Automatic Segmentation of Plant Leaves Disease Detection

Uzum Stanley Ekene
uzums@roehampton.ac.uk, UZU22571175
MSC Data Science, Application of Data Science

ABSTRACT

The proliferation of plant diseases poses a significant threat to global agriculture, leading to reduced crop yields and economic losses. Traditional methods of disease detection and monitoring are often time-consuming and labor-intensive, hindering timely intervention. In response to this challenge, the project "Automatic Segmentation of Plant Leaves Disease Detection" aims to develop an efficient and automated system for detecting diseases in plant leaves.

The project leverages image processing and machine learning techniques to automate the detection process. Specifically, it focuses on the segmentation of plant leaves from images and the subsequent identification of disease-affected regions within these segmented leaves. The proposed system utilizes MATLAB as the primary platform for implementing the algorithms and developing a user-friendly graphical user interface (GUI).

The GUI facilitates the input of leaf images and provides visual feedback on the segmentation and disease detection results. Through a combination of image enhancement, clustering, and feature extraction methods, the system accurately identifies and quantifies disease-affected areas within the plant leaves. Additionally, machine learning classifiers are employed to classify the type of disease present based on extracted features.

This project aims to contribute to the advancement of precision agriculture by providing farmers and researchers with a reliable tool for early disease detection and monitoring. By automating the process and reducing dependence on manual inspection, the system enables proactive disease management strategies, leading to improved crop health and productivity. The project addresses a pressing need in agriculture and demonstrates the

potential of technology to revolutionize disease detection and management in plant pathology.

TABLE OF CONTENT

I. INTRODUCTION

In the World, agriculture faces the perennial challenge of optimizing crop yield while combating the threat of plant diseases. With agriculture serving as the backbone for most of the nation's economy, ensuring the health and productivity of crops is of paramount importance. Timely detection and effective management of plant diseases are critical to safeguarding crop production and sustaining agricultural livelihoods. Manual monitoring of crops for disease symptoms is labor-intensive, time-consuming, and often prone to human error. Hence, there is a pressing need for automated systems that can rapidly and accurately detect plant diseases, enabling proactive interventions to minimize yield losses. In recent years, advancements in computer vision and machine learning have opened new avenues for automated plant disease detection. Leveraging these technologies, researchers have developed innovative solutions for analyzing digital images of plant leaves to identify disease symptoms. Image-based disease detection holds promise for enabling early diagnosis and targeted treatment, thereby enhancing crop resilience and productivity. However, despite the progress made in this field, several challenges remain, including the accurate segmentation of plant leaves from complex backgrounds and the robust identification of disease spots within leaf images.

This proposed project, **"Automatic Segmentation of Plant Leaves Disease Detection,"** seeks to address these challenges by introducing techniques for the automated segmentation and recognition of disease spots in plant leaves. Drawing inspiration from recent advancements in computer vision and machine learning, the project aims to develop a comprehensive framework that encompasses image acquisition, preprocessing, segmentation, feature extraction, and classification steps. By integrating state-of-the-art algorithms and methodologies, this contributed in creating a robust and efficient system capable of detecting a wide range of plant diseases across different crop species.

One of the key innovations is the utilization of the Min-Max Hue Histogram and K-means clustering algorithms for automated leaf segmentation and disease spot identification. The Min-Max Hue Histogram technique provides a robust measure of color distribution in leaf images, enabling precise localization of disease spots based on hue values. Additionally, the K-means clustering algorithm facilitates the segmentation of

plant leaves into distinct regions, allowing for the isolation and analysis of disease-affected areas.
This project introduces the rank-order fuzzy (ROF) filter for image denoising, enhancing leaf image quality by reducing noise. This improves disease spot segmentation and classification accuracy. The proposed algorithms will undergo extensive testing.
The project significantly advances automated plant disease detection with innovative techniques for leaf segmentation and disease spot identification. Leveraging computer vision and machine learning, it aims to develop cost-effective, scalable solutions for crop monitoring, promoting agricultural sustainability worldwide.

Importance of Agriculture in the World: Most of the countries in the world like India, Nigeria, Ghana e.t.c economy is deeply intertwined with agriculture, which employs a significant portion of the population and contributes substantially to the country's GDP. However, the agricultural sector faces numerous challenges, including unpredictable weather patterns, limited access to resources, and the constant threat of pests and diseases. In this context, the timely detection and management of plant diseases play a crucial role in ensuring food security and economic stability.

