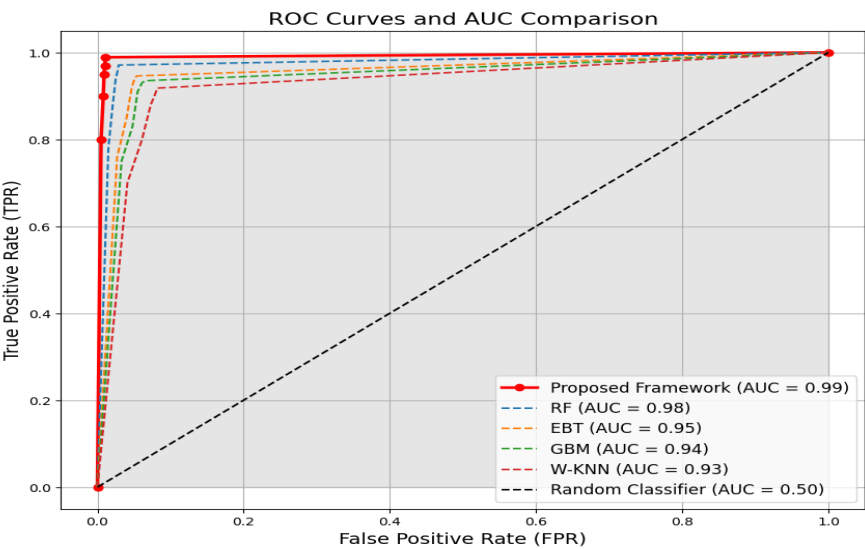


Figure 12. ROC curves and AUC comparison showing the superior performance of the proposed framework (AUC = 0.99) over RF (0.98), EBT (0.95), GBM (0.94), and W-KNN (0.93).

The proposed framework achieves an AUC of 0.9910, outperforming models like RF (AUC = 0.9752) and EBT (AUC = 0.9539). This indicates the framework's strong capacity to separate between target and non-target classes. The steep slope near the origin and the low FPR at high TPR values further demonstrate the framework's reliability and efficiency in managing imbalanced datasets, ensuring accurate predictions essential for practical applications.

Furthermore, Table 20 offers a comparative evaluation of the proposed classification framework with existing methods for all four diseases, Wheat Septoria and wheat rust of Leaf, Stem and Stripe Rust. The results clearly highlight the superiority of the proposed approach in delivering enhanced performance metrics.

Table 20. Comparison of the classification accuracy achieved by the proposed framework with existing methods.

Existing Methods (References)	Year	Accuracy
[43]	2022	92.3
[11]	2021	94.2
[51]	2024	98%
[16]	2024	98.43%
Proposed Classification Framework	-	98.58%

5. Discussion

The proposed framework presents a novel hybrid pipeline that integrates Adaptive K-means, GMM and advanced color manipulation, thereby achieving superior leaf lesion segmentation compared with the use of individual methods in isolation. Building upon this segmentation stage, the framework incorporates hybrid feature engineering and classification using M-SVM, resulting in a holistic approach that ensures both precise lesion segmentation and reliable disease classification. By leveraging features derived from segmented lesion pixels, the M-SVM performs well even on small datasets, mitigating overfitting issues commonly observed in deep learning-based approaches (as seen in Figure 1). The adaptive K-means clustering algorithm effectively separates background pixels from the input image during the segmentation phase by dynamically determining the ideal cluster count. This flexibility accommodates variations in cluster distribution across diverse input images. Subsequently, the segmentation process is further enhanced using Gaussian Mixture Models (GMM) and advanced color manipulation techniques.

The efficiency of the proposed segmentation approach is evident in Fig. 3(d), which showcases accurate lesion segmentation, surpassing other clustering techniques by minimizing the loss or mixing of lesion pixels with background pixels. This robust and adaptable methodology ensures reliable segmentation outcomes, forming a strong basis for the classification stage.

