Tomato Leaf Diseases Prediction and Segmentation Using K Means Clustering

[1]R.Mamallan, [2]P.Chithra,[3]D.Priyadharshini, [4]D.Shobana, [5]S.Subiksha

[1]Assistant Professor, Department of Electronics and Communication Engg., P.T.Lee Chengalvaraya Naicker College of Engineering and Technology, Kancheepuram, Tamilnadu, India.Email-mamallanktr@gmail.com
[2,3,4&5] Final Year Students, Department of Electronics and Communication Engg., P.T.Lee Chengalvaraya Naicker College of Engineering and Technology, Kancheepuram, Tamilnadu, India.

ABSTRACT

Tomato Plant diseases have a significant negative impact on agricultural production, leading to substantial economic losses if not detected and managed in time. Early detection is crucial for the effective prevention and control of leaf diseases, playing a vital role in agricultural decision-making and management. Tomato plants are particularly susceptible to various diseases caused by fungi, bacteria, viruses, and nematodes, which can severely weaken the plant and reduce crop yield. In this study, preprocessing techniques using different filters and segmentation using the K-Means Clustering algorithm are employed to enhance disease detection. The K-Means Clustering algorithm integrates both detection and segmentation tasks within a unified model, enabling more accurate and detailed disease identification. This approach improves the precision of disease classification, aiding in early diagnosis and effective management of tomato leaf diseases.

Key words: Segmentation, image processing, K Means Clustering and Disease.

INTRODUCTION

Image processing involves converting an image into a digital format and applying various operations to enhance its quality or extract valuable information. Image pre-processing techniques are employed to refine image quality before further application processing [1]. These techniques utilize a small pixel neighborhood in the input image to determine a new brightness value for the output image. Often referred to as filtration and resolution enhancement, these methods help mitigate image degradation caused by noise. To preserve the edges and contour details of agricultural plant images, efficient denoising and enhancement techniques are essential. Commonly used filters for image de-noising include Gaussian, Wiener, and Bilateral filters.

Image segmentation is a critical aspect of automatic image processing analysis [2]. It plays a significant role in various image processing applications by dividing an image into its essential components. This process isolates the Region of Interest (ROI), ensuring homogeneity in certain attributes. Segmentation algorithms are generally based on two fundamental intensity properties: (a) Similarity-Based and (b) Discontinuity-Based approaches. In the similarity-based approach, segmentation occurs by grouping pixels according to specific features, while the discontinuity-based approach partitions the image based on abrupt gray-level intensity changes. These methods include (a) Detection of Isolated Points, (b) Detection of Lines, and (c) Edge Detection [3]. The agricultural sector serves as the backbone of any developing economy. To maximize crop yields, it is imperative to equip farmers with the latest technologies and methodologies.

Tomato disease detection plays a crucial role in precision agriculture, ensuring early intervention to mitigate crop losses. Recent advancements in deep learning have significantly improved the accuracy and efficiency of disease identification through preprocessing and segmentation techniques. Various convolutional neural networks (CNNs) and deep learning models have been deployed to enhance image-based disease detection.

An efficient altered Mask Region Convolutional Neural Network (Mask R-CNN) was presented by Kaur et al. [14] for the autonomous segmentation and identification of tomato leaf disease. Their approach achieved an impressive F1-score of 91.20% and an accuracy of 95.00%, demonstrating the model's effectiveness in disease detection. Similarly, Qi et al. [15] introduced an improved YOLOv5 model integrated with a squeeze-and-excitation (SE) module to extract crucial features, yielding an accuracy of 91.07%.

Deep learning-based segmentation methods have also been utilized for disease quantification in tomato crops. Divyanth et al. [8] developed a two-stage deep-learning segmentation model specifically designed for crop disease quantification based on field imagery, enhancing the precision of disease assessment. Furthermore, Karthik et al. [16] introduced an attention-embedded residual CNN for tomato leaf disease detection, leveraging attention mechanisms to refine feature extraction and classification accuracy.

Transfer learning has also played a pivotal role in optimizing tomato disease segmentation. Kaur et al. [22] implemented a novel transfer deep learning method for plant disease detection, demonstrating significant improvements in classification efficiency. Additionally, Paymode and Malode [24] explored the application of pre-trained CNN architectures such as VGG for multi-crop leaf disease classification, proving the effectiveness of transfer learning in agricultural diagnostics.