Significance of Crop Monitoring: Crop monitoring is essential for detecting early signs of disease outbreaks, nutrient deficiencies, water stress, and other factors that can affect crop health and productivity. By continuously monitoring crops, farmers can take proactive measures to mitigate risks and optimize yields. However, traditional methods of crop monitoring, such as visual inspection and manual data collection, are labor-intensive and often unreliable, particularly for large-scale agricultural operations.

Role of Technology in Agriculture: Advances in technology, particularly in the fields of computer vision, machine learning, and remote sensing, have revolutionized agriculture by enabling the development of precision farming techniques and automated monitoring systems. These technologies allow for the collection of high-resolution data on crop health, soil conditions, and environmental factors, facilitating data-driven decision-making and precision agriculture practices.

Challenges in Plant Disease Detection: Plant diseases pose a significant threat to crop yields, causing substantial losses in production and economic revenue. However, detecting and diagnosing plant diseases can be challenging due to the complex nature of plant-pathogen interactions, the variability of disease symptoms, and the diverse range of pathogens that can infect crops. Additionally, manual methods of disease detection are often time-consuming, subjective, and prone to errors, highlighting the need for automated and objective approaches to disease diagnosis.

Advancements in Image-Based Disease Detection: In recent years, there has been a growing interest in image-based methods for plant disease detection, driven by advances in computer vision and machine learning algorithms. These methods leverage digital images of plant leaves to identify disease symptoms based on visual cues such as discoloration, lesions, and abnormal growth patterns. By analyzing leaf images using image processing techniques and machine learning algorithms, researchers can accurately diagnose diseases and provide timely recommendations for disease management.

State-of-the-Art Algorithms and Techniques: Several state-of-the-art algorithms and techniques have been proposed for image-based plant disease detection, including convolutional neural networks (CNNs), support vector machines (SVMs), and decision trees. These algorithms have shown promising results in accurately classifying diseased and healthy plants based on leaf images. However, the accurate segmentation of plant leaves from complex backgrounds and the robust identification of disease spots within leaf images remain challenging tasks that require further research and development.

Motivation for the Proposed Project: The proposed project is motivated by the need to develop automated solutions for the segmentation and detection of plant diseases, with a focus on enhancing the accuracy and efficiency of disease diagnosis. By leveraging innovative algorithms and methodologies, the project aims to overcome the limitations of existing approaches and provide farmers with cost-effective and scalable tools for crop monitoring and disease management. Through collaboration with agricultural stakeholders and field testing in real-world environments, it validate the effectiveness of its proposed solutions and contribute to the advancement of agricultural technology in any part of the worid.

II. MATERIALS AND METHODS

A. Dataset Overview

The datasets used in this project was sourced from the PlantVillage database [4](https://plantvillage.org), which offers a diverse collection of plant leaf images useful for plant disease research. For our experimental setup, we curated a dataset comprising

100 distinct crop leaf images to support our analysis and model development. Fig1 is visualization of some of the dataset

Out of the 100 images, 15 depict healthy leaves from various crop species, offering a reference for non-diseased foliage. The remaining 85 images showcase different types of leaf diseases affecting crops, providing a comprehensive representation of common plant health issues. These diseased leaf images are categorized into several distinct classes based on the type of disease present.

Fig 1.

The primary disease classes included in our dataset are bacterial blight, alternaria alternate (a fungal disease), cercospora leaf spot, and anthracnose.

1. **Bacterial blight:** This is a disease caused by bacteria. It can affect the leaves and stems of plants, causing damage and making the plants less healthy.
2. **Alternaria alternate:** This is a fungal disease. The fungus can infect various parts of a plant, such as the leaves or fruits, leading to dark spots and potentially reducing the health and productivity of the plant.
3. **Cercospora leaf spot:** This is another fungal disease. It creates spots on the leaves of plants, which can lead to the leaves dying and falling off. This weakens the plant and can decrease crop yields.

4. **Anthracnose:** This is a disease caused by a different type of fungus. It can affect many parts of the plant, including leaves, stems, and fruits. Anthracnose can cause dark, sunken lesions on the plant, harming its health.

Each category represents a different threat to crop health and yields.

The diversity and balance of the dataset make it ideal for training machine learning models to accurately detect and classify various leaf diseases. This foundation will enable the development of robust and efficient disease detection systems for agricultural applications.

B. Data Pre-processing Methodologies

1. **Image Denoising:** Image denoising is a critical preprocessing step in the automated detection of plant diseases, aimed at improving the quality of leaf images by reducing digitization noise. The presence of noise can significantly affect the performance of subsequent analysis and classification tasks, as it may obscure important details in the images.

