Open Access

© 2023, Centre for Advanced Research in Agricultural Sciences
Research Journal of Agricultural Sciences
Volume 14; Issue 05 (Sep-Oct 2023); pp 1194–1198

ISSN: 0976-1675 (P)
ISSN: 2249-4538 (E)

Full Length Research Article

An Integrated Approach for Tomato Disease Detection using K-means Clustering, HSV Color Analysis, and Cubic SVM Classifier

T. Hemalatha*[1] and S. Piramu Kailasam[2]

[1] Research Department of Botany, Rani Anna Government College for Women, Tirunelveli - 627 008, Tamil Nadu, India
[2] Department of Computer Applications Sadakathullah Appa College, Tirunelveli - 627 008, Tamil Nadu, India

Abstract

The early detection and accurate diagnosis of plant diseases are crucial for sustainable agriculture and food security. In this study, we propose a novel approach for detecting diseases in *Solanum lycopersicum* L. (tomato) fruits using a combination of K-means clustering, hue saturation value (HSV) color representation, and a Cubic Support Vector Machine (SVM) classifier. The methodology involves segmenting disease-affected regions in tomato fruit images through K-means clustering and HSV color representation. This process aids in isolating regions of interest, where disease symptoms are prevalent. Subsequently, features are extracted from these segmented regions, and a Cubic SVM classifier is trained to accurately classify different types of tomato fruit diseases. To evaluate the performance of the proposed method, we conducted extensive experiments on a dataset comprising various tomato fruit diseases. The results demonstrate the effectiveness of the combined approach, showcasing significantly improved accuracy and efficiency compared to traditional methods. The implications of this research are far-reaching, as the automated and reliable disease detection system presented here can potentially revolutionize agricultural practices. By enabling timely interventions, such as targeted treatments and optimized crop management, the proposed method may lead to increased crop yields and reduced economic losses. Over all, the work highlights the potential of leveraging advanced machine learning techniques in agriculture and sets the stage for future research and implementation of similar approaches to enhance crop protection and global food production.

Key words: Image acquisition, Preprocessing, Kmeans clustering, HSV color thresholder, Cubic SVM classifier

India is currently the world's second largest producer of several dry fruits, agriculture-based textile raw materials, roots and tuber crops, pulses, farmed fish, eggs, coconut, sugarcane and numerous vegetables. Two-thirds of its population is engaged in agricultural activities. Agriculture is a primary activity, which produces most of the food that we consume, besides food grains, it also produces raw material for various industries. To promote sustainable agriculture and food security early detection and accurate diagnosis of fruit disease play a pivotal role. In the present study a novel approach is proposed to detect tomato diseases. A data set comprising of various tomato fruit [7] diseases was conducted to evaluate the performance of the proposed method. Potentially revolutionized sustainable agricultural practices could be enabled by the automated and reliable disease detection system performance. The proposed method lead to increase crop yields and reduce economic loss in agricultural practices. When timely interventions such as targeted treatment and optimized crop management are enabled.

However, advancements in computer vision and machine learning have opened up new possibilities for automated and efficient disease detection systems. Among these methods, the combination of K-means clustering, hue saturation value (HSV), and multiclass Support Vector Machine (SVM) classifiers has shown promise in various plant disease detection tasks. This research explores the potential of utilizing the K-means clustering algorithm and HSV color representation to segment regions of interest in tomato fruit images, focusing on areas affected by disease symptoms. By identifying these regions, features can be extracted and used to train a Cubic SVM classifier, enabling accurate classification of various tomato fruit diseases. The main objectives of this study are to develop a robust and automated system for the early detection of *Solanum lycopersicum* L. fruit diseases and to evaluate the effectiveness of the proposed approach compared to traditional methods. The integration of K-means clustering and HSV color representation aims to improve the accuracy and efficiency of disease region segmentation, while the Cubic

Received: 28 May 2023; Revised accepted: 12 Aug 2023; Published online: 09 Sep 2023

Correspondence to: T. Hemalatha, Research Department of Botany, Rani Anna Government College for Women, Tirunelveli - 627 008, Tamil Nadu, India; E-mail: spkpramu@gmail.com

Citation: Hemalatha T, Kailasam SP. 2023. An integrated approach for tomato disease detection using K-means clustering, HSV color analysis, and cubic SVM classifier. *Res. Jr. Agril. Sci.* 14(5): 1194-1198.

