International Journal of Mathematics and
Computer Science, **16**(2021), no. 4, 1217–1228

Early Blight Disease Segmentation on Tomato Plant Using K-means Algorithm with Swarm Intelligence-based Algorithm

Syaiful Anam, Zuraidah Fitriah

Department of Mathematics
Faculty of Mathematics and Natural Science
Brawijaya University
Malang, Indonesia

email: syaiful@ub.ac.id

(Received January 19, 2021, Accepted February 23, 2021)

Abstract

Early blight commonly attacks the tomato leaf. The drone and image processing technologies offer a tool to detect the symptom of the tomato disease. One of the stages in the early blight detection is the segmentation of the tomato leaf. The K-means algorithm is a known image segmentation method which works simply and quickly. However, randomized initialization of centroids causes the K-means algorithm to easily get stuck in the local optimum and, as a result, gives imprecise segmentation. The swarm intelligence-based algorithm can avoid this problem. For this reason, we propose the early blight disease segmentation method on tomato leaves uses K-means algorithm with swarm intelligence-based algorithm. This research uses the Particle Swarm Optimization (PSO) because PSO is one of the swarm intelligence-based algorithms which has a balance in exploration and exploitation. The Hue of the HSV color space is used as input. From the experimental results, we can conclude that the performance of the early blight disease segmentation method using the K-means algorithm with swarm intelligence-based algorithm is much better.

Key words and phrases: Swarm intelligence, Particle Swarm Optimization, image segmentation, early blight disease, tomato leaves.
AMS (MOS) Subject Classifications: 68W50, 68T45, 68T10.
ISSN 1814-0432, 2021, http://ijmcs.future-in-tech.net

1 Introduction

The tomato is one of the most widely cultivated agrarian products in a tropical country, especially in Indonesia. The most frequently occurring diseases are those of the tomato leaves, such as early blight diseases [1]. Controlling early blight diseases in tomato plants on large land is a very time-consuming and labor-intensive task. Therefore, it is quite important to control tomato diseases by utilizing smart farming 4.0 [2].

The tomato leaf images acquired by the drone are analyzed to detect the disease. Leaf image segmentation is one of the stages in monitoring the health of tomato plants. Segmenting the image of the tomato leaf is one of the important steps to decide the types of diseases that attack tomato plants and to measure the extent of the tomato leaves that are affected by diseases.

Image segmentation methods have been applied in various fields such as medicine [3, 4, 5] and agriculture [6]. The clustering-based methods have been applied for many image segmentation applications. For this reason, a clustering based algorithm for early blight segmentation will be studied in this article.

The K-mean algorithm is a known clustering method [7, 8] which is easy to implement and run and its computation time is fast and easy to adapt. However, randomized initialization of the centroids causes the algorithm to be stuck in the local optima [9].

The swarm intelligence-based algorithm can find the global optimum from many local optima, does not require derivatives, is robust, and easy to apply [10]. For this reason, this papers proposes the early blight disease segmentation method using K-means clustering with swarm intelligence-based algorithm. In this article, we consider Particle Swarm Optimization (PSO) since it is one of the good swarm intelligence-based algorithm which gets the global optimum easily.

2 Material and Method

2.1 Sample Data Preparation

The research data taken from the internet are the tomato leaf images infected by early blight disease. This study only focuses on separating the healthy and diseased parts of the leaves. The separation of the leaves and the leaf background will be focused on in further research. In this study, using the six images shown in Figure 1 which have various colors and varied lighting.