Build a two-dimensional index matrix Fij with element values of 0 or 1. The procedure for determining the value of index 0 or 1 in each index matrix Fij is as follows:

Create a $(2M \times 1)(2M \times 1)$ neighborhood region with Pij as the center for each pixel Pij at position (i, j) in the picture.

If Pij equals 0 or 255, then Pij is a corrupted pixel, and the value of Fij should be set to 1.

If Pij is between 0 and 255, then Pij could be an uncorrupted pixel. So, determine whether or not it is corrupted. To do so,

first translate the window into a 1-D vector S and then sort the elements of S in ascending order.

If $S(\lfloor \alpha/2 \rfloor + 1) \leq Pij \leq S(n - (\lfloor \alpha/2 \rfloor))$ where αan integer and $0 \leq \alpha \leq n$ then Pij is an uncorrupted pixel, and the value of Fij is set to 0. Otherwise, the pixel is mutated, and the value of Fij is set to 1.

Calculate the percentage of impulse noise forecast using the following formula:

$$\rho = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} Fij \times 100\%$$

Denoted:

ρ is the percentage of impulse noise prediction

Fij is index matrix 2D at coordinate (i, j)

M is the total of rows and N is the total of columns.

To address this challenge, a rank-order fuzzy (ROF) filter is applied to the plant images. Fig2 show visualization of Alternaria alternate plant leave of actual image and enhance contrast image(preprocesses)

The ROF filter is designed to enhance the clarity of leaf images by removing noise introduced during image acquisition or processing. This filtering technique operates by evaluating pixel values and their relationships within a defined neighborhood, then adjusting the pixel values based on a set of fuzzy rules.

Set the value of each noise element, where “salt or 255” by “0”. So that Pij only has one noise model that is “pepper or 0”. If Pij is an uncorrupted pixel then its value is left unchanged. Otherwise, follow the following step:

Select 2-D window of size(2M × 1) (2M × 1). Assume that the pixel being filtered isPij.

Matrix elements of the window should be sorted starting from the smallest value to the largest value. If the selected window contains all elements as 0’s then increase the window size by one and again check the increased window. If increased window contains all 0’s, then again increase window size by one. This process is repeated until we have a window with some element (except 0) on it or the maximum window size limit is reached.

Eliminate the 0 from the window and find the average and median of the remaining pixels. Suppose average and median is denoted by Aij and Mij.

Calculate first order absolute differences D′(i + k, j + l) by:

$$D'(i+k, j+l) = |p_{i+k,j+l} - P_{ij}| \text{ with k, l} \neq 0$$

Extract the local information Dij from Wij according to:

$$D_{ij} = max(D'(m))$$

Compute the fuzzy membership value μij based on the local information Dij

$$\mu_{ij} = \begin{cases} 0 & : D_{ij} < T1 \\ \frac{D_{ij} - T1}{T2 - T1} & : T1 \leq D_{ij} < T2 \\ 1 & : D_{ij} \geq T2 \end{cases}$$

Where, T1 and T2 are two predefined thresholds.

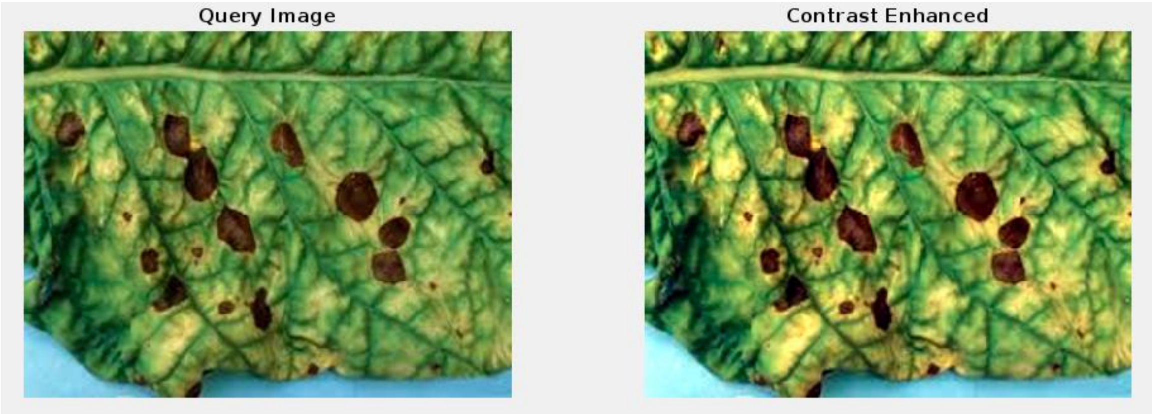

Fig2.

Improved image clarity plays a crucial role in facilitating more accurate disease spot segmentation and classification.