Moreover, several studies have explored hybrid expert systems for post-harvest disease management. Sottocornola et al. [9] developed DSSApple, a hybrid system for diagnosing post-harvest diseases in apples, which could be adapted for tomato disease detection. Similarly, Hughes and Salathé [28] created an open-access plant disease image repository to facilitate mobile disease diagnostics, emphasizing the importance of large-scale datasets in improving segmentation accuracy.

The integration of lightweight networks has also been explored to enhance computational efficiency. Qu and Sun [11] designed a lightweight network for mummy berry disease recognition, which could be adapted for real-time tomato disease detection in resource-constrained environments. Likewise, Alruwaili et al. [20] proposed the RTF-RCNN architecture for real-time tomato leaf disease detection in video streaming applications, ensuring rapid and accurate disease classification.

Overall, preprocessing and segmentation techniques play a fundamental role in improving the accuracy of tomato disease detection models. With the integration of CNNs, attention mechanisms, transfer learning, and hybrid expert systems, these approaches continue to evolve, ensuring robust and scalable solutions for precision agriculture. In this paper, section 1 discusses about the introduction of the proposed work. The major diseases and pests which are affecting tomato leaf are discussed in section 2 and 3 respectively. Section 4 elaborates the implementation of K Means Clustering algorithm for segmentation of tomato leaf images. The results and discussion of the proposed work is discussed in section 5. Section 6 gives the conclusion of the proposed work.

2. TOMATO PLANTS ARE VULNERABLE TO VARIOUS DISEASES CAUSED BY FUNGI, BACTERIA, VIRUSES

2.1. Fungal Diseases

a) Early Blight (Alternaria solani)

Early blight is a fungal disease caused by Alternaria solani, which primarily affects tomato plants in warm and humid conditions. The pathogen spreads through wind, rain, contaminated soil, and infected plant debris, making it a persistent threat in fields with poor crop rotation. The first symptoms appear as small, dark brown or black spots on older leaves, gradually enlarging to form characteristic concentric rings, often referred to as "target spots." As the disease progresses, affected leaves turn yellow and drop prematurely, leading to reduced photosynthesis and weakened plant growth. In severe cases, the fungus can also infect stems and fruits, causing sunken, dark lesions. If left untreated, early blight significantly reduces tomato yield and quality, making effective disease management crucial.

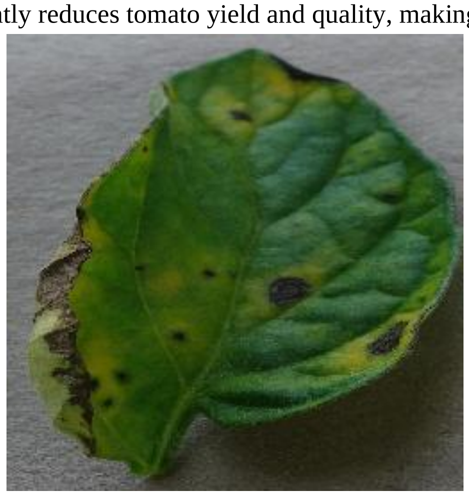

Fig.1. Early_blight

Fig.2. Late Blight

b) Late Blight (Phytophthora infestans)

Late blight, caused by Phytophthora infestans, is one of the most destructive diseases affecting tomato plants. This pathogen, classified as an **oomycete** (water mold), thrives in cool and wet conditions, spreading rapidly through windborne spores and water splashes. Symptoms first appear as large, irregular, water-soaked lesions on leaves, which quickly turn brown and develop a white, moldy growth on the undersides during humid conditions. The infection extends to stems, causing blackened areas and plant collapse. Infected fruits develop dark, sunken lesions, making them unmarketable. Late blight spreads aggressively under favorable conditions, often leading to total crop loss if not managed promptly through fungicides, resistant varieties, and proper field hygiene.

2.2 Bacterial Diseases:

c) Bacterial Spot (Xanthomonas campestris pv. vesicatoria)

Bacterial spot is a serious disease of tomato plants caused by Xanthomonas campestris pv. vesicatoria. It thrives in warm, humid environments and spreads through contaminated seeds, plant debris, and water splashes. The first symptoms appear as small, dark brown to black spots on leaves, often surrounded by a yellow halo. As the infection progresses, the spots enlarge, merge, and cause leaf blight, leading to defoliation. Severe leaf loss weakens the plant and exposes fruits to sunscald. Infected fruits develop raised, rough, and scabby lesions, reducing their marketability. Since bacterial spot spreads through wind-driven rain and overhead irrigation, controlling moisture levels and using copper-based bactericides can help manage the disease.

