Unsupervised Machine Learning Approach in Smart Agriculture to Measure Disease Severity of Tomato Leaf

Trijit Arka Ghosh[1,*,†], Atanu Mondal[2,†]

[1]Ramakrishna Mission Vidyamandira, Belur Math , Howrah-711202, West Bengal, India
[2]Ramakrishna Mission Vidyamandira, Belur Math , Howrah-711202, West Bengal, India

Abstract

Different diseases show different symptoms on plant leaves due to different reasons. Identifying plant leaf diseases without specialist expertise is typically challenging like finding region-areas of interest (ROIs) that is disease affected area on the leaf. The well known unsupervised machineilearningialgorithmsiK-meansiandiFuzzyiC-Meansi(FCM) approaches based on clustering for plant leaf disease identification and severity evaluation is considered as a powerful technique to determine the form of illness has afflicted a particular tomato leaf and the severity. The segmentation technique is used to locate the ROI on a tomatoileaf. The objective of this clustering is to determine the accuracy of the segmented image of disease-affected leaves. The experiment utilizes images obtained from the Plant Village Dataset. The two stages of yellow leaf curl virus (YLCV) disease affected area have been observed. The severity score has been calculated for each type. The optimum cluster number of different severity level has been investigated based on various cluster validation index. Then different optimal cluster numbers of different severity levels have been shown by using Calinski and Harabasz index (CH) and Davies-Bouldin index (DB). The comparative study demonstrates the method's superiority relative to current methodologies.

Keywords

Smart agriculture, Machine Learning, Clustering, Cluster Validation Index

1. Introduction

A report published by the Food and Agriculture Organization[1] in 2017, found that by 2050 world's population will reach 9.7 billion and 11.2 billion by 2100[2] respectively, accordingly as a consequence, the demand for food production willialsoiincrease to satisfy the required demand. The problem in reduced crop production due to different environmental issues as well as agricultural concept is also generated with the increase of demand of food production. Crops produced from agriculture plays a pivotal role to build a nation's economy in different ways like food for human-beings and animals as well as raw materials for industry. A total of four revolutions[3] have occurred in the development of agriculture since the primitive civilization till today. The last revolution in agriculture has taken place over the last two decades along with tremendous development of information technology, network communication and artificial intelligence, thus smart agriculture comes into field, where advanced modern technologies, the tools and devices are used in farming. Crops and weeds can be harvested in different ways with the help of robots, drones are used to accurately monitor crop fertilizer and crop growth levels. In addition, smart agriculture by using information communication technology and artificial intelligence cyber-physical farm management is also conducted. Hence smart agriculture is dependent on IoT technologies which are used to protect plant and make proper irrigation and to improve quality of product, control to ban disease and detection of process etc. Thus IoT is measured as the mainstay of smart agriculture. All the farming devices and equipment can be connected together with the technology of IoT[4] to take right decision at the right time in disease detection, irrigation, fertilizer supply with the help of analytical result of data which are gathered by different sensors[5]

The 2024 Sixth Doctoral Symposium on Intelligence Enabled Research (DoSIER 2024), November 28–29, 2024, Jalpaiguri, India
*Corresponding author.
†These authors contributed equally.
trijitarkaghosh@gmail.com (T. A. Ghosh); atanumondal@hotmail.com (A. Mondal)
0000-0003-4672-1180 (T. A. Ghosh); 0000-0002-1819-0602 (A. Mondal)

connected with different cloud services controlled by satellite. Thus, smart agriculture becomes so important by providing the features like (1) increase the turn up of collected real-time crop data, (2) help to overseeing and managing farmers from a distance, (3) managing water and other natural resources, (4) precise assessment of soil and crops and (5) enhancing agricultural output. One of the important fields of smart agriculture is Precision agriculture which is a new frontier in agriculture that measures the amount of herbicides and pesticides a plant needs and applies them to that plant according to actual needs to achieve economic and environmental benefits. Variable rate application technology is a rapidly expanding precision technology that integrates with systems like global positioning technology, geographic information systems, and other technologies for seeding, weed and pest control, lime distribution, and fertilizer application. In a word, the main application of precision agriculture is to reduce the misuse of fertilizer and pesticides. This concept is actually used in this present work to measure the severity of tomato disease. The present article is structured asifollows: Section 2 deals with the problems of smart agriculture. In sectioni3 the relevant researches and suggested methods are highlighted. The proposed methodology is presentediin section 4. Sectioni5 deals with the analyses. The conclusion and future work is mentioned in section 6.

2. Problems in smart agriculture

In the 21st century, advanced technologies are being used in agriculture all over the world. In addition, to environmental and technical issues, some other issues like diverse soil and cropping patterns, digital divide i.e., unavailability of strong and stable internet connectivity, lack of confidence of farmers in new advanced technologies, lack of co-ordination between stakeholders which in turn highlights the issue of not reaching the farmers from different agricultural universities in the same geographical area at the same time and unreliable data sets respectively.

3. Literature Survey

If the disease attacks the leaves of any tree, the colour and structure of the leaves of that tree will change. Same things happens with tomato plants. For this reason, various researchers have studied computer vision and machine learning algorithms for detecting leaf infections at different times of different plants, classifying diseases and measuring disease severity. This section provides a various computerivision and machineilearning based techniques used by researchers in plant disease detection and disease severity measurement. Hong et al.[6] have used tomato for their research and five deep network architecture of Resnet50, Xception, MobileNet, ShuffleNet, DensenetXception were used for feature extraction. In their research, recognition accuracy was 83.68% for ShuffleNet and 97.10% for DensenetXception. Theistudy by Agarwal et al.[7] discussed Convolution Neural Network (CNN) for tomato leaf disease detectioniand classification. Ashok & Vinod[8] have used Mango Fruits for disease Detection by deep CNN architecture. Wongsila et al.[9] have used CNN for detection of mangoes infected with anthracnose. They were able to achieve more than 70% accuracy to isolate the diseased mango after testing on 364 images. Sabrol & Kumar[10] studied on Adaptive Neuro-Fuzzy Classification for plant leaf disease detection using GLCM matrix.