2. **Color Space Conversion:** Color Space Conversion is a crucial step in the process of segmenting and analyzing plant leaves for disease detection. The RGB images of plant leaves are converted to the Lab* color space, which is known for its ability to represent color information more effectively and intuitively. The Lab* color space consists of three distinct channels: L (lightness), a (green to red), and b (blue to yellow).

Integrate the adjusted H component, the unaltered S component, and the refined I component to generate the ultimate enhanced HSI color image. Subsequently, transform this enhanced HSI color map back into the RGB color map to derive the final enhanced image. Following this procedure, the conversion from HSI to RGB color space is executed to showcase the enhancement outcome. An inverse conversion algorithm can be articulated accordingly.

$$\textbf{\textit{if RG section}} \quad (0^{\circ} \leq H < 120^{\circ})$$

$$R = I\left[1 + \frac{ScosH}{cos(60^{\circ} - H)}\right] G = 1 - (R + G) B = I(1 - S)$$

$$\textbf{\textit{if GB section}} (120^{\circ} \leq H < 240^{\circ})$$

$$R = I(1 - S)$$

$$G = I\left[1 + \frac{ScosH}{cos(60^{\circ} - H)}\right] B = 1 - (R + G)$$

$$\textbf{\textit{if BR section}} (240^{\circ} \leq H \leq 360^{\circ})$$

$$H = H - 240^{\circ}$$

$$R = 1 - (G + B)$$

$$G = I(1 - S)$$

$$B = I\left[1 + \frac{ScosH}{cos(60^{\circ} - H)}\right]$$

By separating the color information into these channels, the Lab* color space conversion improves the accuracy of distinguishing between healthy and diseased areas of plant leaves. This refined representation aids in the subsequent steps of feature extraction and classification, ultimately enhancing the performance of automated plant disease detection systems.

C. Feature Engineering Techniques

Min-Max Hue Histogram: The Min-Max Hue Histogram is a powerful feature engineering technique employed in plant disease detection to improve the accuracy of segmentation and classification. This method involves computing a histogram of hue values for each leaf image, which represents the distribution of colors present in the image. Fig2 shows the Hue base detection flow chart in RGB images, the RGB image will be converted to HSI then convert back to RGB after processing.

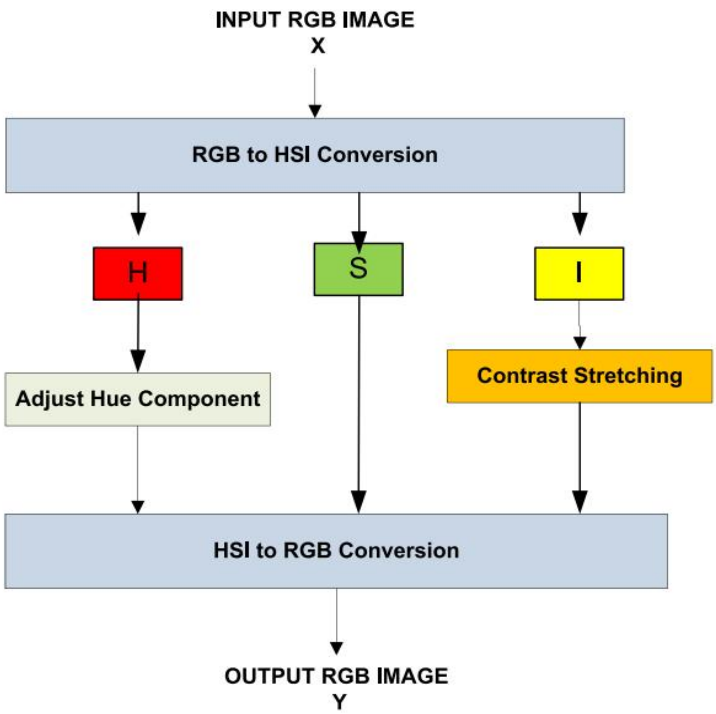

Fig2

The components of image X in RGB format are delineated as follows, with particular focus on the H component.

$$H = \theta, \text{ if } B \leq G$$
$$360\text{-}\theta, \text{ if } B > G$$

$$\theta = cos^{-1}\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{\left[(R-B)^2+(R-B)(G-B)^{1/2}\right]} \right\}$$

$$S = 1-\frac{3}{(R+G+B)}[MIN(R,G,B)]$$

$$I = \frac{1}{3}(R+G+B)$$

The range of intensity and hue are [0, 1] and [0,360], respectively
and The range of saturation is from [0,1].
By examining the range of hue values, particularly the minimum and maximum values within a predefined range, this technique helps to identify disease spots based on their unique color characteristics. Fig3 show the visualization of Alternaria alternate plant leave min-max hue histogram
The distinctive color variations associated with different plant diseases, such as yellowing, browning, or dark spots, can be effectively localized using the Min-Max Hue Histogram. This precise localization allows for the accurate isolation of disease-affected areas from the healthy regions of the leaf. As a result, the segmentation process becomes more effective, facilitating the analysis and classification of plant diseases.