Fig.3. Bacterial Spot

Fig.4. Bacterial Speck

d) Bacterial Speck (Pseudomonas syringae pv. tomato)

Bacterial speck is a foliar disease caused by Pseudomonas syringae pv. tomato, which spreads through infected seeds, plant debris, and water splashes during cool, wet conditions. Symptoms begin as small, dark, water-soaked spots on leaves, which later turn black and may be surrounded by a yellow halo. Unlike bacterial spot, bacterial speck lesions are smaller and more superficial. The infection can spread to fruits, causing small, black specks on the surface, which do not penetrate deeply but still reduce fruit quality and market value. The disease can be managed through the use of disease-free seeds, crop rotation, and copper-based sprays to limit bacterial spread.

2.3 Viral Diseases:

e) Tomato Mosaic Virus (ToMV)

Tomato mosaic virus (ToMV) is a highly contagious viral disease that affects tomato plants, causing significant yield loss. The virus spreads through contaminated seeds, plant debris, human handling, and insect vectors such as aphids. It can survive on surfaces like tools, hands, and clothing, making transmission easy in greenhouses and fields. Symptoms include light and dark green mottling (mosaic patterns) on leaves, leaf curling, and stunted growth. Infected plants often show reduced flowering and fruit deformation, leading to poor-quality tomatoes. Unlike other viral diseases, *ToMV* does not typically kill the plant but significantly impacts production. Since no cure exists, prevention through resistant varieties, proper sanitation, and insect control is essential.

Fig.5. Tomato_mosaic_virus

Fig.6. Tomato_Yellow_Leaf_Curl_Virus

f) Tomato Yellow Leaf Curl Virus (*TYLCV*)

Tomato yellow leaf curl virus (TYLCV) is a devastating viral disease transmitted by the whitefly (Bemisia tabaci). The virus spreads rapidly in warm climates, particularly in regions with high whitefly populations. Infected plants exhibit severe leaf curling, yellowing, and stunted growth, with leaves becoming thick and brittle. The disease drastically reduces fruit production, as infected plants produce fewer flowers and small, malformed fruits. TYLCV can spread quickly across fields, making early detection and whitefly control crucial. Management strategies include planting resistant varieties, using insect-proof nets, and applying biological or chemical controls to reduce whitefly populations.

3.PREPROCESSING UNSING BILATERAL FILTER

Tomato plants are highly vulnerable to various diseases caused by fungi, bacteria, and viruses, leading to significant losses in yield and quality. The early detection of these diseases is crucial for effective management. In digital image processing, preprocessing plays a vital role in enhancing image quality for accurate disease identification. One of the most effective techniques for noise reduction while preserving important features is the bilateral filter. This filter smooths the image while maintaining the edges, ensuring that crucial details such as disease spots, discoloration, and lesion boundaries remain intact.

The bilateral filter is a non-linear, edge-preserving technique that considers both spatial and intensity differences in an image. Unlike conventional smoothing filters such as Gaussian or mean filters, which tend to blur edges, the bilateral filter retains fine details of infected regions while reducing noise. This makes it highly suitable for preprocessing leaf images, where disease symptoms such as yellowing, necrotic spots, and curling need to be accurately preserved for subsequent analysis. The bilateral filter enhances image clarity by reducing unwanted variations caused by lighting conditions, sensor noise, and background interference, thus improving the segmentation and classification accuracy of diseased areas.

After applying the bilateral filter, the preprocessed images can be further analyzed using segmentation techniques such as K-Means clustering. The enhanced image quality ensures that segmentation algorithms effectively distinguish diseased regions from healthy areas, facilitating precise disease identification. By integrating bilateral filtering as a preprocessing step, the overall efficiency of image-based disease detection models can be significantly improved, leading to better agricultural decision-making and disease management strategies.