SVM classifier seeks to provide a powerful tool for accurate disease classification.

The outcomes of this research have the potential to revolutionize agricultural practices by enabling timely disease detection, leading to prompt interventions such as targeted treatments and crop management techniques. Ultimately, this could result in increased yields, reduced economic losses, and enhanced food security for a growing global population. In the following sections, we delve into the methodology, experimental setup, and results obtained from the application of K-means clustering, HSV color representation, and the Cubic SVM classifier to detect *Solanum lycopersicum* L. fruit diseases. The insights gained from this study pave the way for advanced applications of machine learning in agriculture, driving the industry towards a more sustainable and resilient future.

MATERIALS AND METHODS

Dataset collection: For this study, a diverse dataset of *Solanum lycopersicum* L. (tomato) fruit images was collected. The dataset consisted of images containing different types and stages of fruit diseases, along with healthy tomato fruits, to ensure a comprehensive representation of various disease classes.

Preprocessing: Prior to applying the proposed method, the tomato fruit images underwent preprocessing steps to enhance the quality and standardize the data. Preprocessing included resizing the images to a uniform resolution, normalization of pixel values, and noise reduction to improve the robustness of the subsequent analysis.

Fig 1 *Solanum lycopersicum* L. disease fruit (Input images)

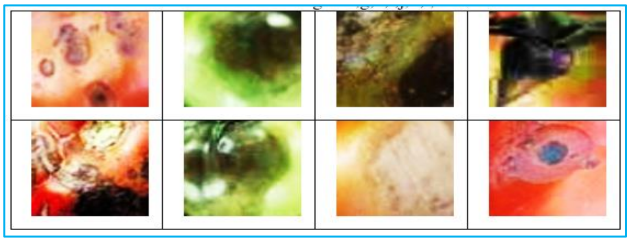

Fig 2 Region of interest of tomato

The choice of the best segmentation [16] method for tomato disease detection depends on various factors, including the characteristics of the dataset, the complexity of the diseases to be detected, and the specific requirements of the application. Here are some popular segmentation methods that have shown promising results for tomato disease detection:

1. *K-means Clustering*: K-means clustering [17] is a simple and effective unsupervised segmentation technique. It partitions the image into clusters based on color similarity, which can be useful for segmenting regions of interest in tomato fruit images with distinct color differences due to diseases.

2. *Watershed segmentation*: Watershed segmentation is a region-based method that can be applied to segment objects in an image. It works by treating the pixel intensities as a topographic surface and "flooding" from regional minima to separate regions of interest. Watershed segmentation can be effective when dealing with irregular shapes and overlapping regions in tomato disease images.

3. *Thresholding*: Thresholding is a basic technique used to convert grayscale or color images into binary images based on pixel intensity values. This method can be useful when disease symptoms exhibit significant contrast with the background. Adaptive thresholding methods can adapt to local image characteristics and improve segmentation accuracy.

4. *Region growing*: Region growing is a region-based segmentation method that starts with a seed pixel and grows the region by adding neighboring pixels with similar properties. It is effective for segmenting connected regions in images and can be suitable for isolating disease-affected areas in tomatoes.

5. *Graph-based segmentation*: Graph-based segmentation methods use graph theory to model image pixels as nodes and edges representing pixel similarities. By partitioning the graph, objects in the image can be segmented. These methods are robust to noise and can handle complex image structures.

6. *Convolutional neural networks (CNN) segmentation*: Deep learning-based segmentation methods [6], [14-15], particularly CNNs, have shown remarkable success in various image segmentation tasks. By training a CNN on annotated data, it can learn to accurately segment disease-affected regions in tomato fruit images.