In order to assess the effectiveness of the proposed segmentation technique, Fig. 4 presents a comparative analysis relative to existing approaches. The quantitative metrics outlined in Table 3 and detailed in Table 4 demonstrate an Average Dice Coefficient of 0.9814, an Average F1-Score of 0.9619, and an Average Jaccard Index of 0.9661, emphasizing high precision and consistency in segmentation. Additionally, Fig. 5 visually and numerically confirms alignment with ground truth images, achieving a segmentation performance index of 98.21%.

The segmented images are analyzed using four feature types: shape, color, texture, and edge, which are integrated to form a feature vector. A feature selection process, depicted in Fig. 6, identifies the most influential features. The M-SVM classifier, which has benefits like better managing small datasets, less overfitting than deep learning, quicker training and classification, and enhanced interpretability of feature-based models, is then fed the chosen features. Using this approach, the proposed framework achieved classification accuracies of 98.4%, 97.9%, 99.1%, and 98.9% among these, rust diseases—such as leaf, stem, and stripe rust—along with wheat leaf septoria, respectively. Performance metrics for the proposed framework and corresponding confusion matrices are detailed in Tables 8, 10, 12, and 14, with supplementary data in Tables 9, 11, 13, and 15.

Based on Table 17 and confirmed by the confusion matrix in Table 18, the overall classification accuracy for all four diseases is 98.58%. This demonstrates the framework's robustness in identifying and categorizing wheat diseases.

A comparative performance analysis is carried out using metrics like Accuracy, F1-Score, FPR, FNR, FDR, and Classification Time in order to guarantee the framework's reliability. With a classification accuracy of 98.58%, an F1-Score of 0.983, and computational efficiency of 5.95 seconds, the framework outperforms models such as RF, GBM, EBT, and W-KNN, as shown in Fig. 11 (a–e).

The ROC and AUC curves, which analyze the trade-offs between true positive and false positive rates, are finally shown in Fig. 12. The proposed framework achieves the highest AUC score of 0.99, surpassing alternative methods such as RF (0.98), EBT (0.95), GBM (0.94), and W-KNN (0.93).

Finally, Table 20 provides a detailed comparison of the proposed algorithm with other techniques, demonstrating its reliability and effectiveness. The findings confirm the superiority of the proposed classification framework, particularly when integrated with M-SVM, over existing approaches, ensuring improved accuracy and efficiency for agricultural disease management.

6. Conclusion and Future Directions

This study proposes a novel framework for identifying and classifying wheat lesions among these rust diseases—such as leaf, stem and stripe rust—along with wheat leaf septoria. The framework combines a strong classification model with sophisticated lesion segmentation approaches. To precisely isolate lesions, segmentation uses Gaussian Mixture Models (GMM), adaptive k-means clustering, and sophisticated color manipulation.

The framework extracts shape, color, texture, and edge-based features, which are further refined using PCA, skewness, entropy, and Mutual Information to eliminate redundancy and enhance classification performance. By deriving effective features from the segmented

lesions, the framework addresses the challenge of small datasets and mitigates the overfitting issues commonly observed in deep learning-based classifiers. As a result, the proposed M-SVM classifier achieves high sensitivity, specificity, and accuracy. Experimental validation on the Wheat Leaf and CIMMYT datasets confirms the superior performance of the framework compared to existing methods in both segmentation and classification.

While the proposed approach is effective, there are opportunities for enhancement. Future research could extend the framework to accommodate additional diseases and crops, incorporate deep learning for automated feature extraction, with large dataset and optimize computational efficiency for real-time field applications. Expanding datasets to include diverse environmental conditions and incorporating temporal data for disease progression tracking could further improve the robustness and applicability of the framework.

This study provides a strong foundation for advancing wheat disease management, with potential to evolve into a scalable, real-time solution for agricultural diagnostics.