Abisha & Jayasree[11] discussed automatic detection andiclassification ofidifferent typesiof brinjal leaf diseasesiusing Artificial Neural Network (ANN) which shows the improvement of classification accuracy. Islam et al.[12] showed the method for paddy leaf disease detection, where they have used 4 typesiof diseases and oneihealthy leaficlass of the paddy. Deep learning CNN models was used for their research and they were able to achieve more than 92.68% accuracy for paddy leaf disease detection. Nalini et al.[13] have usedik-meansiclusteringimethod for pre-processing and DeepiNeuraliNetwork (DNN) classificationimodel for the identificationiof paddy leafidisease using plantiimage data. They have used crow search algorithmi(CSA) to minimize classification error. According to their study, their proposediDNN-CSA model provides better classificationiaccuracy to a support vector machine with multiple cross-foldivalidations. In Jayanthi and Shashikumar[14] discussed probabilistic neuralinetwork

(PNN) for cucumber leafidisease detection. They have first used adaptively regularized kernel-based FCM (ARKFCM) for segmentation, then Hue, Saturation and Value (HSV) Technique and Grey level co-occurrence matrix (GLCM) technique were used for color feature and texture feature extraction respectively. Then extracted features were given to PNN for disease detection. Performance of their proposed method was analysed in terms of accuracy, sensitivity and specificity. RajaKumar et al.[15] have used ANN for cucumber leaf disease detection. They were able to achieve 98.66. Another study by Sanga et al.[16] discussed fiveideep learning architecturesinamely Vgg16, Resnet18, Resnet50, Resnet152 andiInceptionV3ifor banana disease detection and they were able to achieve all high accuracies.

Machine learning and deep learning have been recently emphasized in smart agriculture as tools for disease identification. Methods such as adaptive fuzzy systems and convolutional neural networks are used in previous works. However, unsupervised clustering for severity evaluation has received little attention in the literature. This study fills that need by using clustering algorithms and demanding evaluation metrics to validate the results. Innovative use of clustering indices and an emphasis on agriculturally-oriented, scalable systems are two important contributions.

4. Present Work

The proposed methodology involves clustering-based image segmentation for disease severity evaluation. K-means and Fuzzy C-Means algorithms are used to identify affected regions. The methodology includes:

1. Data preprocessing to enhance image quality:

Images are preprocessed to enhance quality and ensure uniformity across the dataset. This includes resizing, normalization, and noise reduction.

2. Clustering to segment the diseased region.

3. Validation using Calinski-Harabasz and Davies-Bouldin indices to determine optimal clusters.

Comparative analysis with existing benchmarks demonstrates the approach's effectiveness in identifying disease severity levels which is explained at the end of the section 5.

This work is proceeded up under unsupervised machine learning domain. The proposed method in this work evaluates the severity of tomato leaf disease without supervision, utilizing K-means and FCM clustering approach. Clustering is an unsupervised pattern classification method for machine learning. This clustering method is used as a very important tool in data science, where the input space can be divided into one or more clusters to identify the natural structure of data within a dataset[17]. Such an important clustering method is used in image processing, network sensing, pattern recognition, psychology, computer security, recommendation system, biology, text clustering etc. Hence, the detection of tomato leaf disease is also implemented by using clustering which is discussed in the following sections.

4.1. K-Means approach

Let $P = \{p_1, p_2, \ldots, p_n\}$ be a data set in a d-dimensional Euclidean space $\mathbb{R}^d$. Let $B = \{b_1, b_2, \ldots, b_t\}$ be the t cluster centers. Let $Y = [Y_{ik}]_{n\times t}$, where Y_{ik} is a binary variable (i.e., $Y_{ik} \in \{0, 1\}$) indicating if the data point p_i belongs to the k-th cluster ($k = 1, \ldots, t$). The objective function of k-means is as follows:

$$J(Y, B) = \sum_{i=1}^{n} \sum_{k=1}^{t} Y_{ik} \|p_i - b_k\|^2 \tag{1}$$

The k-means algorithm is repeated through essential conditions for minimizing the k-means objective function $J(Y, B)$ with updating equations for cluster centers and memberships, respectively, as:

$$a_k = \frac{\sum_{i=1}^{n} z_{ik} x_{ij}}{\sum_{i=1}^{n} z_{ik}} \tag{2}$$

$$Z_{ik} = \begin{cases} 1, & \text{if } \|x_i - a_k\|^2 = \min_{1 \le k \le c} \|x_i - a_k\|^2 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $\|p_i - b_k\|$ is the Euclidean distance between the cluster center b_k and the data point p_i.

A challenging issue in k-means is that it requires specifying the number of clusters in advance, but this number is often unknown in practical situations. The superiority of the clusters produced by the k-means algorithm depends on the initial cluster centers. This is also a limitation of using this algorithm. To resolve the issue of finding the number of clusters, cluster validity indices have gained much attention. Hence, Fuzzy C-Means (FCM) is used here for overlapping datasets.

4.2. FCM approach

FCM makes partition of a finite collection of n elements $P = \{P_1, \ldots, P_n\}$ into a collection of C fuzzy clusters based on certain principles. Given a finite set of data points from affected tomato leaves, the FCM algorithm produces a list of C cluster centers $C = \{c_1, \ldots, c_C\}$ and a partition matrix:

$$W = \{w_{ij} \mid w_{ij} \in [0,1],\ i = 1, \ldots, n,\ j = 1, \ldots, C\}$$

where each element w_{ij} indicates the degree to which element x_i belongs to cluster c_j.

The goal of FCM is to minimize the following objective function:

$$J(W, C) = \sum_{i=1}^{n} \sum_{j=1}^{C} w_{ij}^m \|x_i - c_j\|^2 \tag{4}$$

where m is the hyper-parameter that controls how fuzzy the clustering will be. A higher value of m results in fuzzier clusters.

The membership degree w_{ij} is calculated as:

$$w_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{5}$$

After determining the number of clusters, cluster validation methodologies are applied. Two types of cluster justification indices, external and internal[18], are available. In this work, widely used internal indices such as Sum of Squares Between Clusters (SSB), Sum of Squares Within Clusters (SSW), CH[19], and DB[20] indices are employed to determine the number of optimal clusters and to analyze the relationship between optimal clusters and disease severity.

4.3. Sum of Squares within Clusters

In the case of numerical data, the amount of SSW variance[21] can be determined with the help of the cluster centroid. SSW is used to measure cluster compactness. From the data obtained from various studies, it is observed that as the number of clusters increases, the value of SSW decreases. The equation to measure cluster compactness is given below:

$$\text{SSW} = \sum_{i=1}^{N} \|X_i - C_{l_i}\|^2 \tag{6}$$

4.4. Sum of Squares between Clusters

SSB is used to measure the degree of separation between clusters. The centroid distance from the mean vector of all objects is first measured, and this is used to compute the degree of separation between clusters. It is observed from the data of various studies[21] that as the number of clusters increases, the value of SSB also increases. The equation to measure the degree of separation between clusters is given below:

$$\text{SSB} = \sum_{i=1}^{K} n_i \|c_i - \bar{x}\|^2 \tag{7}$$

4.5. Calinsk-Harabasz (CH) index

The value of CH is measured as the ratio of separation and compactness of the clusters. Cluster quality is better when the value of the CH index is maximum. SSB's value should be higher, and the value of SSW should be lower as the number of clusters increases. The equation to measure the value of this index is given below:

$$\text{CH} = \frac{\text{SSB}/(K-1)}{\text{SSW}/(N-K)} \tag{8}$$

4.6. Davie-Bouldin (DB) index

The DB criterion is based on a ratio of within-cluster and between-cluster distances. The equation to measure the value for this index is as follows:

$$\text{DB} = \frac{\text{SSW}}{\text{SSB}} \tag{9}$$

Hence, by using these aforementioned approaches, the severity of disease on tomato leaves is identified, and the corresponding results and analysis are mentioned in the following section.