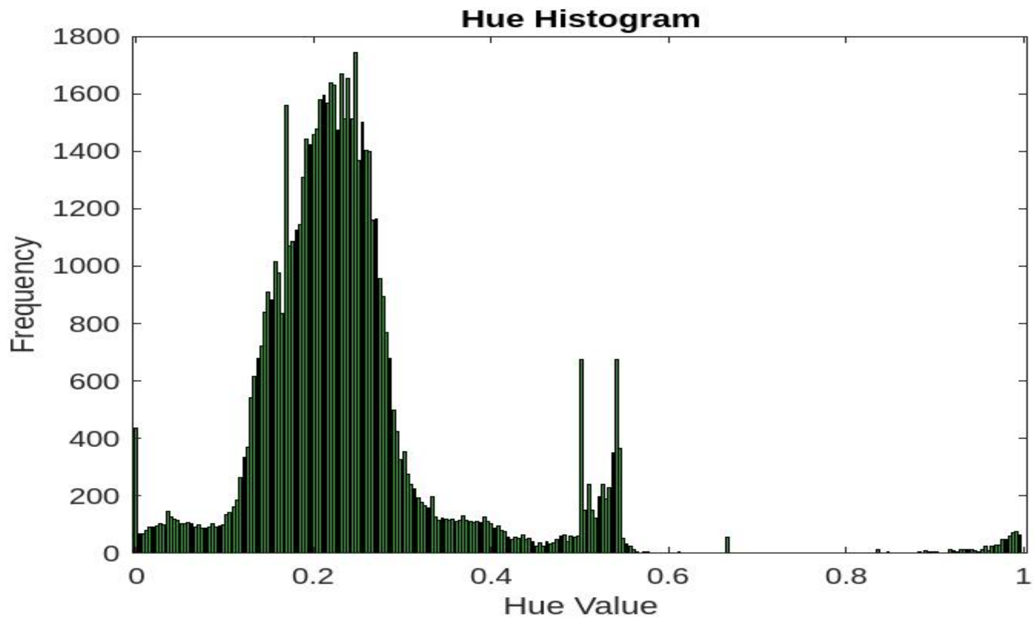

Fig3.

D. Machine Learning Algorithms

K-means Clustering: The K-means clustering algorithm segments plant leaves based on color similarity, efficiently identifying disease-affected areas. By grouping similar pixels into clusters, the algorithm partitions the leaf image into distinct regions corresponding to different leaf parts or lesion types. This segmentation isolates and analyzes areas with disease symptoms, as shown in Fig4 with Alternaria alternate plant leaf K-means segmentation.

The algorithm assigns pixels to predefined clusters, highlighting areas with similar characteristics for precise detection of potential disease spots. This targeted process enhances disease detection accuracy, providing insights for better crop health assessment and effective disease management strategies.