4. SEGMENTATION USING K MEANS CLUSTERING ALGORITHM

Tomato plants are highly susceptible to various diseases caused by fungi, bacteria, and viruses, making early detection essential for effective disease management. Image segmentation is a crucial step in processing diseased leaf images, as it helps in identifying and isolating infected regions from healthy areas. One of the most widely used unsupervised clustering techniques for segmentation is K-Means Clustering. This algorithm groups pixels based on color similarity, allowing diseased regions to be accurately distinguished. By applying K-Means clustering, infected areas exhibiting discoloration, lesions, or necrotic spots can be effectively segmented from the background and other non-relevant parts of the image.

The K-Means clustering algorithm works by dividing an image into K clusters, where K represents the number of color-based groups. The algorithm assigns each pixel to the nearest cluster center based on color similarity and iteratively refines these clusters until optimal segmentation is achieved. In tomato disease detection, setting an appropriate K value is crucial to ensure that infected regions are accurately highlighted. For instance, choosing K=3 or K=5 typically provides a balanced segmentation by differentiating between healthy tissue, diseased areas, and background elements. Preprocessing techniques like bilateral filtering before segmentation help in reducing noise and improving the accuracy of cluster formation.

Once the K-Means clustering process is completed, the segmented image can be analyzed for further classification or feature extraction. The segmented output helps in identifying disease-affected areas more precisely, enabling automated disease recognition models to classify different types of infections such as Early Blight, Late Blight, Bacterial Spot, and Tomato Yellow Leaf Curl Virus. The integration of K-Means clustering in disease segmentation enhances the accuracy of diagnosis, supports precision agriculture, and enables early intervention to prevent widespread crop damage. This method plays a vital role in computer-aided plant disease detection systems, improving the efficiency of agricultural disease monitoring.

5 RESULTS AND DISCUSSION

The sample images depicted in Figures 7(a), 8(a), 9(a), 10(a), 11(a), 12(a), and 13(a) represent different tomato leaf conditions, including Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, Tomato Yellow Leaf Curl Virus, and healthy leaves. These images were initially processed using a Bilateral filter to reduce noise and enhance clarity. The pre-processed images, illustrated in Figures 7(b), 8(b), 9(b), 10(b), 11(b), 12(b), and 13(b), were subsequently fed into segmentation algorithms such as K-Means Clustering. The segmentation outcomes, presented in Figures 7(c), 8(c), 9(c), 10(c), 11(c), 12(c), and 13(c), reveal that K-Means Clustering effectively segmented relevant regions across all categories. This technique successfully delineated diseased regions, including those affected by Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, and Tomato Yellow Leaf Curl Virus, ensuring accurate disease identification by isolating significant areas of interest without distortion.

Table 1. Classification Accuracy

S.No	Disease	Samples taken	K Means
1	Bacterial spot	100	95
2	Early blight	100	96
3	Late blight	100	97
4	Leaf Mold	100	95
5	Septoria Leaf Spot	100	96
6	Tomato Yellow Leaf Curl Virus leaf	100	95
	Total	600	574

Fig.7(a) Sample Image 1 (Bacterial spot)

Fig.7(b) Preprocessed output using Bilateral filter

Fig.7(C) Segmented output using K Means Clustering

Fig.8(a) Sample Image 2 (Early blight)

Fig.8(b) Preprocessed output Bilateral filter

Fig.8(C) Segmented output using K Means Clustering

Fig.9(a) Sample Image 3 (Late blight)

Fig.9(b) Preprocessed output Bilateral filter

Fig.9(C) Segmented output using K Means Clustering

Fig.10(a) Sample Image 3 (Leaf Mold)

Fig.10(b) Preprocessed output Bilateral filter

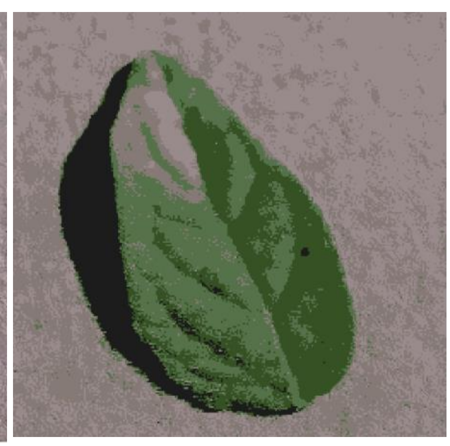

Fig.10(C) Segmented output using K Means Clustering

Fig.11(a) Sample Image 4 (Septoria Leaf Spot)