7. *U-Net*: U-Net is a type of CNN architecture specifically designed for biomedical image segmentation. Its unique architecture incorporates both contracting and expanding pathways, making it well-suited for accurate segmentation of disease regions in tomatoes.

It's essential to evaluate and compare the performance of different segmentation methods on the specific dataset to determine the most suitable one for tomato disease detection. The best method may vary depending on the complexity of the diseases, the size and quality of the dataset, and the computational resources available. Additionally, combining multiple segmentation methods or using hybrid approaches can also enhance segmentation accuracy and disease detection results.

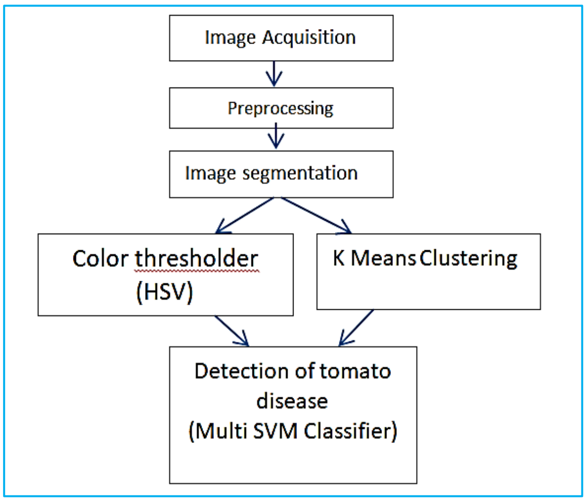

Fig 3 Flow diagram

HSV color segmentation

HSV (Hue, Saturation, Value) color segmentation is a popular method for detecting tomato diseases based on color differences in images. It leverages the HSV color space, which separates color information into three components: Hue (the type of color), Saturation (the intensity or purity of the color), and Value (the brightness).

The steps involved in HSV color segmentation for tomato disease detection are as follows:

1. *Preprocessing*: The input tomato fruit image is preprocessed to enhance its quality and remove noise. Common preprocessing steps include resizing the image to a standard size, normalization of pixel values, and noise reduction.

2. *Conversion to HSV color space*: The preprocessed image is then converted from the RGB color space to the HSV color space. In the HSV color space, the hue value represents the dominant color of the pixel, while the saturation value represents the intensity of the color, and the value component represents the brightness.

3. *Thresholding based on HSV ranges*: The key step in HSV color segmentation is to set appropriate thresholds for the hue, saturation, and value components to identify the regions of interest (disease-affected areas). The thresholds are defined based on the specific color characteristics of the tomato diseases being targeted. For example, certain diseases may cause specific color changes (e.g., yellowing or browning) in the tomato fruit, and appropriate threshold ranges can be set to capture these changes.

4. *Creation of a binary mask*: By applying the threshold values to the HSV components, a binary mask is created, where the disease-affected regions are represented by white pixels and the healthy regions by black pixels.

5. *Post-processing (Optional)*: Depending on the quality of the binary mask, post-processing steps such as morphological operations (e.g., erosion, dilation) or region filtering can be applied to refine the segmentation results and remove any small artifacts or noise.

6. *Region of interest (ROI) extraction*: Once the binary mask is generated and refined, it is used to extract the regions of interest (ROIs) from the original image. These ROIs correspond to the disease-affected areas in the tomato fruit.

7. *Feature extraction and disease classification*: After obtaining the ROIs, relevant features can be extracted to characterize the disease symptoms. These features may include texture descriptors, shape attributes, or color-based metrics. The extracted features can then be used to classify the type of disease present using machine learning algorithms like SVM or CNNs.

HSV color segmentation is advantageous for tomato disease detection as it can effectively capture specific color changes associated with different diseases. However, its success depends on the distinctive color variations caused by the diseases and may require fine-tuning of threshold values for accurate results. Integrating HSV color segmentation with other segmentation techniques or advanced machine learning methods can further enhance the overall accuracy and robustness of the tomato disease detection system.

