Smart Agriculture: Utilizing Image Processing for Real-Time Crop Disease Monitoring

Ms. Priyanka Lalasaheb Naikal[1] , Ms. Sujata Gaikwad [2] , Mr. Parmeshwar Suresh Deshmukh[3]
[2] TPCT's College of Engineering , Osmanabad (Dharashiv), India.
priyankanaikal12@gmail.com, sujatagaikwad414@gmail.com, psdeshmukh111995@gmail.com

Abstract: The K-Means clustering technique is a widely recognized method for low-level image segmentation challenges, converging iteratively to refine partitioning decisions based on an initial user-specified cluster set that updates with each iteration. Initially, our approach identifies predominantly green-colored pixels in the image, which are then masked based on threshold values computed using Otsu's method. Subsequent masking eliminates any remaining predominantly green pixels. Additionally, pixels with zero values for red, green, and blue channels, as well as those along the boundaries of the infected cluster, are systematically removed. Experimental findings highlight the robustness of this technique in detecting plant leaf diseases, demonstrating its efficacy in real-world scenarios.

Keywords: Image Segmentation, K-Means clustering, Crop disease detection, Image processing, Agriculture, Machine learning,

I. INTRODUCTION

India is an agricultural country where about 70% of the population depends on farming. Farmers grow a variety of fruits and vegetables, but achieving the best yield and quality requires technical knowledge, which can be improved with technology. Managing fruit crops, especially by monitoring for diseases, is crucial as they can greatly affect production and post-harvest quality. In plants, a disease is any problem that impairs normal functions and produces symptoms, which are signs of the disease. Diseases are caused by pathogens, which are disease-causing agents, and are often visible on the leaves or stems. Therefore, identifying plants, leaves, and stems, and determining the extent of pest or disease presence is important for successful crop cultivation.

In biological research, experiments can generate thousands of images that are needed for studies like classifying lesions, scoring traits, and measuring areas eaten by insects. Most of these tasks are done manually or with different software, which involves a lot of work and has two major problems: it takes a long time and results can vary between individuals. Therefore, plant biologists need efficient computer software to automatically extract and analyze important information. This is where image processing becomes important.

II.LITERATURE SURVEY

In paper [1] authors focused on Rice disease identification and considered the two diseases, namely Leaf Blast & Brown Spot. Boundary detection & spot detection methods are used for feature extraction of the infected parts of plant's leaves. Authors introduced SOM (Self Organizing Map) neural network in zooming algorithm for classification of rice diseased images. Method of making of input vector in SOM is padding of zeros & interpolation of missing points, zooming algorithm gives satisfactory result.

In paper [2] authors considered five plant diseases namely Late scorch, Cottony mold, Early scorch, Ashen mold and Tiny whiteness from Jordan's Al-Ghor area for testing. K-Means clustering method is used for segmentation of leaf images and the CCM (Colour Co-occurrence Method) method is used for infected leaf texture analysis. For classification of plant diseases, back propagation algorithm in neural network is used.

In paper [3] authors used LABVIEW vision & MATLAB for detection of chili plant disease. Leaf inspection in early stage is possible due to combined technique of two software's . The LABVIEW is used for capturing images of leaf and MATLAB is used as image processing software. Edge detection, Fourier filtering, morphological operations are done with help of image pre-processing and color clustering method is used for separating chili and non-chili leaves in feature extractions. Image recognition and the classification shows chili plant healthiness.

In paper [4] authors introduced technique for detection of mauls Domestic leaves disease. Gray scale images are obtained by histogram equalization and the texture analysis in image segmentation is done with help of co-occurrence matrix method algorithm also color analysis is obtained using K-means clustering algorithm. In threshold matching process, there is comparison between individual pixels value and threshold value. For detection of plant diseases, texture & color images are compared with previously obtained images of leaf.

In paper [5] authors described technique for detection of Bacterial leaf scorch infection in plant. In image segmentation, K-means clustering algorithm is applied for separating foreground and background images. Clustering in segmentation is based on subtracting the clustered leaf images and intensity mapping for highlighting leaf area. K-means is very effective and simple for detection of infected area.