Fig.11(b) Preprocessed output using Bilateral filter

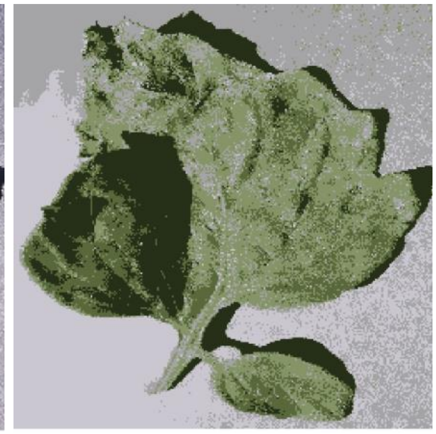

Fig.11(C) Segmented output using K Means Clustering

Fig.12(a) Sample Image 5 (Yellow Leaf Curl Virus leaf)

Fig.12(b) Preprocessed output using Bilateral filter

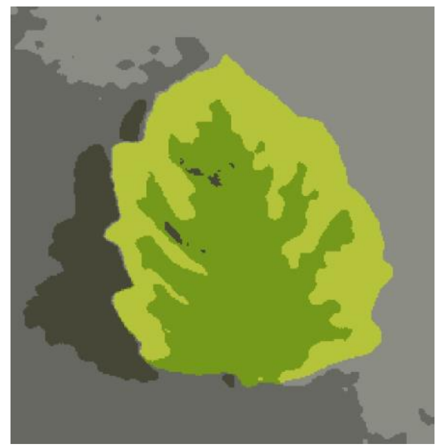

Fig.12(C) Segmented output using K Means Clustering

Fig.13(a) Sample Image 6 (Healthy Image)

Fig.13(b) Preprocessed output using Bilateral filter

Fig.13(C) Segmented output using K Means Clustering

6 CONCLUSION

This paper discussed various tomato diseases caused by fungi, bacteria, and viruses that affect tomato leaves. The tomato leaf dataset was collected, pre-processed, and segmented to facilitate accurate disease detection. The preprocessing phase employed a bilateral filter for noise reduction, while the segmentation process was carried out using K-Means Clustering. Experimental results demonstrate the effectiveness of the proposed system, achieving an impressive accuracy of 95.66%. Future research could aim to enhance segmentation precision by incorporating deep learning-based refinements and investigating hybrid models for improved disease identification and classification.