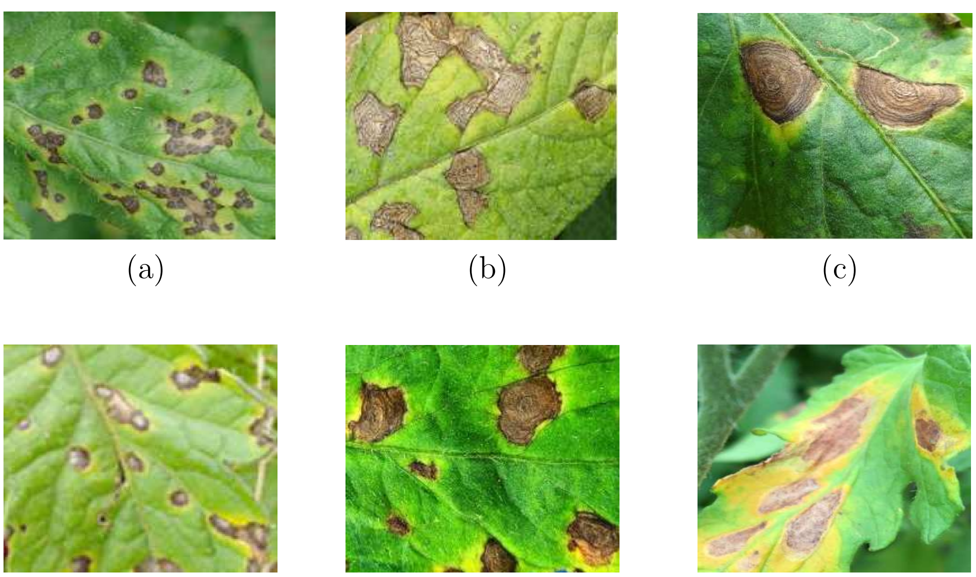
The varied leaf image is intended so that we can measure the robustness of the proposed segmentation method.

2.2 Preprocessing

Data from tomato leaf images cannot be processed directly by the proposed method because the images still require feature extraction to distinguish healthy and diseased leaves. The characteristics of the leaves are influenced by the representation of the image's color space. RGB and HSV color spaces are often used in many applications. Figure 2 shows that Hue is better than others in distinguishing diseased areas from healthy ones.

2.3 Early Blight Disease Segmentation Method Using K-means Algorithm with Swarm Intelligence-based Algorithm

In a previous study, the disease segmentation in apple leaves with a variation level set has been proposed [11]. This method does not work well for images where the object boundaries are not clear or vague.



(a) (b) (c)

(d) (e) (f)

Figure 1: Tomato leaves with early blight disease. (a) 1st data. (b) 2nd data. (c) 3rd data. (d) 4th data. (e) 5th data. (f) 6th data

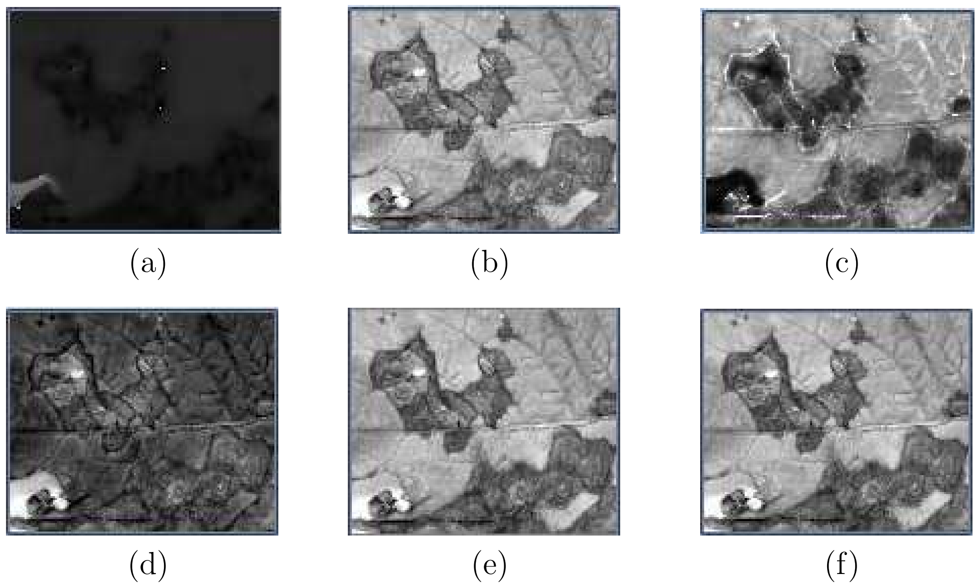


Figure 2: Component of HSV and RGB Color Spaces. (a) Hue. (b) Saturation (c) Value. (d) Red. (e) Green. (f) Blue.