In paper [6] authors introduced technique of Citrus leaf disease detection and diseases are: Anthracnose, Citrus canker, Overwatering and Citrus greening. Image pre-processing involved color space conversion by applying YCbCr color system & L*a*b* color space also color image enhancement by applying discrete cosine transform. Gray-Level Co-Occurrence Matrix is used for feature extraction to see various statistics such as energy, contrast, homogeneity and entropy. Lastly SVMRBF and SVMPOLY are used for citrus leaf diseases detection.

In paper [7] authors presented technique for detection of Sun scorch Orchid Black leaf & spot leaf disease. Preprocessing is obtained by histogram equalization, intensity adjustment and filtering for image enhancement. Segmentation involved thresholding process and three morphological processes which are applied for removing & preserving the small & large object respectively. Finally classification is done by calculation of white pixels in leaf image and diseases are recognized.
In paper [8] authors described technique of Tomato leaves diseases detection and diseases are: Powdery mildew & Early blight. Image pre-processing involved various techniques such as smoothness, remove noise, image resizing, image isolation and background removing for image enhancement. Gabor wavelet transformation is applied in feature extraction for feature vectors also in classification. Cauchy Kernel, Laplacian Kernel and Invmult Kernel are applied in SVM for output decision & training for disease identification.

In paper [9] authors presented technique in which pre-processing involved conversion RGB images to grey using the equation $f(x)=0.2989*R+0.5870*G+0.114*B$ and removing objects and noise in image. Boundary & spot detection algorithms are configured in segmentation to find leaf infected part. After that H&B components and color co-occurrence methods are used to extract various features. Binary images are created from grey images by Otsu threshold algorithm and diseases are classified and identified using both artificial neural network and back propagation network along with K-means method.

In paper [10] authors described technique to detect Spot & Scorch disease in which by creating color transformation structure, color values are converted to space value in image pre-processing. Masked cells inside the boundaries are removed by masking of green-pixels after applying K-means method. Color co-occurrence method extracts the features such as color, texture & edge and lastly neural network is used for recognition and disease classification.

III. PROPOSED SYSTEM

In 1967, Macqueen proposed the K-Means clustering algorithm, a popular method for grouping data into clusters. This algorithm is widely used because it is efficient, scalable, and quickly converges, especially with large datasets. However, traditional K-Means has some problems, such as needing to set the number of clusters (K) beforehand, randomly picking initial cluster centers, and being affected by noisy data.

To fix these problems, this paper introduces an improved K-Means algorithm with a noise filter. This enhanced algorithm uses methods based on data density to detect and handle noise. It adds steps to identify and remove noisy data within the original algorithm. By filtering out noise before clustering, the algorithm improves the quality of the clusters and reduces the impact of noise, resulting in more accurate clustering results.

The proposed methodology involves several steps:

1. RGB image acquisition.
2. Create the color transformation structure.
3. Convert the color values in RGB to the space specified in the Color transformation structure.
4. Apply K-means clustering.
5. Masking green-pixels.
6. Remove the masked cells inside the boundaries of the infected Clusters.
7. Convert the infected (cluster / clusters) from RGB to HIS Translation.
8. SGDM Matrix Generation for H and S.
9. Calling the GLCM function to calculate the features.
10. Texture Statistics Computation.

By following this methodology, the algorithm effectively addresses the challenges posed by noise data in K Means clustering, leading to improved clustering accuracy and robustness.