REFERENCE PAPER

1. K.A. Awan, I. Ud Din, A. Almogren, H. Almajed, Agritrust—a trust management approach for smart agriculture in cloud-based internet of agriculture things, Sensors 20 (21) (2020) 6174.
2. A. Gomez-Zavaglia, J.C. Mejuto, J. Simal-Gandara, Mitigation of emerging implications of climate change on food production systems, Food Res. Int. 134 (2020) 109256.
3. S. Dasgupta, E.J. Robinson, Attributing changes in food insecurity to a changing climate, Sci. Rep. 12 (1) (2022) 1–11.
4. J.A. Wani, S. Sharma, M. Muzamil, S. Ahmed, S. Sharma, S. Singh, Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: methodologies, applications, and challenges, Arch. Comput. Methods Eng. 29 (1) (2022) 641–677.
5. M. Brahimi, K. Boukhalfa, A. Moussaoui, Deep learning for tomato diseases: classification and symptoms visualization, Appl. Artif. Intell. 31 (4) (2017) 299–315.
6. A. Venkataraju, D. Arumugam, C. Stepan, R. Kiran, T. Peters, A review of machine learning techniques for identifying weeds in corn, Smart Agric. Technol. (2022) 100102.
7. A. Ahmad, D. Saraswat, A. El Gamal, A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools, Smart Agric. Technol. (2022) 100083.
8. L. Divyanth, A. Ahmad, D. Saraswat, A two-stage deep-learning based segmentation model for crop diseasequantification based on corn field imagery, Smart Agric. Technol. 3 (2023) 100108.
9. G. Sottocornola, S. Baric, M. Nocker, F. Stella, M. Zanker, Dssapple: a hybrid expert system for the diagnosis of post-harvest diseases of apple, Smart Agric. Technol. 3 (2023) 100070.
10. I. Grijalva, B.J. Spiesman, B. McCornack, Image classification of sugarcane aphid density using deep convolutional neural networks, Smart Agric. Technol. 3 (2023) 100089.
11. H. Qu, M. Sun, A lightweight network for mummy berry disease recognition, Smart Agric. Technol. 2 (2022) 100044.
12. A. Kamilaris, F.X. Prenafeta-Boldú, A review of the use of convolutional neural networks in agriculture, J. Agric. Sci. 156 (3) (2018) 312–322.
13. M. Kaur, R. Bhatia, Development of an improved tomato leaf disease detection and classification method, in: 2019 IEEE Conference on Information and Communication Technology, IEEE, 2019, pp. 1–5.
14. P. Kaur, S. Harnal, V. Gautam, M.P. Singh, S.P. Singh, An approach for characterization of infected area in tomato leaf disease based on deep learning and object detection technique, Eng. Appl. Artif. Intell. 115 (2022) 105210.
15. J. Qi, X. Liu, K. Liu, F. Xu, H. Guo, X. Tian, M. Li, Z. Bao, Y. Li, An improved yolov5 model based on visual attention mechanism: application to recognition of tomato virus disease, Comput. Electron. Agric. 194 (2022) 106780.
16. R. Karthik, M. Hariharan, S. Anand, P. Mathikshara, A. Johnson, R. Menaka, Attention embedded residual cnn for disease detection in tomato leaves, Appl. Soft Comput. 86 (2020) 105933.
17. N. Aishwarya, N. Praveena, S. Priyanka, J. Pramod, Smart farming for detection and identification of tomato plant diseases using light weight deep neural network, Multimed. Tools Appl. (2022) 1–12.
18. M. Moussafir, H. Chaibi, R. Saadane, A. Chehri, A.E. Rharras, G. Jeon, Design of efficient techniques for tomato leaf disease detection using genetic algorithm-based and deep neural networks, Plant Soil 479 (1) (2022) 251–266.
19. T. Anandhakrishnan, S. Jaisakthi, Deep convolutional neural networks for image based tomato leaf disease detection, Sustain. Chem. Pharm. 30 (2022) 100793.
20. M. Alruwaili, M.H. Siddiqi, A. Khan, M. Azad, A. Khan, S. Alanazi, Rtf-rcnn: an architecture for real-time tomato plant leaf diseases detection in video streaming using faster-rcnn, Bioengineering 9 (10) (2022) 565.
21. G. Sakkarvarthi, G.W. Sathianesan, V.S. Murugan, A.J. Reddy, P. Jayagopal, M. Elsisi, Detection and classification of tomato crop disease using convolutional neural network, Electronics 11 (21) (2022) 3618.
22. P. Kaur, S. Harnal, V. Gautam, M.P. Singh, S.P. Singh, A novel transfer deep learning method for detection and classification of plant leaf disease, J. Ambient Intell. Humaniz. Comput. (2022) 1–18.
23. A.K. Rangarajan, R. Purushothaman, A. Ramesh, Tomato crop disease classification using pre-trained deep learning algorithm, Proc. Comput. Sci. 133 (2018) 1040–1047.

24. A.S. Paymode, V.B. Malode, Transfer learning for multi-crop leaf disease image classification using convolutional neural network vgg, Artif. Intell. Agric. 6 (2022) 23–33.
25. M. Agarwal, S.K. Gupta, K. Biswas, Development of efficient cnn model for tomato crop disease identification, Sustain. Comput. Informatics Syst. 28 (2020) 100407.
26. P. Ujawe, S. Nirkhi, Comparative study of tomato crop disease detection system using deep learning techniques, in: Intelligent Communication Technologies and Virtual Mobile Networks, Springer, 2023, pp. 493–499.
27. H.M. Chong, X.Y. Yap, K.S. Chia, Effects of different pretrained deep learning algorithms as feature extractor in tomato plant health classification, Pattern Recognit. Image Anal. 33 (1) (2023) 39–46.
28. D. Hughes, M. Salathé, et al., An open access repository of images on plant health to enable the development of mobile disease diagnostics, arXiv preprint, arXiv :1511 .
08060, 2015.Sann,
29. R. Mamallan and E. Sivaraman "Coconut tree disease segmentation using Fuzzy Rough C-Means clustering algorithm" *Journal of Xidian University,* VOLUME 16, ISSUE 5, 2022
30. R. Mamallan and A.Sujitha "Coconut tree disease segmentation using YOLOv8 (You Only Look Once version 8)" *Journal of Xidian University,* VOLUME 19 – ISSUE 2 – February 25,Page No: 55 – 62