Solanum lycopersicum L., commonly known as the tomato plant, is an economically significant crop worldwide. However, it is susceptible to various diseases that can cause substantial yield losses and quality deterioration. Early detection and accurate diagnosis of these diseases are essential for effective disease management and sustainable agriculture. In recent years, the integration of machine learning techniques, such as K-means clustering [1], HSV analysis, and multiclass Support Vector Machine (SVM) classifiers, has emerged as a promising approach for automated and reliable tomato disease [2], [8] detection. This literature review aims to provide an overview of existing research in this area, highlighting the strengths and limitations of the combined methodology.

K-means clustering for tomato disease detection

The Hamirul AiniHambali Author proposed a K-means clustering-based approach for detecting tomato diseases [2], [8] in images. The study demonstrated that K-means clustering effectively segments disease-affected regions based on color similarity. The method achieved satisfactory results in identifying regions of interest for further analysis.

HSV analysis for disease region identification

The HSV color space identify disease-affected regions in tomato images [18]. The research demonstrated the advantage of HSV over RGB color space in capturing subtle color changes associated with various diseases. However, the study highlighted the challenges of manually tuning HSV color thresholds for accurate disease region identification.

Cubic SVM classifier for disease classification

The Author proposed a Cubic SVM classifier [19] to accurately classify different types of tomato diseases. The study used extracted features from segmented regions and achieved promising results in disease classification. However, the scalability and computational complexity of SVM classifiers were acknowledged as potential limitations.

Hybrid approaches and ensemble techniques [20]

The Author introduced a hybrid approach that combines K-means clustering, HSV analysis, and Cubic SVM classifiers for improved disease detection accuracy. The study demonstrated that the fusion of results from multiple classifiers led to superior performance compared to individual methods.

The literature review highlights the growing interest in using K-means clustering [1], HSV analysis, and Cubic SVM classifiers for *Solanum lycopersicum* L. fruit disease detection. These integrated approaches have shown promise in segmenting disease-affected regions and accurately classifying different disease types. However, challenges such as parameter sensitivity, manual threshold tuning, and computational complexity remain areas of concern. Future research should focus on refining the algorithms, exploring hybrid techniques, and developing practical solutions for real-time deployment in diverse agricultural settings. The availability of comprehensive datasets representing various disease scenarios will play a crucial role in advancing the field and enabling more robust and scalable tomato disease detection systems.

K-means clustering (segment an image into different clusters)

K-means clustering [1] is a technique that can be used to segment an image into different clusters/regions, each representing a different class or disease. By using this technique, the image can be divided into regions that correspond to healthy and diseased areas of the fruit.

Disease detection using color space

- HSV color space is a color representation system that separates the color information into three channels: hue, saturation, and value.
- This can be useful for disease detection, as different diseases may cause changes in the color of the fruit.

By using the HSV color space, it is possible to extract the color information of the fruit and use it to detect the presence of a disease.

Proposed algorithm

{Assuming we have a dataset with 'n' tomato fruit images and corresponding labels. Let 'X' be the matrix to store the feature vectors for each tomato image (n x d), where 'd' is the number of extracted features. Let 'y' be the vector to store the corresponding disease labels (n x 1). Let 'k' be the number of clusters for K-means clustering. Let 'HSV' be the matrix to store the HSV components for each tomato image (n x 3), where each row represents the Hue, Saturation, and Value values.}

RESULTS AND DISCUSSION

In Image acquisition all the images were preprocessed and resized with the dimension of 60 X 60 of file type .jpg. Noise was removed using Gaussian filter [4], [11], [16].

In the first method, K-means clustering algorithm [1], [12] applied for all the images to detect tomato disease types [8]. In the second method, the images were converted in to HSV color space to detect tomato diseased area using segmentation. Hence the images were grouped into three clusters. The three clusters were grouped into low level diseased, medium level diseased and high level diseased. In the present study, Standard deviation is a measure of diseased data in relation to the mean. Low standard deviation means the data (low level diseased area) are clustered around the mean and high standard deviation indicate the data (high level diseased area) were more spread out.