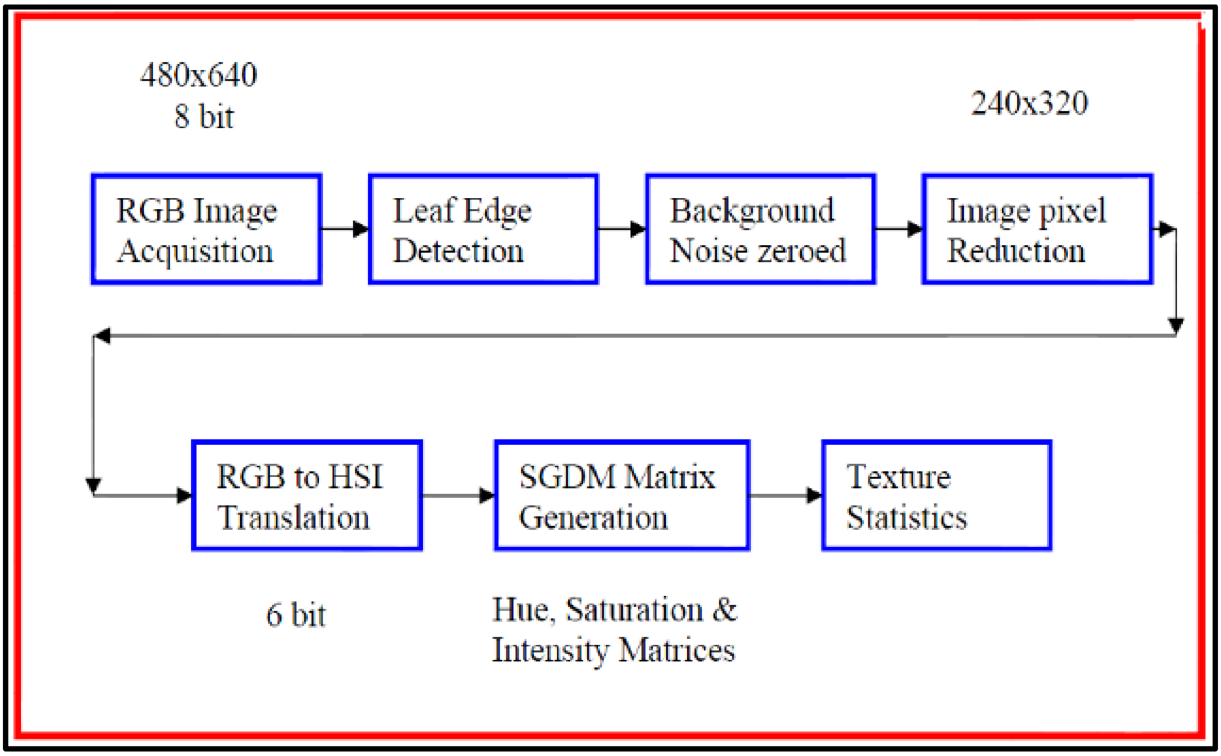

Fig 1. Image acquisition and Classification Flow chart

Our proposed step-by-step image segmentation and recognition process is presented in Algorithm 1. In the initial step, RGB images of all leaf samples were selected. Some real samples of these diseases are shown in Figure 2. It's evident from Figure 2 that leaves belonging to early scorch, cottony mold, ashen mold, and late scorch exhibit significant differences from greasy spot leaves in terms of color and texture. Additionally, the figure shows two images; the left one is infected with tiny whiteness disease, and the right one is a normal image. However, leaves related to these six classes (early scorch, cottony mold, ashen mold, late scorch, tiny whiteness, and normal) have very subtle differences discernible to the human eye, which might justify misclassifications based on naked eye observation.

In detail, in Step 2, a color transformation structure is created for the RGB leaf image, and then, a device-independent color space transformation is applied to the color transformation structure in Step 3. Steps 2 and 3 are necessary for carrying out Step 4, where the images are segmented using the K-Means clustering technique. These four steps constitute Phase 1, where the infected object(s) is/are determined. In Step 5, we identify mostly green-colored pixels.

Subsequently, based on specified and varying threshold values computed for these pixels using Otsu's method, these mostly green pixels are masked as follows: if the green component of pixel intensities is less than the pre-computed threshold value, the red, green, and blue components of this pixel are assigned a value of zero. This is done because these pixels have no significant contribution to disease identification and classification and likely represent healthy areas in the leaf. Moreover, this reduces the overall image processing time. In Step 6, pixels with zero red, green, and blue values and pixels on the boundaries of the infected cluster (object) are completely removed. Steps 5 and 6 form Phase 2, which enhances disease classification and identification accuracy while significantly reducing computation time. The rationale behind Steps 5 and 6 was experimentally validated.