5. Result and Analysis

In order to measure tomato plant leaf disease severity, it is needed a large,iverifiedidataset ofiimages of healthy andidiseased tomato leaves. In this work, images of diseases of Tomato leaves have been taken from PlantVillage dataset[22]. There are 16,012 tomato leaves images in this dataset. The size of all the images is 256×256 and format is png. The PlantVillage data set contains a total of 1591 tomato healthy leaf images; out of 1591 only 50 images are randomly picked. The first six healthy leaf images from those are shown in Figure 1.

Similarly, 50 leaf images were randomly picked from 329 tomato yellow leaf curl virus (YLCV) leaf images in PantVillage dataset and stored in YLCV folder. Then from YLCV folder six images were picked in such a way that each leaf image has less yellow part than green part. These 10 leaf images are placed in the Grade-I folder which are shown in Figure 2.

The k-Means clustering algorithm is applied on Sample-1 in Figure 1. A total of 9 cluster images with values of k spanning from 2 to 10 are shown in Figure 4. The FCM clustering algorithm is applied on Sample-1 in Figure 1. A total of 9 cluster images with values of k spanning from 2 to 10 are shown in Figure 5. The k-Means clustering algorithm is applied on Sample-1 in Figure 2. A total of 9 cluster images are shown in Figure 6, with values of k spanning from 2 to 10. The FCM clustering algorithm is applied on Sample-1 in Figure 2. A total of 9 cluster images with values of k spanning from 2 to 10 are shown in Figure 7. Figure 3 shows the severity score δ of each sample image of Grade-I. Equation (10) is used to measure δ for each sample image belonging to Grade-I, where d indicates the pixel number only within the regions of a disease-affected leaf, and n indicates the total pixel number in that disease-affected leaf.

$$\delta = \frac{d}{n} \times 100\% \tag{10}$$

After obtaining the clustered image using K-Means and FCM clustering algorithms, internal validity index, DB is used to know how good the cluster structure is. Various graphs with optimum cluster number (k) obtained using DB are shown from Figures 12-17.

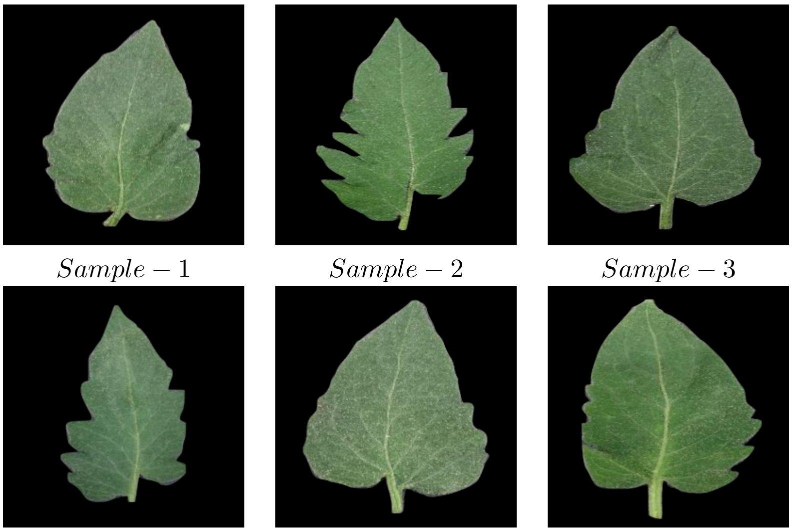


Figure 1: Samples of healthy tomato leaf images.

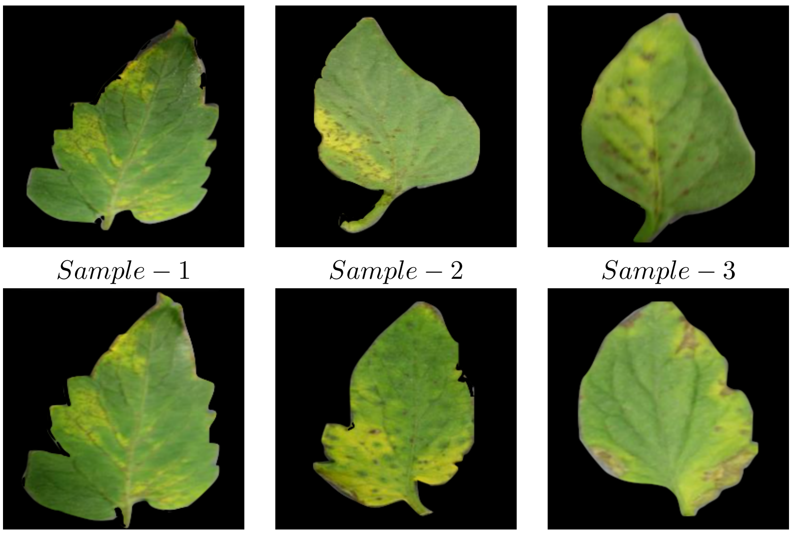


Figure 2: Samples of YLCV disease affected tomato leaf images of Grade-I.

5.1. Relationship between optimal K and disease severity

Clustering is commonly related to optimization problems by considering the optimization criteria to assess the class of clustering which belong to the internal validity index and related to static number of k clusters. Determining the optimal clusters number in a dataset is a very important task in partition clustering such as K-Means. SSW method is used to measure the compactness of a cluster. The smaller value of SSW indicates better quality in that cluster. Various experiments have shown that the value of SSW decreases as the clusters number increases. On the other side, SSB method is used to measure the separation between clusters. CH index is used to measure the best possible separation and compactness values between clusters. From Equation (8), it is easily understood that the worth of CH index increases when the value of SSB increases and the value of SSW decreases. So it can be said that with high separation the worth of CH index is maximum due to less error of compactness. The highest value of this CH index indicates the best clustering. On the other hand the value of DB index is shown in Equation (9). 50 Healthy leaves, 50 Grade-I diseased leaves, and 50 Grade-II diseased leaves (total 150 leaves) were used for generating CH index values as well as for DB index and then scatter diagrams are plotted in Figure 15 and Figure 16 respectively.