One of the optimization algorithms that is often used is the Swarm Intelligence-based Algorithm which is a global optimization algorithm with good exploration capabilities [12]. The early blight disease segmentation method using K-means algorithm with swarm intelligence-based algorithm can be defined as follows.

1. Determine a feature matrix X, the number of swarms s, the maximum number of iterations ($t_{\max}$), epsilon, the number of clusters k, and the parameters c_1, c_2, w of PSO.

2. Assign the position $\mathbf{z}_i(t)$ and velocity $\mathbf{v}_i(t)$ of the swarm randomly. Each swarm position represents the candidate of the centroid of clusters.

3. The next step is that the distance between each data and each centroid is calculated. Therefore, locate each data in the nearest centroids for each swarm position.

4. Compute the fitness of each swarm by using Equations 2.1.

$$f_i = \sum_{m=1}^{k} \sum_{\mathbf{X}_l \in CL_m} d\left(\mathbf{X}_l, \mathbf{z}_{i,m}\right) \tag{2.1}$$

f_i is i^{th} fitness value of swarm. $d(\mathbf{X}_l, \mathbf{z}_{i,m})$ represents the distance between data to l to the swarm position i by considering centroids m.

5. Therefore, the best swarm position and the best global position of all swarms are updated.
6. The position and velocity of each swarm are updated.
7. The variance of the fitness values of all swarm is calculated.
8. When we get convergence in the first step, matrix IDX is created based on clustering results. The results of the program are a matrix IDX, the global best of fitness value, and the computational time.

2.4 Evaluation

The objective function values is used for evaluation. In addition, the success of the program in segmenting early blight diseases was measured by comparing the results of the disease segmentation of the proposed method with manual disease segmentation. The segmentation results are evaluated by measuring the F-measure [13]. Recall is a measure of the ratio of the number of data classified correctly in the positive class with the amount of data predicted to enter the positive class. Precision is a measure of the ratio of the number of data correctly classified in the positive class to the number of data in the true positive class. F_{measure} provides a single score that balances the attention of precision and recall in one number.

2.5 Experimental Result and Discussion

In this section, we will show and discuss the results of the experiments. Figure 3 shows that Hue can differentiate the leaf with disease areas from the healthy ones for almost all dataset. The color of the leaf affected by the disease are black while that of the healthy one is gray. Therefore, Hue is used for segmentation process.

In this study, the experiments were repeated 15 times for each data because both the segmentation method by the K-means algorithm with swarm intelligence-based algorithm and the segmentation method by the K-means algorithm involve random numbers to find the optimum value of the objective function. Table 1 shows that the average of objective function value generated from 15 experiments by the K-means algorithm with swarm intelligence-based algorithm is smaller than the the K-means algorithm which means that the performance of the proposed method is better than the image segmentation method using the K-means algorithm.

Figure 4 shows the segmentation results of tomato leaves into healthy and diseased area using the proposed method, while the results of the segmentation of early blight disease using the K-means algorithm are shown in Figure 5. These figures show that the proposed method gives better results than the early blight disease segmentation method by using the K-means algorithm. The segmentation method by the K-means algorithm usually gives mixed results because the the K-means algorithm becomes trapped at a local optimum and so is unable to segment healthy leaves and diseased ones as shown in Figures 5.