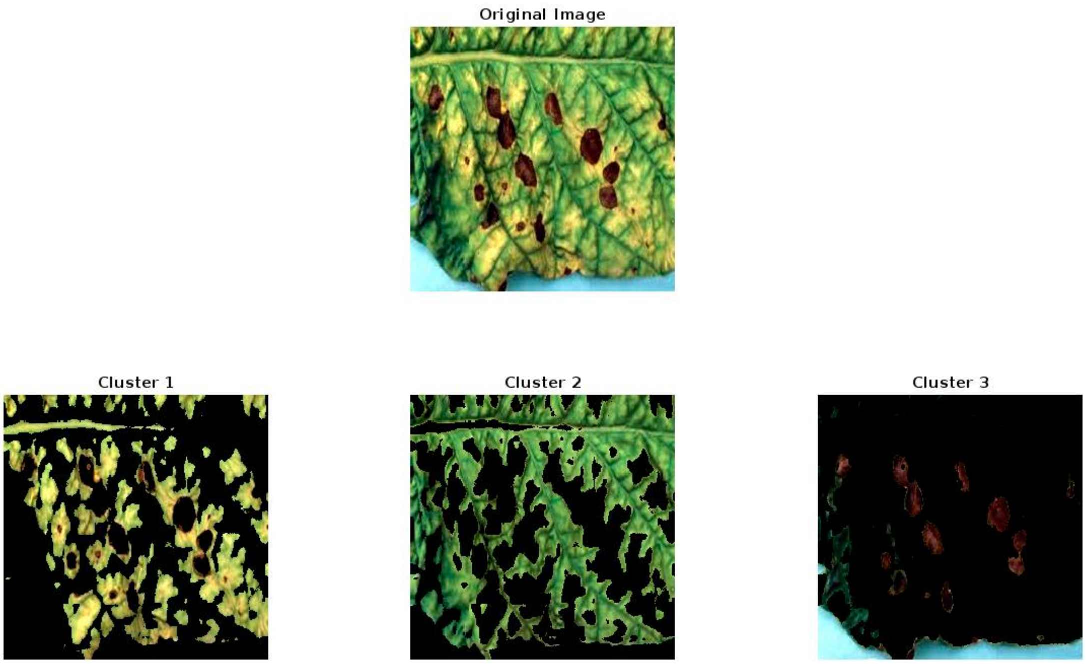

Fig4

Classification: Implementing a multi-class Support Vector Machine (SVM) is a powerful method for classifying the type of plant disease based on the extracted features. Multi-class SVMs are capable of handling the classification of multiple classes of diseases by constructing a set of binary SVM classifiers and leveraging the "one-vs-all" or "one-vs-one" strategy.

The one-vs-all approach involves training one SVM for each class, where the SVM classifies between that class and all other classes combined. On the other hand, the one-vs-one approach involves training one SVM for every possible pair of classes, and the final classification is determined based on majority voting or other aggregation techniques.

By using a multi-class SVM, the model can effectively classify different types of plant diseases such as Alternaria alternata, Anthracnose, Bacterial Blight, Cercospora leaf spot, and healthy leaves. The model takes in the features extracted from the leaf image (such as contrast, correlation, energy, and other textural and statistical properties) and predicts the type of disease affecting the plant.

Evaluation Metrics: Various evaluation metrics play a crucial role in assessing the performance of segmentation and detection algorithms in plant leaf disease detection. These metrics, including contrast, correlation, energy, homogeneity, mean, standard deviation, entropy, root mean square (RMS), variance, smoothness, kurtosis, and skewness, offer quantitative insights into the algorithms' ability to accurately identify disease spots and differentiate them from healthy leaf regions.

By evaluating these metrics, researchers can measure the precision and efficiency of the algorithms in processing leaf images and detecting abnormalities. For instance, contrast and homogeneity assess the sharpness and uniformity of the segmented regions, while metrics like entropy and variance capture the complexity and diversity within the image. Smoothness and kurtosis provide information on the distribution and behavior of pixel values, offering a comprehensive understanding of the segmentation quality. Overall, these metrics collectively determine the robustness and accuracy of the disease detection process. Fig5 is the Flow chart for the methodology.

This project employs a combination of data pre-processing methodologies, feature engineering techniques, and machine learning algorithms to automate the segmentation and detection of plant diseases in leaf images. By integrating these techniques into a cohesive framework, it will help to develop a robust and efficient system capable of accurately identifying and classifying a wide range of plant diseases, thereby enabling proactive interventions to mitigate crop losses and ensure food security.