Next, in Step 7, the infected cluster is converted from RGB format to HSI format. In the subsequent step, SGDM matrices are generated for each pixel map of the image for only H and S images. SGDM is a measure of the probability that a given pixel at one particular gray-level will occur at a distinct distance and orientation angle from another pixel, given that pixel has a second particular gray-level. From the SGDM matrices, texture statistics for each image are generated. In essence, the feature set is computed only for pixels inside the boundary of the infected areas of the leaf. In other words, healthy areas inside the infected areas are also removed. Steps 7 – 10 form Phase 3, where texture features for the segmented infected objects are calculated.

Finally, the recognition process in Phase 4 is performed on the extracted features through a pre-trained neural network. For each image in the dataset, the subsequent steps in Algorithm 1 are repeated. The image data of the selected leaves for this study would be collected. Algorithms based on image processing techniques for feature extraction and classification would be designed. Manual feeding of the datasets, in the form of digitized RGB color photographs, would be performed for feature extraction and training the SAS statistical classifier. After training the SAS classifier, the test datasets would be used to analyze the performance of accurate classification. This entire analysis procedure would be replicated for three alternate classification approaches, including statistical classifiers using the Mahalanobis minimum distance method, neural network-based classifiers using the back propagation algorithm, and neural network-based classifiers using radial basis functions. The results obtained from the three approaches would be compared, and the best approach for the problem at hand would be determined.

K-means clustering algorithm

K-means clustering algorithm is simply described as follows: Input: N objects to be cluster (x1, x2, xn), the number of clusters k; Output: k clusters and the sum of dissimilarity between each object and its nearest cluster centre is the smallest;

1. Arbitrarily select k objects as initial cluster centres (m1, m2 ... mk);
2. Calculate the distance between each object Xi and each cluster centre, and then assign each object to the nearest

cluster, formula for calculating distance as:

$$d(x_i,m_i)=\sqrt{\sum_{j=1}^{d}(x_{i1}-m_{j1})^2}\text{ , i=1… N; j=1… k;}$$

d (Xi, mJ) is the distance between data i and cluster j;

3. Calculate the mean of objects in each cluster as the new cluster centers,

$$m_i=\frac{1}{N_i}\sum_{j=1}^{N_i}x_{ij}\text{ , i=1, 2… k;}$$

Ni is the number of samples of current cluster i;

4. Repeat 2 & 3 until the criterion function E converged, return (m),m2 . . . mk).

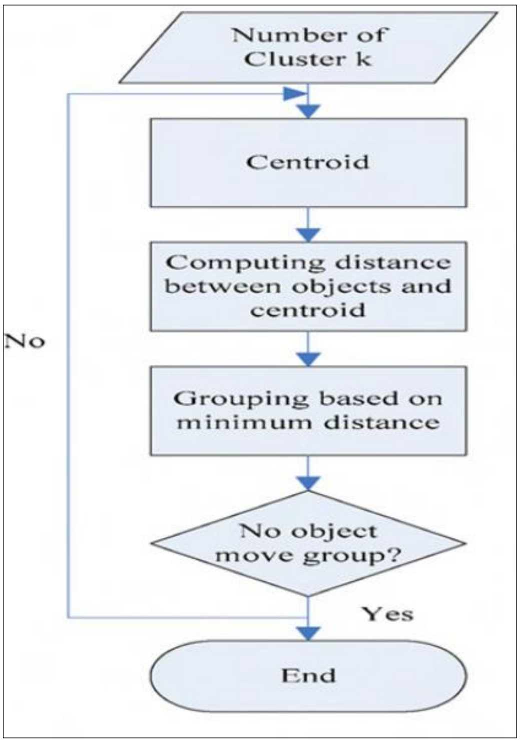

Fig 3. Flow chart for K Means Clustering

Advantages of K-Means Clustering:

1. **Scalability and Efficiency:** The algorithm demonstrates scalability and efficiency in processing large datasets. Its computational complexity is represented by O(nkt), where n is the number of objects, k is the number of clusters, and t is the number of iterations. This efficiency makes it suitable for handling large volumes of data.
2. **Effective with Compact and Well-Separated Clusters**: K-Means performs well when dealing with clusters that are compact and well-separated from each other. This characteristic makes it particularly suitable for datasets where clusters exhibit clear boundaries.
3. **Simplicity and Interpretability:** The algorithm is simple to understand and implement, making it accessible to users with varying levels of expertise. Additionally, the results produced by K-Means are easily interpretable, allowing for straightforward analysis and decision-making. Moreover, it can be adapted to handle streaming data, offering flexibility in real-time data processing scenarios.
4. **Continuous Improvement and Generalization:** K-Means has undergone continual improvements and generalizations over time, ensuring its relevance and effectiveness across various applications. These advancements have contributed to its versatility and applicability in diverse domains, making it a widely used clustering algorithm.

EXISTING SYSTEM

The Crop disease detection in image processing involves the use of computer algorithms to analyze images of crops and identify signs of diseases or abnormalities. These algorithms utilize techniques such as machine learning and deep learning to recognize patterns in the images that correspond to known diseases. The process typically begins with acquiring images of crops using cameras or drones, which are then processed by the system. Through this analysis, the system can provide farmers with early warnings about potential diseases affecting their crops, allowing for timely intervention and management strategies to mitigate crop losses.