Table 2 shows that the values of F_{measure} of the proposed method are not much different between experiments; this is indicated by the small standard deviation of each, namely 6.427E-06, 2.681E-16, 3.772E-06 and 2.873 E-16. Meanwhile, the average of standard deviations of F_{measure} from the K-means algorithm are 0.339, 0.370, 0.288 and 0.330, respectively which shows that the results of the K-means algorithm between experiments differ greatly and are trapped at local optima. Table 2 also shows that the value of F_{measure} of the proposed method is 0.900 while that of F_{measure} of the tomato leaf disease segmentation method using the K-means algorithm is 0.418. These results show that the performance of the proposed method is much better than the K-means algorithm.

The average of computation time of the proposed method was 142.062 seconds and the tomato leaf disease segmentation method using the K-means algorithm was 7.184 seconds. Although the proposed method is slower, the F_{measure} performance is much better.

Table 1: The average and standard deviation of F_{measure}

Data	**The early blight disease segmentation method using the K-means algorithm with swarm intelligence-based algorithm**		**The early blight disease segmentation based on the K-means method**	
	Average	**Standard deviation**	**Average**	**Standard deviation**
1^{st} Data	0.8771	$2.263E-05$	0.257	0.300
2^{nd} Data	0.9289	$4.597E-16$	0.591	0.260
3^{rd} Data	0.8087	$0.000E+00$	0.291	0.282
4^{th} Data	0.9103	$0.000E+00$	0.580	0.361
5^{th} Data	0.9384	$3.448E-16$	0.121	0.148
6^{th} Data	0.9388	$2.298E-16$	0.669	0.376
Average	0.900	$3.772E-06$	0.418	0.288

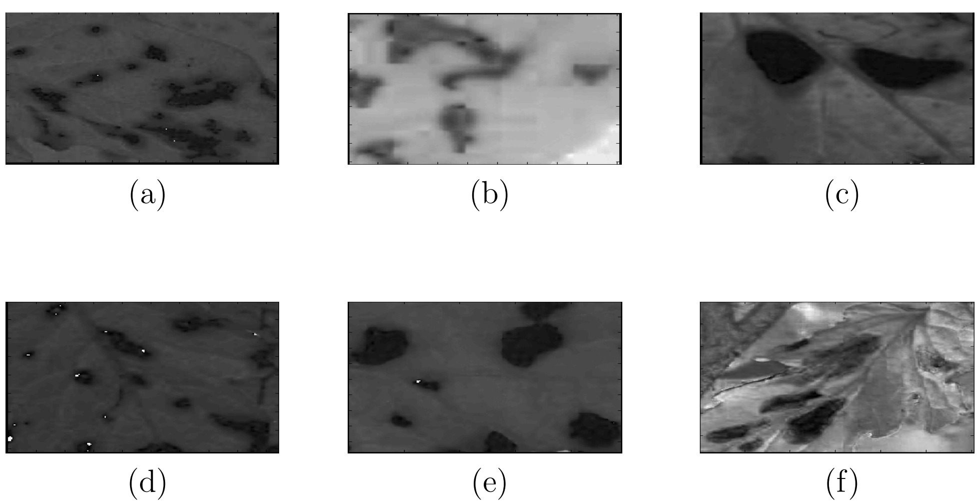


Figure 3: Hue component of the HSV color space. (a) 1^{st} data. (b) 2^{nd} data. (c) 3^{th} data. (d) 4^{th} data. (e) 5^{th} data. (f) 6^{th} data.