References:

- [1] Basavaiah, Jagadeesh, and Audre Arlene Anthony. 2020. "Tomato Leaf Disease Classification Using Multiple Feature Extraction Techniques." *Wireless Personal Communications* 115 (1): 633–51. <https://doi.org/10.1007/s11277-020-07590-x>.
- [2] Bi, Zhicheng, and Peng Cao. 2021. "Color Space Conversion Algorithm and Comparison Study." *Journal of Physics: Conference Series* 1976 (1). <https://doi.org/10.1088/1742-6596/1976/1/012008>.
- [3] Bukhari, Hassan Raza, Rafia Mumtaz, Salman Inayat, Uferah Shafi, Ihsan Ul Haq, Syed Mohammad Hassan Zaidi, and Maryam Hafeez. 2021. "Assessing the Impact of Segmentation on Wheat Stripe Rust Disease Classification Using Computer Vision and Deep Learning." *IEEE Access* 9: 164986–4. <https://doi.org/10.1109/ACCESS.2021.3134196>.
- [4] Choi, D., W. S. Lee, R. Ehsani, J. Schueller, and F. M. Roka. 2016. "Detection of Dropped Citrus Fruit on the Ground and Evaluation of Decay Stages in Varying Illumination Conditions." *Computers and Electronics in Agriculture* 127: 109–19. <https://doi.org/10.1016/j.compag.2016.05.020>.
- [5] Dhingra, Gittaly, Vinay Kumar, and Hem Dutt Joshi. 2018. "Study of Digital Image Processing Techniques for Leaf Disease Detection and Classification." *Multimedia Tools and Applications* 77 (15): 19951–0. <https://doi.org/10.1007/s11042-017-5445-8>.
- [6] Diao, Zhihua, Anping Zheng, and Yuanyuan Wu. 2012. "Shape Feature Extraction of Wheat Leaf Disease Based on Invariant Moment Theory." *IFIP Advances in Information and Communication Technology* 369 AICT (PART 2): 168–73. https://doi.org/10.1007/978-3-642-27278-3_18.
- [7] Eriksson, Inger, and Natalia Tabachnikova. 2022. "Learning Models." *LUMAT: International Journal on Math, Science and Technology Education* 10 (2). <https://doi.org/10.31129/lumat.10.2.1681>.

- [8] Fahim-Ul-Islam, Md, Amitabha Chakrabarty, Sarder Tanvir Ahmed, Rafeed Rahman, Hyun Han Kwon, and Md Jalil Piran. 2024. "A Comprehensive Approach Toward Wheat Leaf Disease Identification Leveraging Transformer Models and Federated Learning." *IEEE Access* 12 (July): 109128–56. <https://doi.org/10.1109/ACCESS.2024.3438544>.
- [9] Figueroa, Melania, Kim E Hammond-Kosack, and Peter S Solomon. 2018. "A Review of Wheat Diseases—a Field Perspective." *Molecular Plant Pathology* 19 (6): 1523–36. <https://doi.org/https://doi.org/10.1111/mpp.12618>.
- [10] Gao, Chunfeng, Xingjie Ji, Qiang He, Zheng Gong, Heguang Sun, Tiantian Wen, and Wei Guo. 2023. "Monitoring of Wheat Fusarium Head Blight on Spectral and Textural Analysis of UAV Multispectral Imagery." *Agriculture (Switzerland)* 13 (2): 1–16. <https://doi.org/10.3390/agriculture13020293>.
- [11] Genaev, Mikhail A., Ekaterina S. Skolotneva, Elena I. Gulyaeva, Elena A. Orlova, Nina P. Bechtold, and Dmitry A. Afonnikov. 2021. "Image-Based Wheat Fungi Diseases Identification by Deep Learning." *Plants* 10 (8). <https://doi.org/10.3390/plants10081500>.
- [12] Grosse-Heilmann, Malin, Elena Cristiano, Roberto Deidda, and Francesco Viola. 2024. "Durum Wheat Productivity Today and Tomorrow: A Review of Influencing Factors and Climate Change Effects." *Resources, Environment and Sustainability* 17 (September): 100170. <https://doi.org/10.1016/j.resenv.2024.100170>.
- [13] Hasan, Md Zahid, Nahid Zeba, Abdul Malek, and Sanjida Sultana Reya. 2021. "A Leaf Disease Classification Model in Betel Vine Using Machine Learning Techniques." *International Conference on Robotics, Electrical and Signal Processing Techniques*, 362–66. <https://doi.org/10.1109/ICREST51555.2021.9331142>.
- [14] Hatice Catal Reis, Veysel Turk. 2024. "Integrated Deep Learning and Ensemble Learning Model for Deep Feature-Based Wheat Disease Detection." *Microchemical Journal* 197. <https://doi.org/https://doi.org/10.1016/j.microc.2023.109790>.
- [15] Hayıt, Tolga, Hasan Erbay, Fatih Varçın, Fatma Hayıt, and Nilüfer Akci. 2023. "The Classification of Wheat Yellow Rust Disease Based on a Combination of Textural and Deep Features." *Multimedia Tools and Applications* 82 (30): 47405–23. <https://doi.org/10.1007/s11042-023-15199-y>.
- [16] Islam, Usman, Wakeel Ahmad, Syed Adnan, and Ijst Jr. 2024. "Leveraging CIELAB Segmentation and CNN for Wheat Fungi Disease Classification," no. July.
- [17] Jebarani, P. Esther, N. Umadevi, Hien Dang, and Marc Pomplun. 2021. "A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection." *IEEE Access* 9: 146153–62. <https://doi.org/10.1109/ACCESS.2021.3123425>.
- [18] Kakumanu, P., S. Makrogiannis, and N. Bourbakis. 2007. "A Survey of Skin-Color Modeling and Detection Methods." *Pattern Recognition* 40 (3): 1106–22. <https://doi.org/10.1016/j.patcog.2006.06.010>.
- [19] Kaur, Sukhvir, Shreelekha Pandey, and Shivani Goel. 2019. "Plants Disease Identification and Classification Through Leaf Images: A Survey." *Archives of*