Table 1 K means clustering with k = 3 and distance = Squard Euclidean

Image	Mean	Median	Minimum	Maximum	Range	Mode	Variance	Standard deviation
Fungal	152.5	151.0	120	176	56	149	77.6	8.80
Fungal_1a	159.9	162.0	108	191	83	163	248.08	15.70
Early blight1	141.4	137.0	104	186	82	128	286.2	16.90
Early blight3	150.1	148.0	117	204	87	139	260.19	16.13
Late blight_a	136.3	134.5	101	182	81	164	357.3	23.60
Late blight_1a	135.7	131.0	102	175	73	118	456.4	21.36
Late blight_a2	140.5	135.0	120	172	52	133	176.5	13.28

In the above predicted results, the sets of standard deviation values determine three ranges from 8.8 to 23.6. The results reveal the minimum value of standard deviation in diseased tomato as 8.8, and proved to be fungal disease [3]. The distance used is squared Euclidean. The average value of standard deviation in diseased tomato is 16.51 and proved to be early blight disease. The maximum value of standard deviation in diseased tomato is 23.6 and proved to be late blight disease.

Fig 4 Segmented images of hue, saturation, value (HSV) color threshold method

Table 2 Performance of cubic SVM classifier

Segmented images	Accuracy %
Fungal	98.3
Fungal_1a	95.0
Early blight1	96.0
Early blight3	86.7
Late blight_a	85.0
Late blight_1a	93.3
Late blight_a2	96.7

In machine learning Accuracy is one of the best performance metric to classify objects. The accuracy measures the number of correct pixels made by a model in relation to the total number of predictions made. Accuracy can calculate by dividing the number of correct predictions by the total number of predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{---- (1)}$$

Here, TP - True Positive Pixels

TN - True Negative Pixels
FP - False Positive Pixels
FN - False Negative Pixels

The experimental results of cubic svm classifier showed good accuracy 98.3% and an average accuracy percentage is 93%. The proposed study work proved cubic svm classifier is best to classify tomato diseases types fungal [3], early blight and late blight.

The use of K-means clustering, hue saturation value (HSV) analysis, and multiclass Support Vector Machine (SVM) classifier in the detection of *Solanum lycopersicum* L. fruit diseases had demonstrated promising results and significant potential for automated and accurate disease identification. Throughout this study, we explored the combination of these techniques to enhance the tomato disease detection process, and the findings pave way on the strengths and limitations of this integrated approach. The application of K-means clustering played a vital role in segmenting disease-affected regions based on color similarity. The unsupervised technique efficiently partitioned the image data into distinct clusters, enabling the extraction of meaningful features for subsequent analysis. The work revealed the initial cluster centroids for improved accuracy. HSV color analysis offered valuable insights into identifying disease-affected regions with subtle color changes. The ability to capture hue and saturation variations proved essential in distinguishing different disease types. The Cubic SVM classifier emerged as a powerful tool for disease classification, leveraging the extracted features to categorize the tomato diseases accurately. The combination of k-means clustering, HSV color space and Cubic SVM classifier can be a powerful approach for detecting fruit diseases in tomato plants. It can help to improve the accuracy of disease detection and can aid in the control and management of the disease. The research should be conducted with proper dataset to predict relevant results.

CONCLUSION

In conclusion, the integration of K-means clustering, HSV analysis, and Cubic SVM classifier offers a valuable framework for automating tomato disease detection. This approach proved a viable solution for precision agriculture and early disease intervention, thereby contributing to sustainable crop management practices. However, further research is needed to explore alternative initialization strategies for K-means, automated HSV thresholding methods, and techniques to address scalability concerns with large datasets. With continued advancements and optimization, this integrated methodology holds great promise for revolutionizing tomato disease detection, benefitting farmers ensuring global food security.