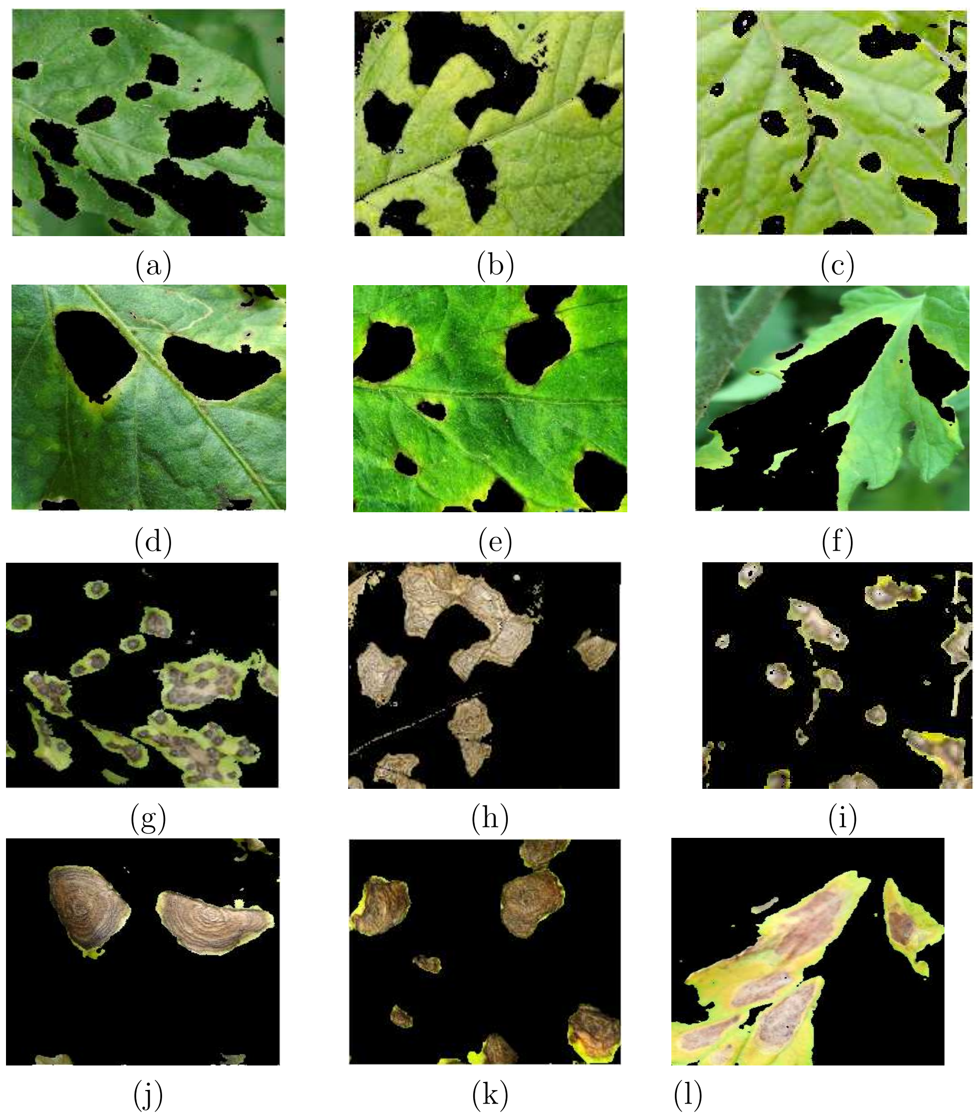


Figure 4: The results of the early blight disease segmentation method by using K-means algorithm with swarm intelligence-based algorithm. (a)-(f) Healthy area. (g)-(l) Disease area.

Table 2: The average and standard deviation of F_{measure}

Data	**The early blight disease segmentation method by using K-means algorithm with swarm intelligence-based algorithm**		**The early blight disease segmentation based K-means method**	
	Average	**Standard deviation**	**Average**	**Standard deviation**
1^{st} Data	0.8771	$2.263E-05$	0.257	0.300
2^{nd} Data	0.9289	$4.597E-16$	0.591	0.260
3^{rd} Data	0.8087	$0.000E+00$	0.291	0.282
4^{th} Data	0.9103	$0.000E+00$	0.580	0.361
5^{th} Data	0.9384	$3.448E-16$	0.121	0.148
6^{th} Data	0.9388	$2.298E-16$	0.669	0.376
Average	0.900	$3.772E-06$	0.418	0.288