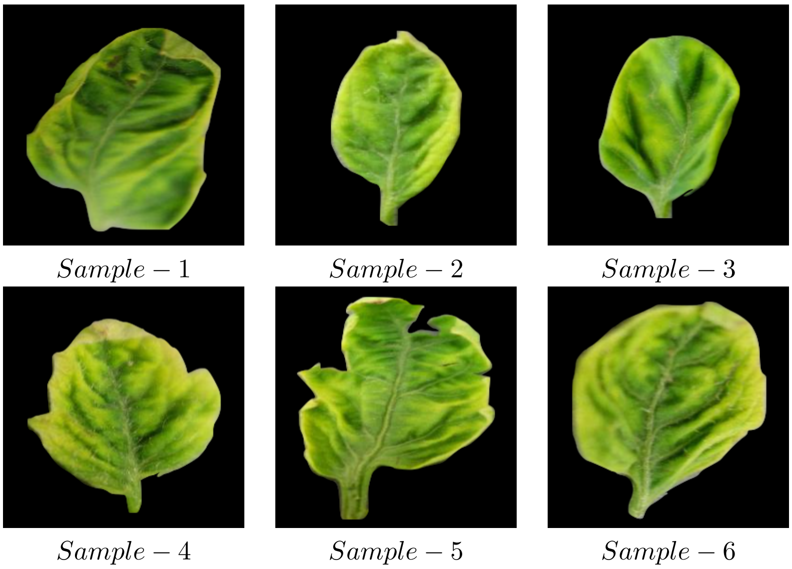


Figure 3: Samples of YLCV disease affected images of tomato leaves of Grade-II.

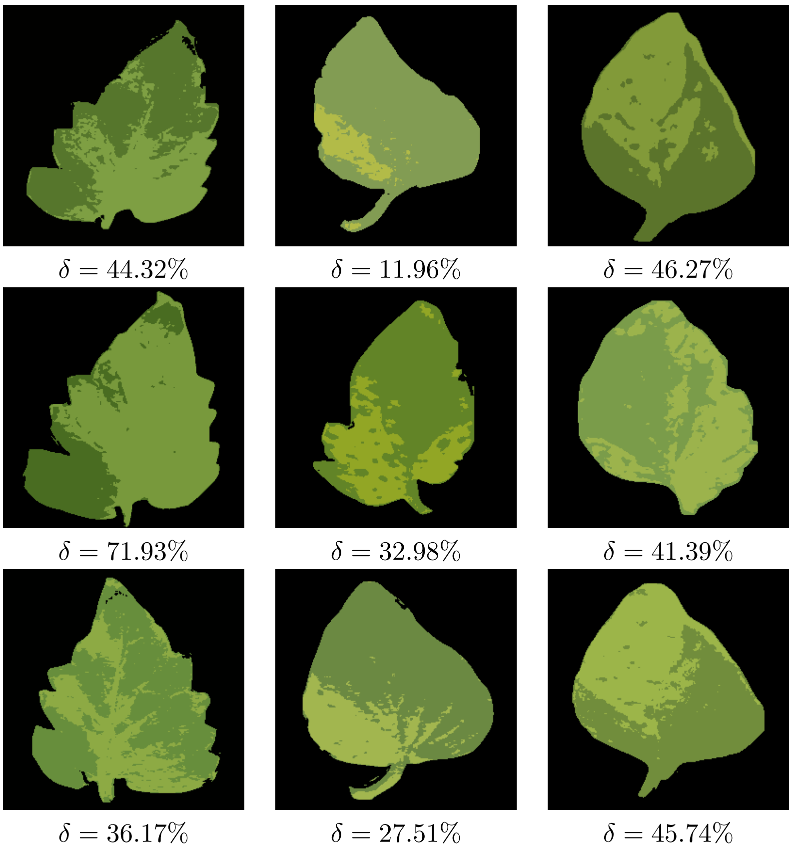


Figure 4: Severity score δ of images of tomato leaves affected by disease of Grade-I.

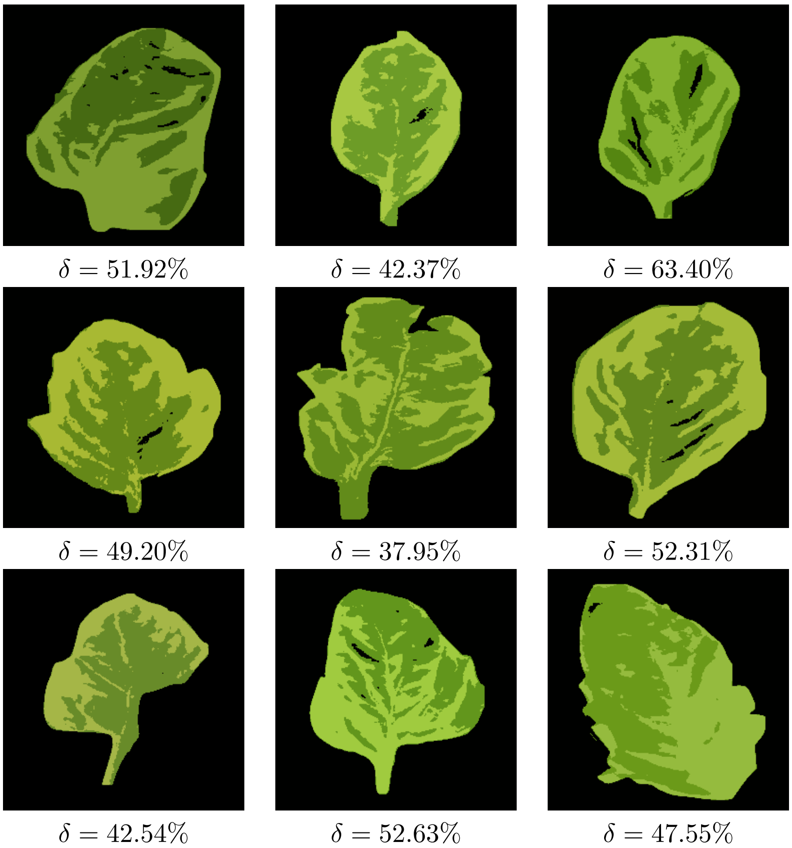


Figure 5: Severity score δ of images of tomato leaves affected by disease of Grade-II.