- Computational Methods in Engineering* 26 (2): 507–30.
<https://doi.org/10.1007/s11831-018-9255-6>.
- [20] Kebapci, Hanife, Berrin Yanikoglu, and Gozde Unal. 2011. “Plant Image Retrieval Using Color, Shape and Texture Features.” *Computer Journal* 54 (9): 1475–90.
<https://doi.org/10.1093/comjnl/bxq037>.
- [21] Khan, Habib, Ijaz Ul Haq, Muhammad Munsif, Mustaqeem, Shafi Ullah Khan, and Mi Young Lee. 2022. “Automated Wheat Diseases Classification Framework Using Advanced Machine Learning Technique.” *Agriculture (Switzerland)* 12 (8): 1–20.
<https://doi.org/10.3390/agriculture12081226>.
- [22] Khan, M. Attique, Tallha Akram, Muhammad Sharif, Aamir Shahzad, Khursheed Aurangzeb, Musaed Alhussein, Syed Irtaza Haider, and Abdualziz Altamrah. 2018. “An Implementation of Normal Distribution Based Segmentation and Entropy Controlled Features Selection for Skin Lesion Detection and Classification.” *BMC Cancer* 18 (1): 1–20. <https://doi.org/10.1186/s12885-018-4465-8>.
- [23] Kumar, Deepak, Vinay Kukreja, and Amitoj Singh. 2024. “A Novel Hybrid Segmentation Technique for Identification of Wheat Rust Diseases.” *Multimedia Tools and Applications* 83 (29): 72221–51. <https://doi.org/10.1007/s11042-024-18463-x>.
- [24] Li, Xinwu. 2010. “Research on Color Space Conversion Model between XYZ and RGB.” *Key Engineering Materials* 428–429: 466–69.
<https://doi.org/10.4028/www.scientific.net/KEM.428-429.466>.
- [25] Liu, Haiyun, Hongbo Chen, Jianming Du, Chengjun Xie, Qiong Zhou, Rujing Wang, and Lin Jiao. 2024. “Auto-Adjustment Label Assignment-Based Convolutional Neural Network for Oriented Wheat Diseases Detection.” *Computers and Electronics in Agriculture* 222: 109029.
<https://doi.org/https://doi.org/10.1016/j.compag.2024.109029>.
- [26] Liu, Yangyang, Li Ren, Xuehong Cao, and Ying Tong. 2020. “Breast Tumors Recognition Based on Edge Feature Extraction Using Support Vector Machine.” *Biomedical Signal Processing and Control* 58: 101825.
<https://doi.org/10.1016/j.bspc.2019.101825>.
- [27] May, Keith A., and Mark A. Georgeson. 2007. “Blurred Edges Look Faint, and Faint Edges Look Sharp: The Effect of a Gradient Threshold in a Multi-Scale Edge Coding Model.” *Vision Research* 47 (13): 1705–20.
<https://doi.org/10.1016/j.visres.2007.02.012>.
- [28] Moya, Luis, Homa Zakeri, Fumio Yamazaki, Wen Liu, Erick Mas, and Shunichi Koshimura. 2019. “3D Gray Level Co-Occurrence Matrix and Its Application to Identifying Collapsed Buildings.” *ISPRS Journal of Photogrammetry and Remote Sensing* 149 (January): 14–28. <https://doi.org/10.1016/j.isprsjprs.2019.01.008>.
- [29] Nema, Sumit, and Aarju Dixit. 2018. “Wheat Leaf Detection and Prevention Using Support Vector Machine.” *2018 International Conference on Circuits and Systems in Digital Enterprise Technology, ICCSDET 2018*, 2294–98.
<https://doi.org/10.1109/ICCSDET.2018.8821098>.