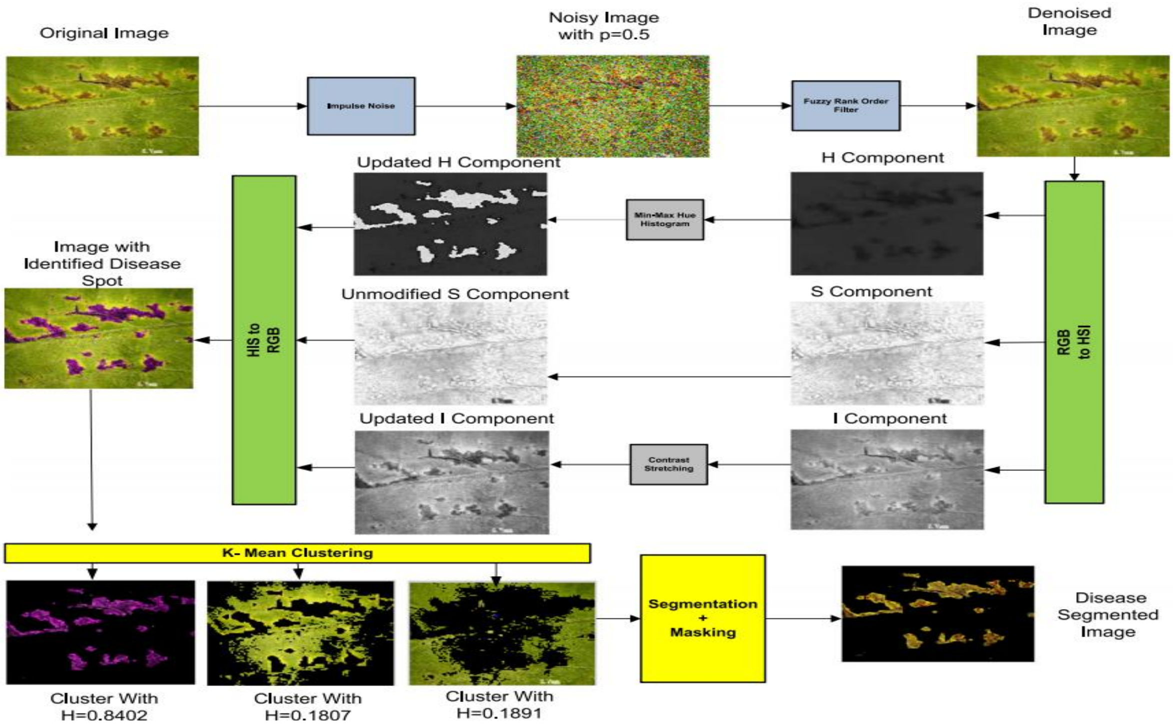

Fig5 Flow chart for the methodology.

III. RESULTS AND DISCUSSION

The project, "Automatic Segmentation of Plant Leaves for Disease Detection," presents a comprehensive MATLAB script for identifying and classifying leaf diseases. Utilizing advanced image processing, the script enhances contrast for better visibility and employs algorithms such as Otsu's method and K-means clustering to isolate diseased areas from healthy leaf tissue. After segmentation, the script extracts features like color, texture, and shape for disease classification. The multi-class SVM model classifies various plant diseases, offering a robust, efficient solution for automated detection, improving crop management and sustainability.

Affected Area

The algorithm found that 49.8002% of the leaf is affected by disease for the test image, indicating severe disease presence. This level of infection raises concerns about the plant's health and productivity.

Such extensive damage may lead to complications, including disease spread to other plants or parts of the same plant. Immediate action, such as targeted treatment or removal of affected areas, is necessary to prevent further spread. This level of damage can compromise photosynthesis and plant vitality. Prompt monitoring and management are crucial for maintaining plant health and successful crop yields.

Disease Classification

The disease has been classified as Alternaria alternata, a widespread fungal pathogen affecting various plant species. This classification informs effective treatment strategies to control the disease's spread and mitigate its impact on plant health.

Alternaria alternata causes leaf spot diseases, which can lead to defoliation and reduced photosynthesis. Airborne spores spread easily, necessitating prompt infection control. Recommended management strategies include fungicide application, pruning infected areas, and improving air circulation around plants.

Early identification and treatment are essential to contain the pathogen and protect crops. Monitoring and preventive measures like crop rotation and using resistant varieties can help manage future infection risks.

Model Performance

The model's performance with a Linear Kernel and 100 iterations yielded an accuracy of 91.9355%, demonstrating the model's robustness and effectiveness in distinguishing between healthy and diseased leaf tissue. This high accuracy is indicative of the model's ability to identify and classify various leaf diseases accurately, allowing for precise and reliable diagnosis.

Such a high level of performance supports the model's potential application in real-world scenarios, providing valuable insights for agricultural professionals in monitoring crop health. The model's success can be attributed to its ability to leverage relevant features extracted from leaf images, including texture, color, and shape characteristics.

Moreover, the model's strong performance suggests its adaptability to various plant species and diseases, making it a versatile tool in agricultural disease management. As a result, farmers and researchers can use this model to implement targeted treatments and preventative measures, ultimately contributing to sustainable agricultural practices and improved crop yields.

Feature Analysis

The extracted features from the diseased leaf areas provide a comprehensive understanding of the texture and color characteristics of the disease, offering insights into the overall condition and severity:

Contrast: The contrast value of 1.0227 highlights a distinct separation between the diseased and healthy areas of the leaf, demonstrating the model's ability to detect differences in visual quality.

Correlation: A correlation value of 0.8214 points to a consistent pattern in the texture of the diseased areas, which can aid in identifying specific types of diseases with distinctive texture characteristics.

Energy: The energy value of 0.2433 suggests a certain level of uniformity within the diseased regions. This measurement is useful for distinguishing different disease types based on their textural smoothness or coarseness.