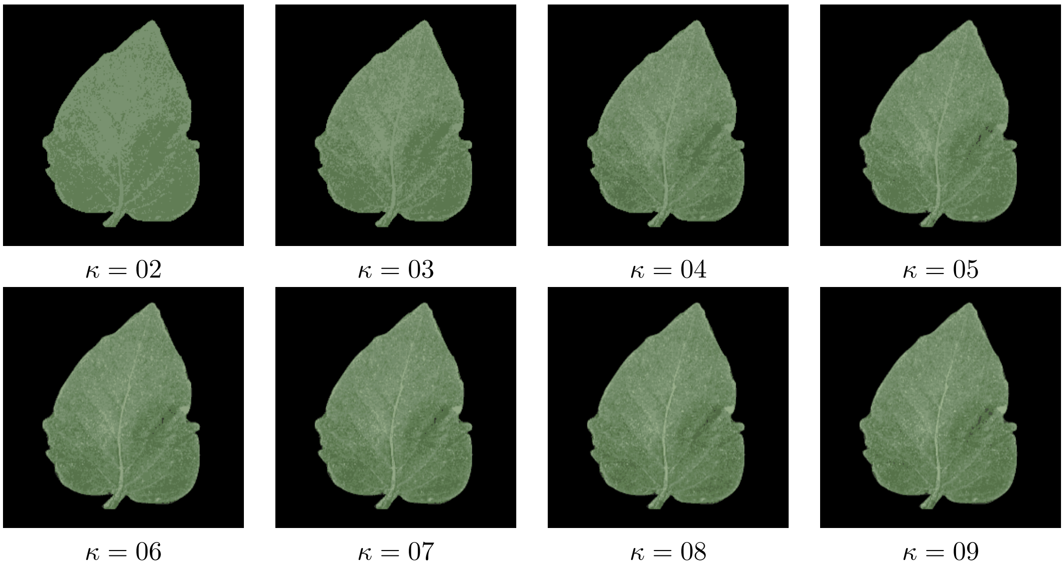


Figure 6: K-Means Cluster images of different value of κ of $Sample - 1$ of Figure 1.

Figure 15 shows the scatter plot of the CH index. The CH index value was measured on images of 50 healthy leaves. A histogram of healthy leaves is plotted in Figure 17, where optimal k values along X-axis and the frequency of optimal values of k along Y-axis are shown respectively. Similarly, two more histogram plots are shown with Grade I and Grade II diseased leaves, in Figure 17.

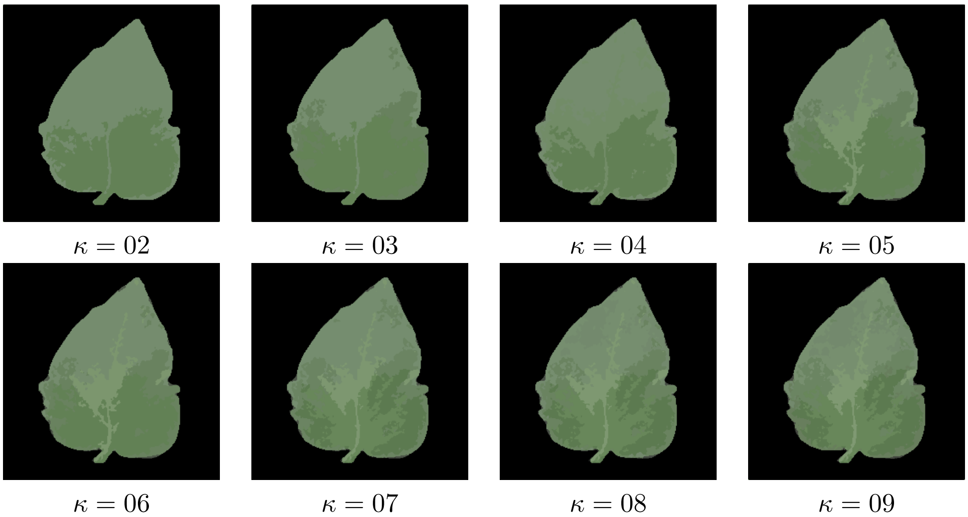


Figure 7: FCM Cluster images of different value of κ of $Sample - 1$ of Figure 1.

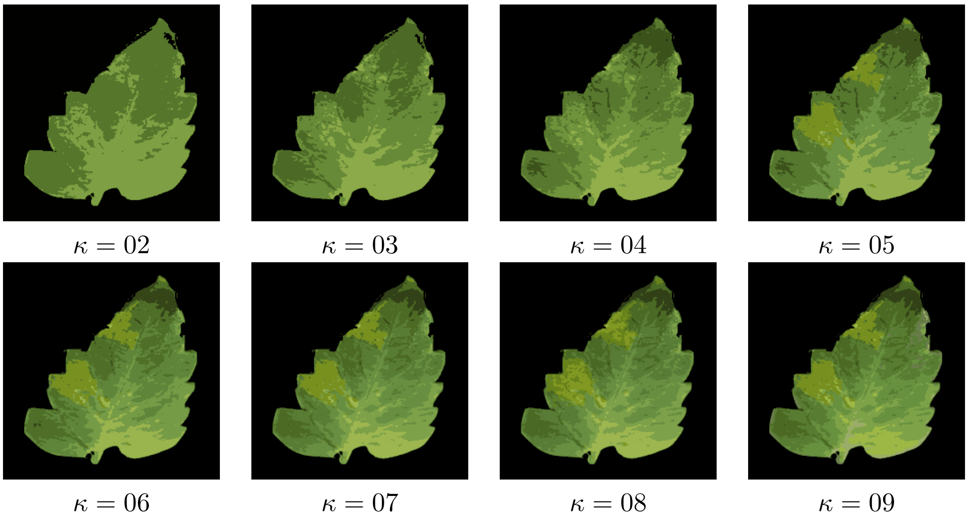


Figure 8: K-Means Cluster images of different value of κ of $Sample - 1$ of Figure 2.