- [30] Nezhad, S.S. Amiri. 2016. "Modeling and Mapping Leaf Area Index Using Spectral and Spatial Information of High - Resolution Multispectral Images : Case Study of Bavarian Forest National Park," no. February: 68p.
- [31] Ngugi, Lawrence C., Moataz Abelwahab, and Mohammed Abo-Zahhad. 2020. "Tomato Leaf Segmentation Algorithms for Mobile Phone Applications Using Deep Learning." *Computers and Electronics in Agriculture* 178 (August): 105788. <https://doi.org/10.1016/j.compag.2020.105788>.
- [32] Nugroho, Hanung Adi, Dina Arifatul Khusna, and Eka Legya Frannita. 2019. "Detection and Classification of Breast Nodule on Ultrasound Images Using Edge Feature." *International Journal of Advances in Soft Computing and Its Applications* 11 (1): 64–80.
- [33] Patil, Jayamala Kumar, and Raj Kumar. 2017. "Analysis of Content Based Image Retrieval for Plant Leaf Diseases Using Color, Shape and Texture Features." *Engineering in Agriculture, Environment and Food* 10 (2): 69–78. <https://doi.org/10.1016/j.eaef.2016.11.004>.
- [34] Pradhan, Debasmita, Sudarsan Padhy, and Biswajit Sahoo. 2017. "Enzyme Classification Using Multiclass Support Vector Machine and Feature Subset Selection." *Computational Biology and Chemistry* 70: 211–19. <https://doi.org/https://doi.org/10.1016/j.compbiolchem.2017.08.009>.
- [35] Prasad, Pramod, Siddanna Savadi, S. C. Bhardwaj, and P. K. Gupta. 2020. "The Progress of Leaf Rust Research in Wheat." *Fungal Biology* 124 (6): 537–50. <https://doi.org/10.1016/j.funbio.2020.02.013>.
- [36] Pydipati, R., T. F. Burks, and W. S. Lee. 2006. "Identification of Citrus Disease Using Color Texture Features and Discriminant Analysis." *Computers and Electronics in Agriculture* 52 (1–2): 49–59. <https://doi.org/10.1016/j.compag.2006.01.004>.
- [37] Revathi, P., and M. Hemalatha. 2012. "Advance Computing Enrichment Evaluation of Cotton Leaf Spot Disease Detection Using Image Edge Detection." *2012 3rd International Conference on Computing, Communication and Networking Technologies, ICCCNT 2012*, no. July: 1-5. IEEE, 2012. <https://doi.org/10.1109/ICCCNT.2012.6395903>.
- [38] Reza, Md Nasim, In Seop Na, Sun Wook Baek, and Kyeong Hwan Lee. 2019. "Rice Yield Estimation Based on K-Means Clustering with Graph-Cut Segmentation Using Low-Altitude UAV Images." *Biosystems Engineering* 177 (2018): 109–21. <https://doi.org/10.1016/j.biosystemseng.2018.09.014>.
- [39] Ribeiro-Gomes, Krishna, David Hernandez-Lopez, Rocío Ballesteros, and Miguel A. Moreno. 2016. "Approximate Georeferencing and Automatic Blurred Image Detection to Reduce the Costs of UAV Use in Environmental and Agricultural Applications." *Biosystems Engineering* 151: 308–27. <https://doi.org/10.1016/j.biosystemseng.2016.09.014>.
- [40] Sarayloo, Zahra, and Davud Asemani. 2015. "Designing a Classifier for Automatic Detection of Fungal Diseases in Wheat Plant: By Pattern Recognition Techniques." *ICEE 2015 - Proceedings of the 23rd Iranian Conference on Electrical Engineering*