RESULTS

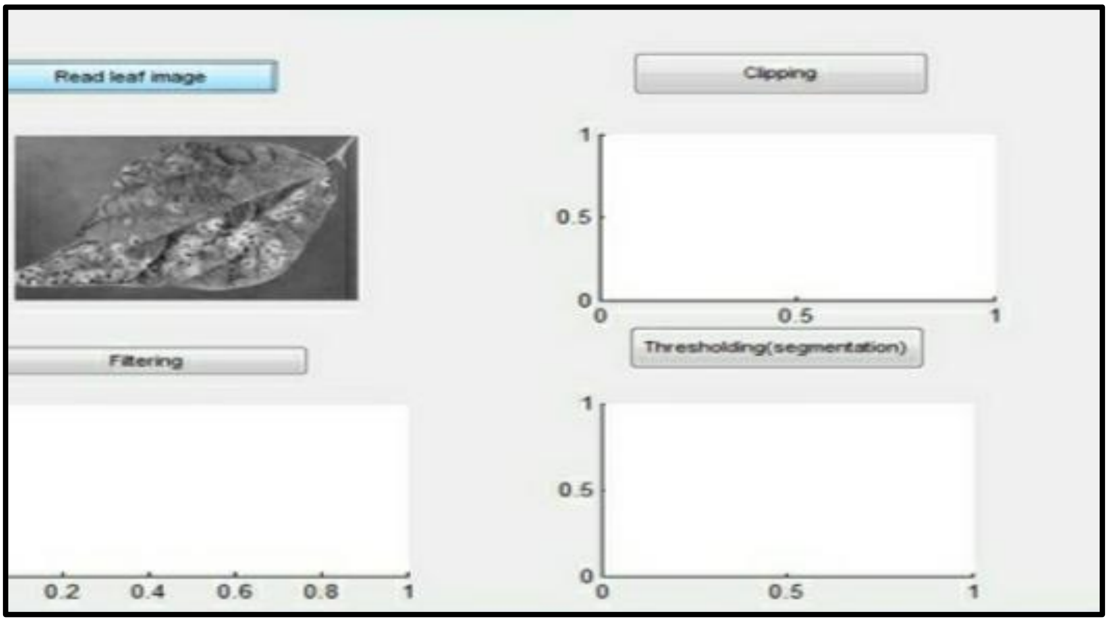

Fig 4.Taking infected image as input

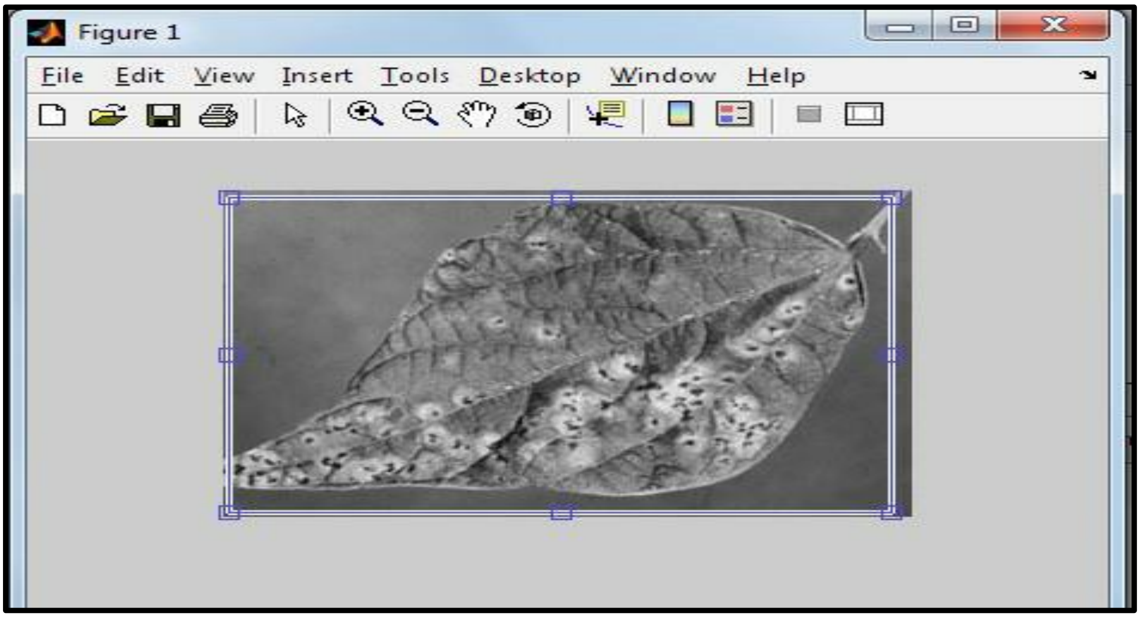

Fig 5. Selected Crop section

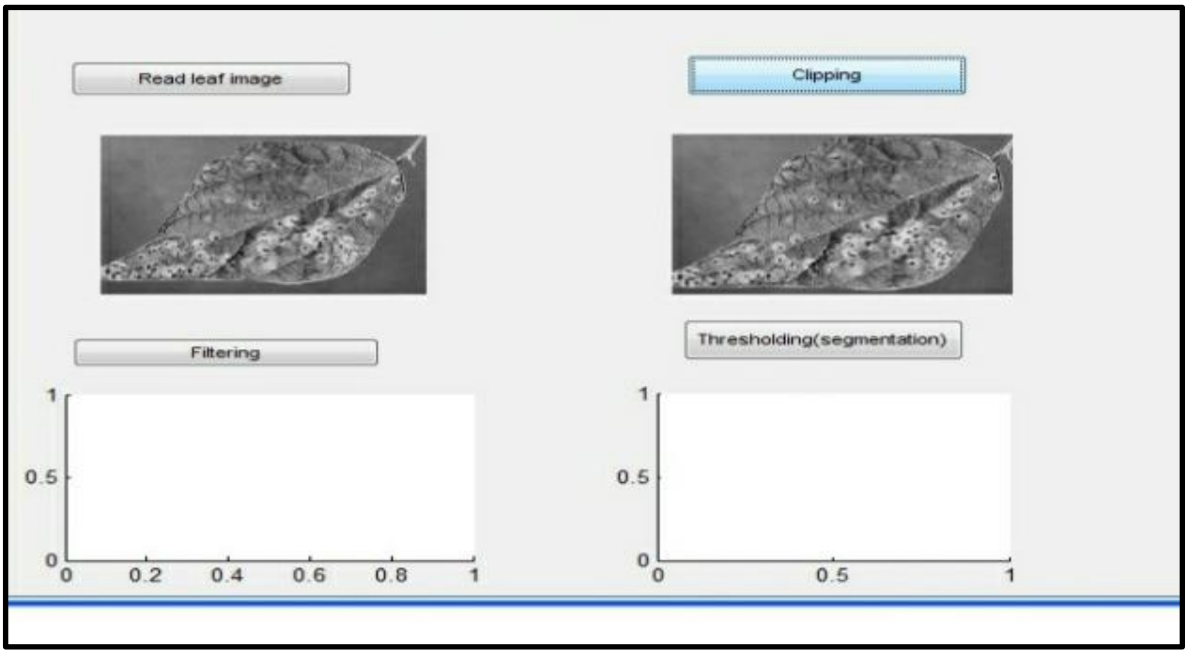

Fig 6. Clipping section of diseased Crop.