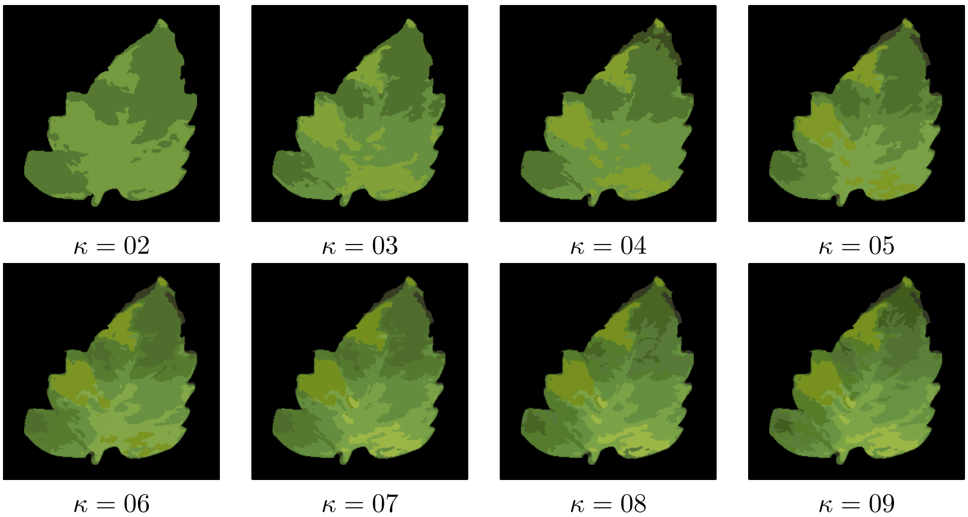
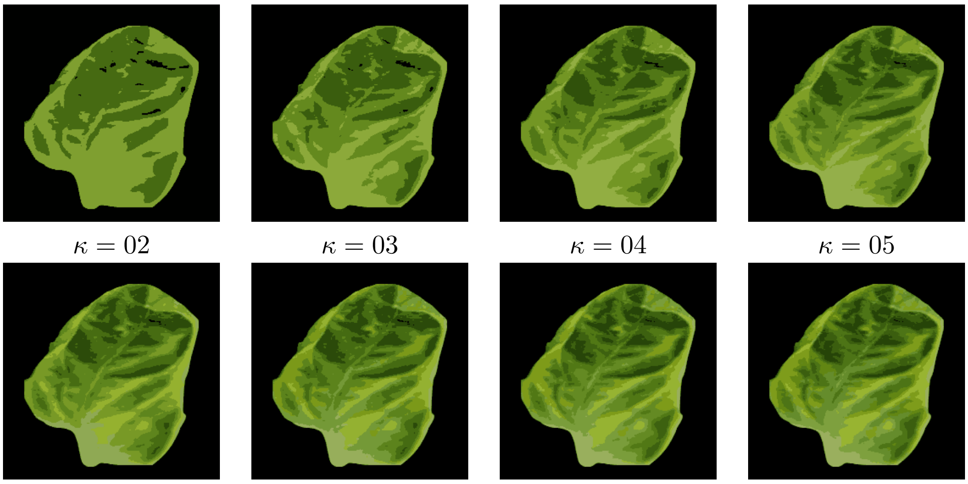


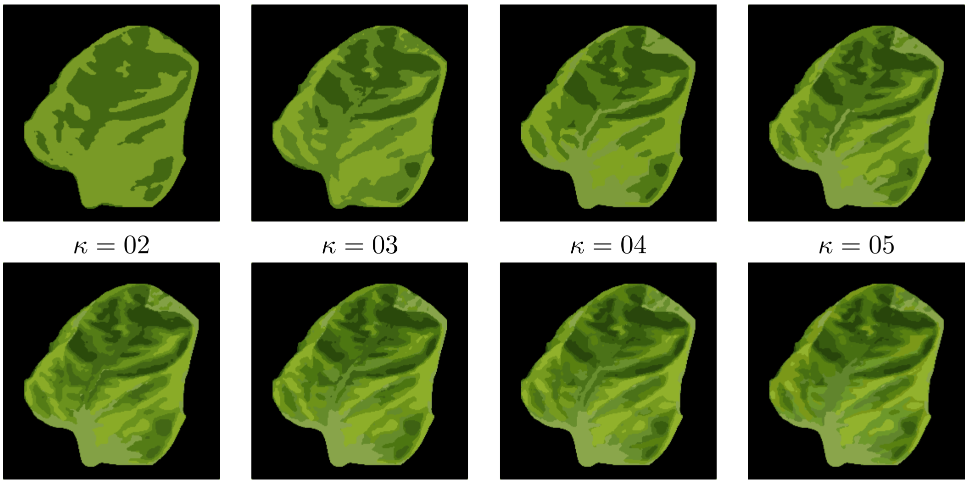
Figure 9: FCM Cluster images of different value of κ of $Sample - 1$ of Figure 2.



$\kappa = 02$ $\kappa = 03$ $\kappa = 04$ $\kappa = 05$

$\kappa = 06$ $\kappa = 07$ $\kappa = 08$ $\kappa = 09$

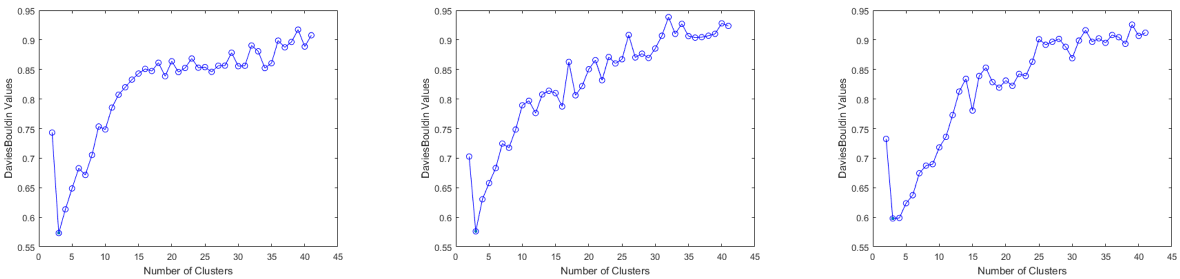
Figure 10: K-Means Cluster images of different value of κ of $Sample - 1$ of Figure 3.



$\kappa = 02$ $\kappa = 03$ $\kappa = 04$ $\kappa = 05$

$\kappa = 06$ $\kappa = 07$ $\kappa = 08$ $\kappa = 09$

Figure 11: FCM Cluster images of different value of κ of $Sample - 1$ of Figure 3.



$Sample - 1, \kappa^{*} = 3$ $Sample - 2, \kappa^{*} = 3$ $Sample - 3, \kappa^{*} = 3$

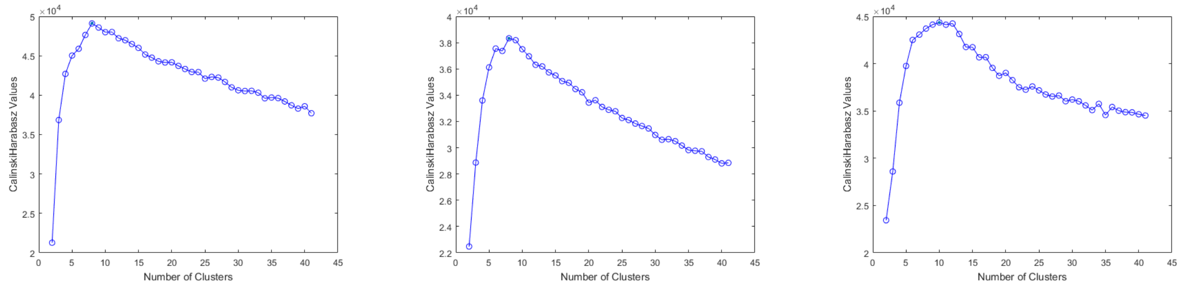
Figure 12: Samples of healthy tomato leaf images with DB index and optimum cluster number (κ^{*}).

5.2. Comparative Analysis

Below is a table presenting the comparative analysis between the proposed methodology and existing methods.