Table 3: The average and standard deviation of computational time

Data	**The early blight disease segmentation method by using K-means algorithm with swarm intelligence-based algorithm**		**The early blight disease segmentation based K-means method**	
	Average	**Standard deviation**	**Average**	**Standard deviation**
1^{st} Data	149.482	32.471	5.807	2.317
2^{nd} Data	123.904	38.614	3.826	1.921
3^{rd} Data	101.678	15.530	2.927	0.865
4^{th} Data	110.515	71.284	8.471	2.863
5^{th} Data	107.872	82.7324	4.491	1.808
6^{th} Data	258.919	117.251	17.581	6.256
Average	142.062	59.647	7.184	2.672

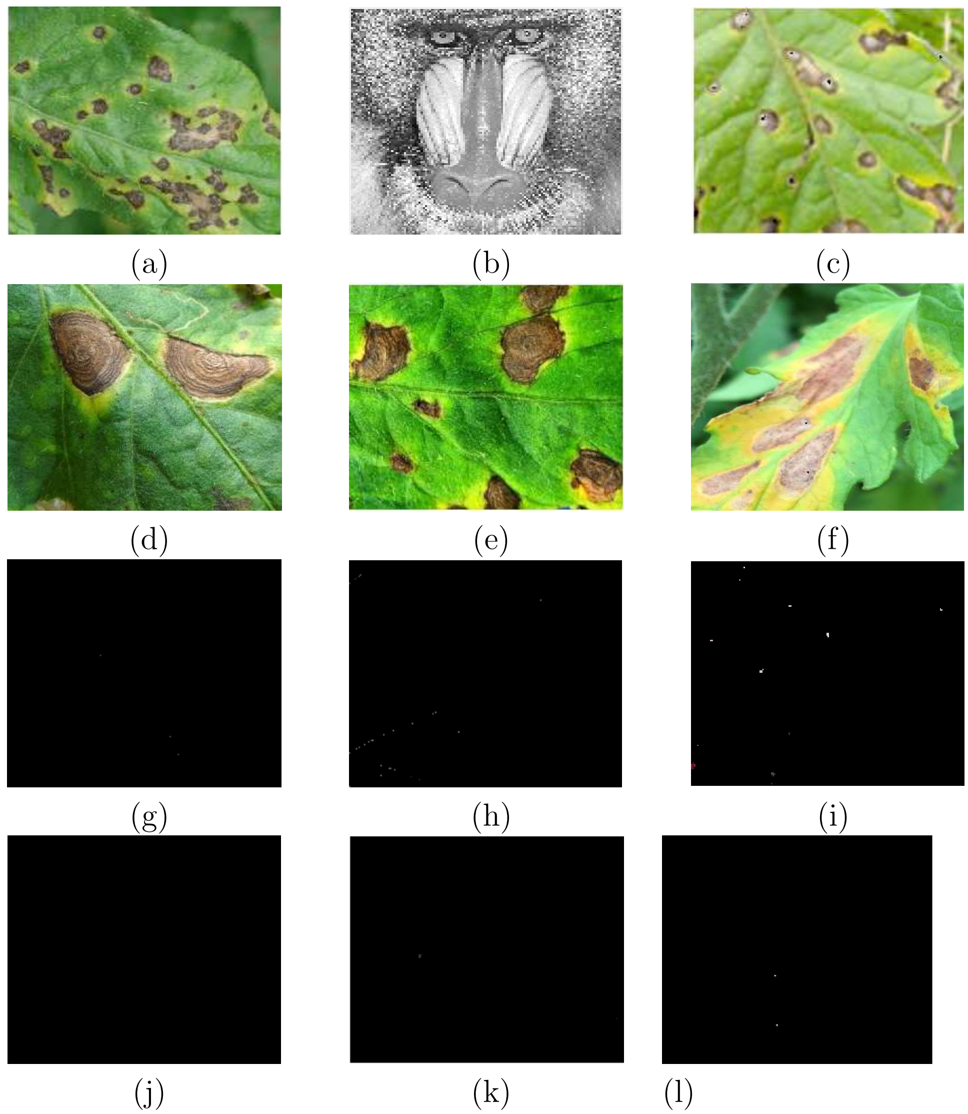


Figure 5: The worst results of the early blight disease segmentation method based K-means method. (a)-(f) Healthy area. (g)-(l) Disease area.