Entropy: An entropy value of 4.6495 reveals the complexity of the leaf's texture, which is typical in diseased conditions. This complexity can arise from varying disease patterns and the heterogeneity of affected tissues.

Homogeneity: The homogeneity value of 0.8562 indicates that the diseased areas have a relatively uniform appearance, which can simplify the process of identifying diseased regions.

Kurtosis and Skewness: The kurtosis value of 2.2957 and skewness of 0.7777 provide insights into the distribution shape of grayscale values in the diseased areas, aiding in understanding the variation in disease impact.

Mean and RMS: The mean value of 47.7238 and RMS of 70.3130 reflect the average intensity and root mean square of the grayscale values, respectively. These measures help assess the overall brightness and consistency of diseased areas.

Smoothness: A smoothness value of 1.0000 suggests a lack of texture or noise in the segmented disease areas, aiding in the precise localization of affected regions.

Standard Deviation: The standard deviation value of 56.5511 quantifies the variability of grayscale values in the diseased areas. High variability may indicate areas with different disease stages or varying disease severity.

Fig6 shows the complete view of the Automatic Segmentation of Plant Leaves Disease Detection GUI(Graphic User Interface).

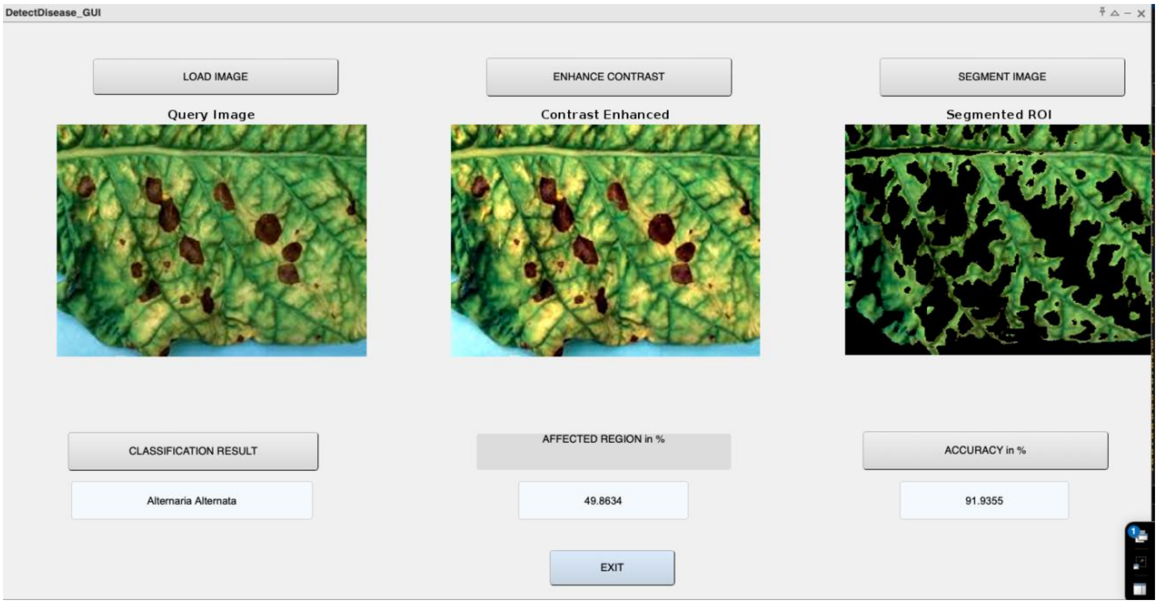

Fig6

IV. CONCLUSION

In conclusion the project demonstrates the potential of machine learning and image processing for early plant disease detection and classification. By identifying disease patterns at an early stage, the system enables timely intervention and may prevent disease spread. The model's high accuracy distinguishes healthy from diseased tissue using texture and color analysis, offering insights into various plant diseases.

The project's success supports future research, suggesting refinement with larger, diverse datasets for improved model generalizability and adaptability. Integrating the model into user-friendly tools for farmers could enable real-time monitoring and early diagnosis, promoting sustainable agricultural practices and food security worldwide.

V. REFERENCES

1. Archana KS, Sahayadhas A (2018) Automatic rice leaf disease segmentation using image processing
techniques. Int J Eng Technol 7:182–185

2. Trivedi, V. K., Shukla, P. K., & Pandey, A. (2022). Automatic segmentation of plant leaves disease using min-max hue histogram and k-mean clustering. Multimedia Tools and Applications, 81(20201-20228). https://doi.org/10.1007/s11042-022-12518-7

3. Dataset: (https://plantvillage.org)