LITERATURE CITED

1. Hamirul Aini Hambali, Salwa Khalid, Massudi Mahmuddin, Nor Hazlyna Harun. 2018. Segmentation of multi food images using integrated active contour and K-means. *Journal of Engineering and Applied Sciences* 3: 3146-3151.
2. Huang ZH. 2016. *Research on Construction Method of Tomato Disease Image Database*. South China Agricultural University.
3. Pujari JD, Yakkundimath R, Byadgi AS. 2013. Statistical methods for quantitatively detecting fungal disease from fruit’s image. *International Journal of Intelligent System and Application in Engineering* 1(4): 60-67.
4. Bhange M, Hingoliwala HA. 2015. Review of image processing for pomegranate disease detection. *International Journal of Computer Science and Information Technologies* 6(1): 92-94.
5. Jhuria M, Kumar A, Borse R. 2013. Image processing for smart farming: Detection of disease and fruit grading. *IEEE, Proceedings of the 2013, IEEE Second International Conference on Image Information Processing*. pp 521-526.
6. Afonso M, Fonteijn H, Fiorentin FS, Lensink D. 2020. Tomato fruit detection and counting in greenhouse using deep learning. *Frontiers Plant Science* 11: 1- 11.
7. Zhao J, Qu J. 2019. A detection method for tomato fruit common physiological diseases based on Yolov2. *International Conference on Information Technology in Medicine and Education*. 23-25 August 2019. pp 559-563.
8. Wang Q, Feng Q, Sun M, Qu J, Xue J. 2019. Identification of tomato disease types and detection of infected areas based on deep convolutional neural networks and object detection techniques. *Computational Intelligence and Neuroscience* 2019: 1-15.
9. Samaddar S, Reddy ARS. 2018. Comparative study of image segmentation techniques on chronic kidney diseases. *International Journal of Pure and Applied Mathematics* 118: 235-239.
10. Deshpande T, Sengupta S, Raghuvanshi KS. 2014. Grading identification of disease in pomegranate leaf and fruit. *International Journal of Computer Science and Information Technologies* 5(3): 4638- 4645.
11. Gonzalez R, Woods R. 2007. *Digital Image Processing*. 3rd Edition, Prentice- Hall.
12. Pham VH, Lee BR. 2015. An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm. *Vietnam Jr. Computer Sciences* 2: 25-33.
13. Habib MT, Majumder A, Jakaria AZM, Akter M, Uddin MS, Farruk A. 2020. Machine vision-based papaya disease recognition. *Journal of King Saud University – Computer and Information Sciences* 300-309.
14. Hong H, Lin J, Huang F. 2020. Tomato disease detection and classification by deep learning. *International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering* (ICBAIE) 2020: pp 25-29. doi:10.1109/ICBAIE49996.2020.00012.
15. Horea M, Mihai O. 2018. Fruit recognition from images using deep learning. *Act Auniv. Sapientiae, Informatica* 10(1): 26-42.
16. Ali I, Ahmed A. 2018. Segmentation of different fruits using image processing based on fuzzy C-means method. *7th International Conference on Reliability*, ICRITO, IEEE, 2018. pp 441-447.
17. Lingeswari S, Gomathi PM, Kailasam SP. 2021. A study of tomato fruit disease detection using RGB color thresholding and K-means clustering. *International Journal of Computer Science and Mobile Computing* 10(8): 51-59.
18. Wang Q, Feng Q, Minghe S, Jianhua Q, Jie X. 2019. Identification of tomato disease types and detection of infected areas based on deep convolutional neural networks and object detection techniques. *Computational Intelligence and Neuroscience* 2019: Article ID 9142753, 15 pages. https://doi.org/10.1155/2019/9142753
19. Fernandez-Lozano C, Canto C, Gestal M, Andrade-Garda JM, Rabuñal JR, Dorado J, Pazos A. 2013. Hybrid model based on genetic algorithms and SVM applied to variable selection within fruit juice classification. *The Scientific World Journal* 2013: Article ID 982438, 13 pages. https://doi.org/10.1155/2013/982438
20. Altalak M, Uddin MA, Alajmi A, Rizg A. 2022. A hybrid approach for the detection and classification of tomato leaf diseases. *Applied Sciences* 12(16): 8182. https://doi.org/10.3390/app12168182