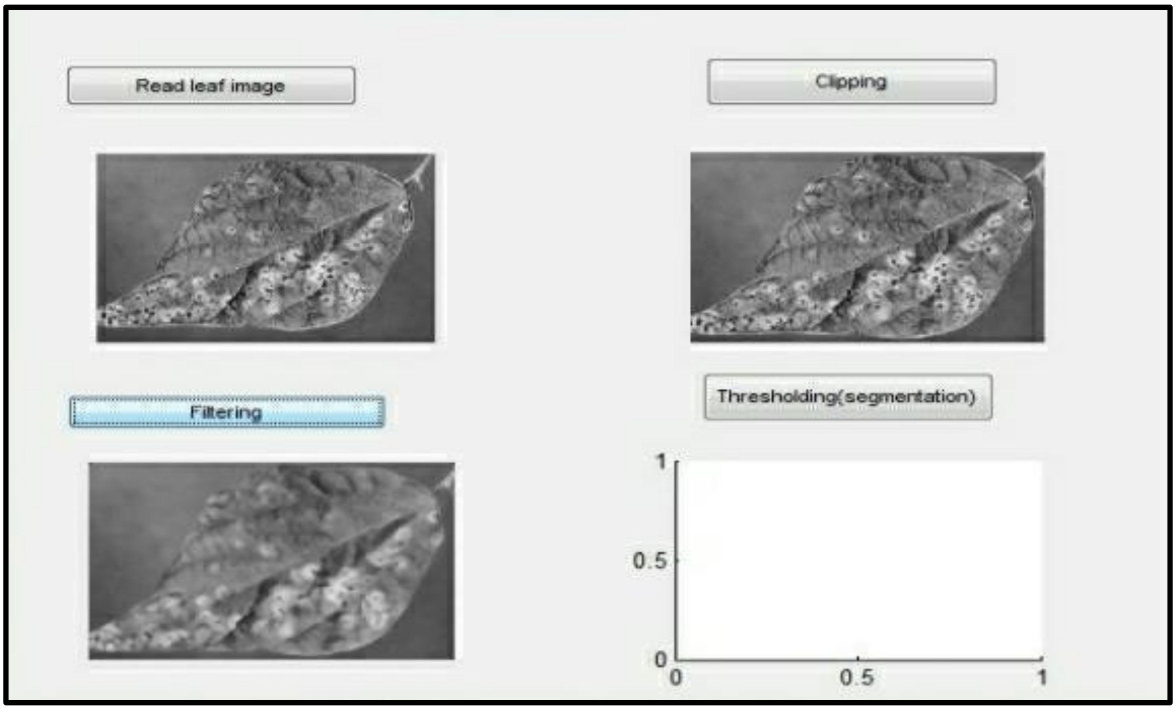

Fig 7 Filtering of diseased crop.

Segmented by k-means algorithm.

CONCLUSION

I introduce a flexible clustering algorithm derived from k-means, adept at discerning inherent clusters within datasets, be they in the original space or subspaces. Like traditional k-means, our method maintains linear time complexity relative to data points, data dimensionality, and cluster count. Empirical findings underscore its efficacy and precision in clustering analysis. Given that clustering analysis is pivotal in data mining, the selection of a suitable algorithm profoundly influences clustering outcome quality.

FUTURE SCOPE

The general k-means-based clustering algorithm showcases its ability to discern natural clusters within datasets, regardless of their presence in the original space or subspaces. Similar to traditional k-means algorithms, our approach maintains a time complexity that scales linearly with the number of data points, the dimensionality of the data, and the number of clusters. Our experimental results underscore the efficiency and high accuracy of our algorithm. Comparing standard and genetic versions of k-means algorithms reveals that standard versions excel in discovering solutions with high fitness, while both versions perform similarly in terms of validity indices. Interestingly, genetic versions often yield solutions with slightly lower fitness values but demonstrate exceptionally high values in individual validity indices during extensive exploration of the solution space. This observation hints at a potential avenue for enhancing k-means-based image clustering techniques, laying groundwork for future improvements in such methods.