- 10: 1193–97. <https://doi.org/10.1109/IranianCEE.2015.7146396>.
- [41] Scharr, Hanno, Massimo Minervini, Andrew P. French, Christian Klukas, David M. Kramer, Xiaoming Liu, Imanol Luengo, et al. 2016. “Leaf Segmentation in Plant Phenotyping: A Collation Study.” *Machine Vision and Applications* 27 (4): 585–606. <https://doi.org/10.1007/s00138-015-0737-3>.
- [42] Sepúlveda, Jorge, and Sergio A. Velastín. 2015. “F1 Score Assesment of Gaussian Mixture Background Subtraction Algorithms Using the MuHAVi Dataset.” *IET Seminar Digest* 2015 (5): 1–6. <https://doi.org/10.1049/ic.2015.0106>.
- [43] Setiawan, Agung W. 2020. “Image Segmentation Metrics in Skin Lesion: Accuracy, Sensitivity, Specificity, Dice Coefficient, Jaccard Index, and Matthews Correlation Coefficient.” *CENIM 2020 - Proceeding: International Conference on Computer Engineering, Network, and Intelligent Multimedia 2020*, no. Cenim 2020: 97–102. <https://doi.org/10.1109/CENIM51130.2020.9297970>.
- [44] Shafi, Uferah, Rafia Mumtaz, Ihsan Ul Haq, Maryam Hafeez, Naveed Iqbal, Arslan Shaukat, Syed Mohammad Hassan Zaidi, and Zahid Mahmood. 2022. “Wheat Yellow Rust Disease Infection Type Classification Using Texture Features.” *Sensors* 22 (1): 1–21. <https://doi.org/10.3390/s22010146>.
- [45] Sharif, Muhammad, Muhammad Attique Khan, Zahid Iqbal, Muhammad Faisal Azam, M. Ikram Ullah Lali, and Muhammad Younus Javed. 2018. “Detection and Classification of Citrus Diseases in Agriculture Based on Optimized Weighted Segmentation and Feature Selection.” *Computers and Electronics in Agriculture* 150 (April): 220–34. <https://doi.org/10.1016/j.compag.2018.04.023>.
- [46] Singh, Ranjeet, and Umesh Chandra Jaiswal. 2025. “An Innovative Machine Learning Framework for Precise Quantification of Leaf Disease Severity via Proposed Hybrid Segmentation and Color Manipulation Technique.” *Engineering Journal* 29 (5): 37–59. <https://doi.org/10.4186/ej.2025.29.5.37>.
- [47] Singh, Ravi P., David P. Hodson, Yue Jin, Evans S. Lagudah, Michael A. Ayliffe, Sridhar Bhavani, Matthew N. Rouse, et al. 2015. “Emergence and Spread of New Races of Wheat Stem Rust Fungus: Continued Threat to Food Security and Prospects of Genetic Control.” *Phytopathology* 105 (7): 872–84. <https://doi.org/10.1094/PHYTO-01-15-0030-FI>.
- [48] Singh, Sanskriti. 2022. “Emphasis on the Minimization of False Negatives or False Positives in Binary Classification.” <http://arxiv.org/abs/2204.02526>.
- [49] Sinha, Aditya, and Rajveer Singh Shekhawat. 2020. “Olive Spot Disease Detection and Classification Using Analysis of Leaf Image Textures.” *Procedia Computer Science* 167 (2019): 2328–36. <https://doi.org/10.1016/j.procs.2020.03.285>.
- [50] Sun, Weizhong, Bohan Fu, and Zhao Zhang. 2023. “Maize Nitrogen Grading Estimation Method Based on UAV Images and an Improved Shufflenet Network.” *Agronomy* 13 (8). <https://doi.org/10.3390/agronomy13081974>.
- [51] Tian, Kai, Jiu hao Li, Jiefeng Zeng, Asenso Evans, and Lina Zhang. 2019. “Segmentation of Tomato Leaf Images Based on Adaptive Clustering Number of K-