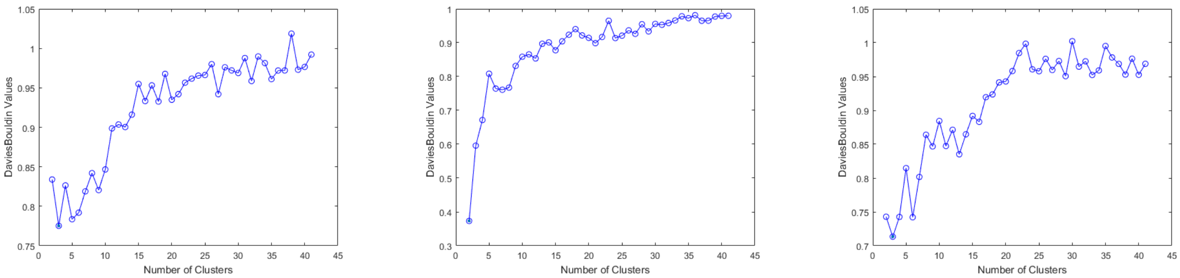
Key findings include:

• Optimal clustering results with an effective segmentation of diseased regions.



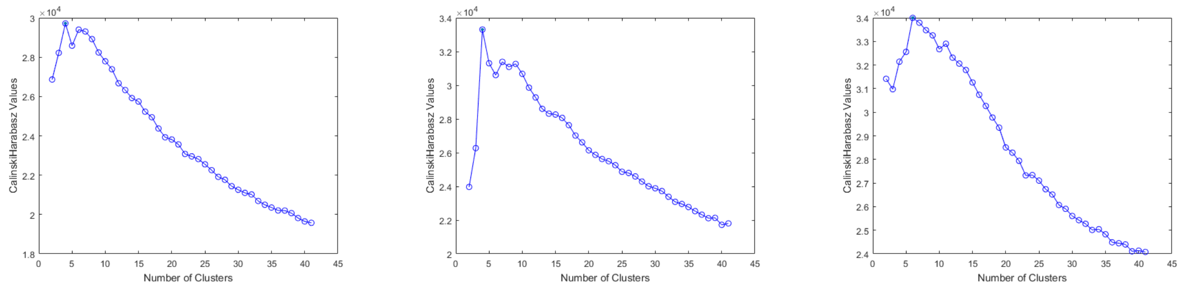
$Sample - 1, \kappa^* = 8$ $Sample - 2, \kappa^* = 8$ $Sample - 3, \kappa^* = 10$

Figure 13: Samples of healthy tomato leaf images with CH index and optimum cluster number (κ^*).



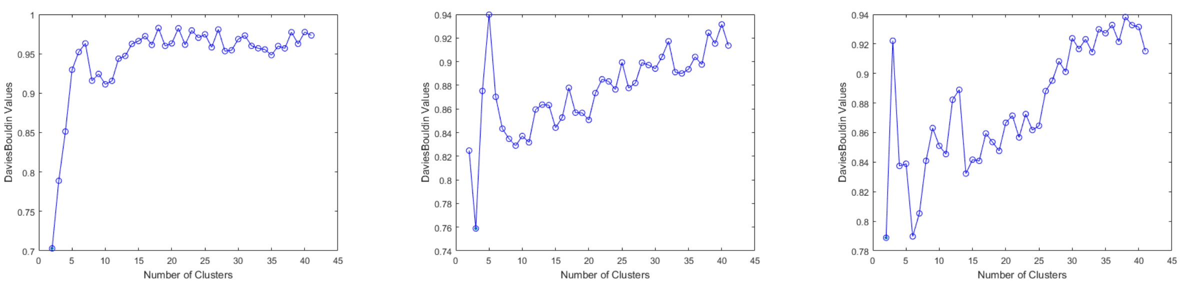
$Sample - 1, \kappa^* = 3$ $Sample - 2, \kappa^* = 2$ $Sample - 3, \kappa^* = 3$

Figure 14: Samples of YLCV Grade-I tomato leaf images with DB index and optimum cluster number (κ^*).



$Sample - 1, \kappa^* = 4$ $Sample - 2, \kappa^* = 4$ $Sample - 3, \kappa^* = 6$

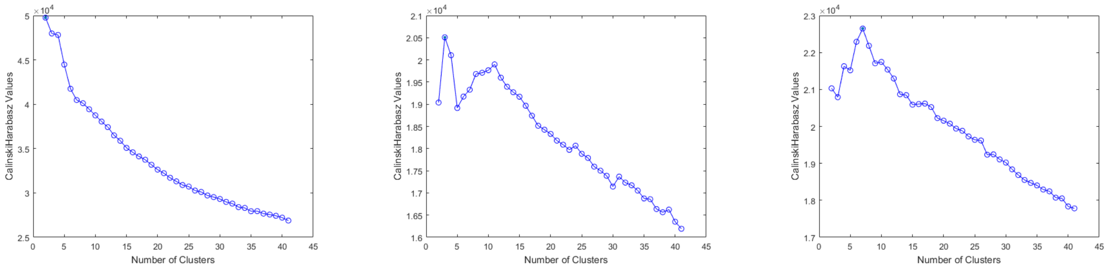
Figure 15: Samples of YLCV Grade-I tomato leaf images with CH index and optimum cluster number (κ^*).



$Sample - 1, \kappa^* = 2$ $Sample - 2, \kappa^* = 3$ $Sample - 3, \kappa^* = 2$

Figure 16: Samples of YLCV Grade-II tomato leaf images with DB index and optimum cluster number (κ^*).

• Comparative analysis highlights superior performance over existing methods.



$Sample-1, \kappa^*=2$ $Sample-2, \kappa^*=3$ $Sample-3, \kappa^*=7$

Figure 17: Samples of YLCV Grade-II tomato leaf images with CH index and optimum cluster number (κ^*).

Table 1
Comparative Analysis

Criterion	Proposed Methodology (K-Means & FCM)	Existing Methods
Accuracy of Disease Detection	Focus on severity evaluation.	Limited focus on severity evaluation.
Robustness to Data Variability	FCM handles overlapping data well, providing soft clustering.	Threshold-based methods are sensitive to image noise and variability.
Computational Complexity	K-Means: Computationally efficient but sensitive to centroid initialization. FCM: Handles complex data but is more resource-intensive.	Deep learning models require high computational power and large datasets for training.
Validation Metrics	Uses Calinski-Harabasz (CH) and Davies-Bouldin (DB) indices. CH: Higher scores indicate better clustering. DB: Lower scores indicate compact, well-separated clusters.	Rarely report cluster validation metrics, focusing instead on model accuracy metrics.
Applicability in Resource-Limited Settings	Requires no labeled data. Suitable for low-resource environments.	Demands extensive labeled datasets and high-performance computational systems.