REFERENCES

1. Santanu Phadikar and Jaya Sil “Rice Disease identification using Pattern Recognition Techniques” IEEE Proceedings of 11th InternationalConference on Computer and Information Technology (ICCIT 2008), Khulna, Bangladesh, pp. 1-4244-2136-7/08, 25-27 December, 2008.
2. Dheeb Al Bashish, Malik Braik and Sulieman Bani-Ahmad “A Framework for Detection and Classification of Plant Leaf and Stem Diseases” IEEE International Conference on Signal and Image Processing, pp. 978-1-4244-8594-9/10, 2010.
3. Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff and Rohani Binti S Mohamed Farook “Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques” IEEE Third International Conference on Intelligent Systems Modelling and Simulation, pp. 978-0-7695-4668-1/12, 2012.
4. Sabah Bashir and Navdeep Sharma “Remote Area Plant Disease Detection Using Image Processing” IOSR Journal of Electronics and Communication Engineering (IOSRJECE) ISSN : 2278-2834 Volume 2, Issue 6, PP 31-34,Sep-Oct 2012.
5. [Murali Krishnan and Dr.M.G.Sumithra “A Novel Algorithm for Detecting Bacterial Leaf Scorch (BLS) of Shade Trees Using Image Processing” IEEE 11th Malaysia International Conference on Communications, Kuala Lumpur, Malaysia pp. 978-1-4799-1532-3/13, 26th -28th November 2013.
6. Ms. Kiran R. Gavhale, Prof. Ujwalla Gawande and Mr. Kamal O. Hajari “Unhealthy Region of Citrus Leaf Detection Using Image Processing Techniques” IEEE International Conference for Convergence of Technology, pp. 978-1-4799-3759-2/14, 2014.
7. Wan Mohd Fadzil W.M.N, Shah Rizam M.S.B and R. Jailani, Nooritawati M.T “Orchid Leaf Disease Detection using Border Segmentation Techniques” IEEE Conference on Systems, Process and Control (ICSPC 2014), Kuala Lumpur, Malaysia, pp. 978-1-4799-6106-1/14, 12-14 December 2014. ISSN(Online) : 2319-8753 ISSN (Print) : 2347-6710 International Journal of Innovative Research in Science, Engineering and Technology (An I SO 3 2 9 7 : 2 0 0 7 Cert if ied Organizat ion) Website: w w w . ij irset .com Vol. 6 , I ssue 6, June 20 1 7 Copyright to IJIRSET DOI:10.15680/IJIRSET.2017.0606034 10381
8. Usama Mokhtar, Mona A. S. Alit, Aboul Ella Hassenian, Hesham Hefny “Tomato leaves diseases detection approach based on support vector machines” IEEE pp. 978-1-5090-0275-7/15, 2015.

9. Sachin D. Khirade, A. B. Patil, “Plant Disease Detection Using Image Processing” IEEE International Conference on Computing Communication Control and Automation, pp. 978-1-4799-6892-3/15, 2015.
10. Ghulam Mustafa Choudhary and Vikrant Gulati “Advance in Image Processing for Detection of Plant Diseases” International Journal of Advanced Research in Computer Science and Software Engineering 5(7), [ISSN: 2277 128X], pp. 1090-1093, July- 2015.
11. Digital image processing Using MATLAB codes by Dhananjay Theckedath Tech-Max Publication.
12. Ali, S. A., Sulaiman, N., Mustapha, A. and Mustapha, N., (2009). K-means clustering to improve the accuracy of decision tree response classification. Inform. Technol. J., 8: 1256-1262. DOI: 10.3923/itj.2009.1256.1262
13. Hillnhuetter, C. and A.-K. Mahlein, Early detection and localisation of sugar beet diseases: new approaches, Gesunde Pfianzen 60 (4) (2008) Jaware et al., 194
14. Otsu, N. (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9: 62–66.DOI:10.1109.