3 Conclusion

The early blight disease segmentation method using the K-means algorithm with swarm intelligence algorithm has successfully segmented the tomato leaf disease in almost all images used which was indicated by the high value F_{measure}. This method is much better than the tomato leaf disease segmentation method using the K-means algorithm. The computational time of this method was relatively slower than the early blight disease segmentation method using the K-means algorithm.

Acknowledgment. This research was supported by Brawijaya University through a 2020 Hibah Doktor Non Lektor Kepala Grant Program. We would like to thank the Computer and Data Science Laboratory, Mathematics Department, Brawijaya University for providing the research facility.

References

[1] J. Lu, R. Ehsani, Y. Shi, A. I. D. Castro, S. Wang, Detection of Multi-Tomato Leaf Diseases (Late Blight, Target and Bacterial Spots) in Different Stages by Using a Spectral-based Sensor, Scientific Report 8, no. 2793, (2018).

[2] R. Bongiovanni, J. L. Deboer, Precision Agriculture and Sustainability, Precision Agriculture 5, (2004), 359–387.

[3] S. Anam, E. Uchino, N. Suetake, Hybrid Boundary Detection Method for Image with Application to Coronary Plaque, Int. Journal of Digital Information and Wireless Communications, **4,** no. 4, (2014), 428–437.

[4] S. Anam, E. Uchino, N. Suetake, Coronary Plaque Boundary Enhancement in IVUS Image by Using a Modified Perona-Malik Diffusion Filter, Int. Journal of Biomedical Imaging, (2014), 1–9.

[5] Antonidoss, K. P. Kaliyamurthie, Segmentation from Images Using Adaptive Threshold, Middle-East Journal of Scientific Research, **20,** no. 4, (2014), 489–484.

[6] K. Tian, J. Li, J. Zeng, A. Evans, L. Zhang, Segmentation of Tomato Leaf Images Based on Adaptive Clustering Number of K-means Algorithm, Computers and Electronics in Agriculture, **165,** (2019), 1–7.

[7] M. Mancas, B. Gosselin, Fuzzy Tumor Segmentation Based on Iterative Watersheds, Proc. STW Conf. of Pro RISC, Veldhoven, Netherlands, (2003).

[8] K. Tian, J. Li, J. Zeng, A. Evans, L. Zhang, Segmentation of Tomato Leaf Images Based on Adaptive Clustering Number of K-means Algorithm, Computers and Electronics in Agriculture, **165,** (2019), 1–7.

[9] K. Singh, D. Malik, N. Sharma, Evolving Limitations in K-means Algorithm in Data Mining and Their Removal, International Journal of Computational Engineering & Management, **12,** (2011), 105–109.

[10] A. K. Kordon, Swarm Intelligence: the Benefits of Swarms, Applying Computational Intelligence, (2009), 145–174.

[11] S. Anam, Z. Fitriah, F. Daniyal, D. Y. Saitritama, Segmentation of Leaf Spot Disease on Apples Plants using Variational Level Set, Proceeding of the 5th IIAE International Conference on Industrial Application Engineering, Kitakyushu, Japan, (2017), 24–29.

[12] A. Ahmadyfard, H. Modares, H. Combining PSO and K-Means to Enhance Data Clustering, Proceeding of the IEEE Int. Symposium on Telecommunications, Tehran, Iran, (2008), 688–691.

[13] M. Hossin, M. N. Sulaiman, A Review on Evaluation Metrics for Data Classification Evaluations, Int. Journal of Data Mining & Knowledge Management Process, **15,** no. 2, (2015), 1–11.