- Means Algorithm.” *Computers and Electronics in Agriculture* 165 (March): 104962. <https://doi.org/10.1016/j.compag.2019.104962>.
- [52] Tian, Yunong, Guodong Yang, Zhe Wang, En Li, and Zize Liang. 2020. “Instance Segmentation of Apple Flowers Using the Improved Mask R-CNN Model.” *Biosystems Engineering* 193: 264–78. <https://doi.org/10.1016/j.biosystemseng.2020.03.008>.
- [53] Vinta, Surendra Reddy, Ashok Kumar Koshariya, S. Sampath Kumar, Aditya, and Annantharao Gottimukkala. 2024. “Investigation of Early Symptoms of Tomato Leaf Disorder by Using Analysing Image and Deep Learning Models.” *EAI Endorsed Transactions on Internet of Things* 10: 1–5. <https://doi.org/10.4108/eetiot.4815>.
- [54] Vishnoi, Vibhor Kumar, Krishan Kumar, and Brajesh Kumar. 2021. *Plant Disease Detection Using Computational Intelligence and Image Processing*. *Journal of Plant Diseases and Protection*. Vol. 128. Springer Berlin Heidelberg. <https://doi.org/10.1007/s41348-020-00368-0>.
- [55] Wang, Hongli, Qian Jiang, Zhenyu Sun, Shiqin Cao, and Haiguang Wang. 2023. “Identification of Stripe Rust and Leaf Rust on Different Wheat Varieties Based on Image Processing Technology.” *Agronomy* 13 (1). <https://doi.org/10.3390/agronomy13010260>.
- [56] Wang, Jianjia, J. Huo, and L. Zhang. 2019. “Thermodynamic Edge Entropy in Alzheimer’s Disease.” *Pattern Recognition Letters* 125: 570–75. <https://doi.org/10.1016/j.patrec.2019.06.026>.
- [57] Wang, Z, K Wang, S Pan, and Y Han. 2018. “Segmentation of Crop Disease Images with an Improved K-Means Clustering Algorithm.” *Applied Engineering in Agriculture* 34 (2017): 277–89. <https://doi.org/https://doi.org/10.13031/aea.12205>.
- [58] Zeng, Huanzhao. 2011. “Preferred Skin Colour Reproduction.” <http://search.ebscohost.com/login.aspx?direct=true&db=ddu&AN=4CF55E3A0B64F905&site=ehost-live>.
- [59] Zhang, Haixia, and Qingxiu Peng. 2022. “PSO and K-Means-Based Semantic Segmentation toward Agricultural Products.” *Future Generation Computer Systems* 126: 82–87. <https://doi.org/10.1016/j.future.2021.06.059>.
- [60] Zhang, Shanwen, Subing Zhang, Chuanlei Zhang, Xianfeng Wang, and Yun Shi. 2019. “Cucumber Leaf Disease Identification with Global Pooling Dilated Convolutional Neural Network.” *Computers and Electronics in Agriculture* 162 (February): 422–30. <https://doi.org/10.1016/j.compag.2019.03.012>.

Received 01.11.2025, Accepted 14.04.2026