6. Conclusion and future work

This work illustrates the usefulness of unsupervised machine learning methods, particularly K-means and Fuzzy C-Means (FCM), to determine disease severity in tomato leaves. The suggested method utilizes clustering-based image segmentation and validates the outcomes through Calinski-Harabasz and Davies-Bouldin indices, resulting in high accuracy in segmenting diseased areas. The comparative analysis demonstrates its superior effectiveness compared to existing approaches, especially in resource-constrained environments. These findings confirm the methodology's potential for practical applications in smart agriculture. Subsequent study will concentrate on bringing together this methodology with Internet of Things (IoT) devices for continuous monitoring and decision-making in agriculture. Furthermore, overcoming limitations such as dependence on high-quality datasets and exploring its applicability to additional crops and diseases would improve its robustness and scalability. These innovations promise to enhance the accessibility and utility of precision agriculture, thereby largely contributing to sustainable farming techniques.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] FAO, The future of food and agriculture–trends and challenges. annual report, 2017. URL: https://openknowledge.fao.org/server/api/core/bitstreams/2e90c833-8e84-46f2-a675-ea2d7afa4e24/content.

[2] A. El-Zeiny, A. El-Hefni, M. Sowilem, Geospatial techniques for environmental modeling of mosquito breeding habitats at suez canal zone, egypt, The Egyptian Journal of Remote Sensing and Space Science 20 (2017) 283–293. URL: http://dx.doi.org/10.1016/j.ejrs.2016.11.009. doi:10.1016/j.ejrs.2016.11.009.

[3] S. E. Mohamed, A. Razek, Agricultural revolution, egypt, Sustainability 13 (2021) 1824. URL: https://doi.org/10.3360/su30118. doi:10.3390/su30118.

[4] T. U. Kumar, A. Periasamy, Iot based smart farming (e-farm)'s, International Journal of Recent Advances in Multidisciplinary Topics 2 (2021) 85–87. URL: https://journals.ijramt.com/index.php/ijramt/article/view/646.

[5] R. Kumar, R. Mishra, H. P. Gupta, T. Dutta, Smart sensing for agriculture: Applications, advancements, and challenges, IEEE Consumer Electronics Magazine 10 (2021) 51–56. URL: http://dx.doi.org/10.1109/mce.2021.3049623. doi:10.1109/mce.2021.3049623.

[6] H. Hong, J. Lin, F. Huang, Tomato disease detection and classification by deep learning, in: 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), IEEE, 2020, p. 25–29. URL: http://dx.doi.org/10.1109/ICBAIE49996.2020.00012. doi:10.1109/icbaie49996.2020.00012.

[7] M. Agarwal, A. Singh, S. Arjaria, A. Sinha, S. Gupta, Toled: Tomato leaf disease detection using convolution neural network, Procedia Computer Science 167 (2020) 293–301. URL: http://dx.doi.org/10.1016/j.procs.2020.03.225. doi:10.1016/j.procs.2020.03.225.

[8] V. Ashok, D. S. Vinod, A Novel Fusion of Deep Learning and Android Application for Real-Time Mango Fruits Disease Detection, Springer Singapore, 2020, p. 781–791. URL: http://dx.doi.org/10.1007/978-981-15-5400-1_74. doi:10.1007/978-981-15-5400-1_74.

[9] S. Wongsila, P. Chantrasri, P. Sureephong, Machine learning algorithm development for detection of mango infected by anthracnose disease, in: 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, IEEE, 2021, p. 249–252. URL: http://dx.doi.org/10.1109/ECTIDAMTNCON51128.2021.9425737. doi:10.1109/ectidamtncon51128.2021.9425737.

[10] H. Sabrol, S. Kumar, Plant Leaf Disease Detection Using Adaptive Neuro-Fuzzy Classification, Springer International Publishing, 2019, p. 434–443. URL: http://dx.doi.org/10.1007/978-3-030-17795-9_32. doi:10.1007/978-3-030-17795-9_32.

[11] S. Abisha, T. Jayasree, Application of image processing techniques and artificial neural network for detection of diseases on brinjal leaf, IETE Journal of Research 68 (2019) 2246–2258. URL: http://dx.doi.org/10.1080/03772063.2019.1696716. doi:10.1080/03772063.2019.1696716.

[12] M. A. Islam, M. Nymur, M. Shamsojjaman, S. Hasan, M. Shahadat, T. Khatun, An automated convolutional neural network based approach for paddy leaf disease detection, International Journal of Advanced Computer Science and Applications 12 (2021). URL: http://dx.doi.org/10.14569/IJACSA.2021.0120134. doi:10.14569/ijacsa.2021.0120134.

[13] S. Nalini, N. Jayasankar, et al., Paddy leaf disease detection using an optimized deep neural network, Computers, Materials and Continua 68 (2021) 1117–1128.

[14] M. Jayanthi, D. R. Shashikumar, Cucumber disease detection using adaptively regularised kernel-based fuzzy c-means and probabilistic neural network, International Journal of Computational Vision and Robotics 10 (2020) 385. URL: http://dx.doi.org/10.1504/IJCVR.2020.109390. doi:10.1504/ijcvr.2020.109390.

[15] RajaKumar.R, Pandian.R, PremJacob.T, Detection and Classification of Diseases in Cucumber Leaves, IOS Press, 2021. URL: http://dx.doi.org/10.3233/APC210008. doi:10.3233/apc210008.

[16] S. L. Sanga, D. Machuve, K. Jomanga, Mobile-based deep learning models for banana disease detection, Engineering, Technology and Applied Science Research 10 (2020) 5674–5677. URL:

http://dx.doi.org/10.48084/etasr.3452. doi:10.48084/etasr.3452.
[17] K. P. Sinaga, M.-S. Yang, Unsupervised k-means clustering algorithm, IEEE Access 8 (2020) 80716–80727. URL: http://dx.doi.org/10.1109/access.2020.2988796. doi:10.1109/access.2020.2988796.
[18] S. Lloyd, Internal versus external cluster validation indexes, IEEE Transactions on Information Theory 28 (2011) 129–137. URL: http://dx.doi.org/10.1109/tit.1982.1056489. doi:10.1109/tit.1982.1056489.
[19] T. Calinski, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics - Theory and Methods 3 (1974) 1–27. URL: http://dx.doi.org/10.1080/03610927408827101. doi:10.1080/03610927408827101.
[20] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (1979) 224–227. URL: http://dx.doi.org/10.1109/tpami.1979.4766909. doi:10.1109/tpami.1979.4766909.
[21] Q. Zhao, Cluster validity in clustering methods, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2012) 224–227. URL: https://doi.org/10.1109/tpami.2012.47909. doi:10.1109/tpami.2012.47909.
[22] Emmarex, Plant disease dataset, 2024. URL: https://www.kaggle.com/datasets/emmarex/plantdisease, accessed on 1 September 2024.