**Accredited Ranking SINTA 2**
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

**JURNAL RESTI**
**(Rekayasa Sistem dan Teknologi Informasi)**
Vol. 6 No. 2 (2022) 240 - 245 ISSN Media Electronic: 2580-0760

# The Clustering Rice Plant Diseases Using Fuzzy C-Means and Genetic Algorithm

Faza Adhzima[1], Yandra Arkeman[2], Irman Hermadi[3]
[1,2,3]Department of Computer Science, FMIPA, IPB University
[1]fazaadhzima@apps.ipb.ac.id, [2]yandra@apps.ipb.ac.id, [3]irmanhermadi@apps.ipb.ac.id

***Abstract***

*Rice is an agricultural sector that is very important for Indonesia's economy. The main problem with rice plants is pest and disease control which has a very dangerous impact as well as economic losses for farmers. The characteristics that are very visible on rice leaves have a greater area than other plant structures, rice leaves can be applied for early diagnosis of rice plant diseases. Fuzzy C-Means (FCM) and Genetic Algorithm-Fuzzy C-Means are the approaches employed (GA-FCM). The center of the cluster is obtained while adopting genetic algorithms for optimization. The primary dataset used in this research is Teaching Sawah Farm IPB, and the secondary dataset is UCI Rice Leaf Diseases. According to the results of the comparison the GA-FCM optimization results in a higher level of clustering precision with a 65% optimal cluster center point on the silhoutte coefficient value compared to just 60% for FCM. This research shows the results that the proposed method can add 5% accuracy to the clustering results in terms of identifying the types of rice plant diseases properly.*

*Keywords: Clustering, Fuzzy C-Means, Genetic Algorithm, Image Processing, Rice Plants Diseases.*

## 1. Introduction

The agricultural sector in a country with a huge land area, such as Indonesia. The majority of Indonesians lives are based on rice farming as a basic need [1]. Rice is one of Indonesia's most important economic sectors. According to data from the Central Bureau of Statistics (Badan Pusat Statistik), rice farming harvested area and production decreased by 6,15% and 7,76 % respectively in 2019 compared to 2018 [2]. The main issues that exist in rice plants, namely: pest and disease control because they have very dangerous consequences as well as losses for farmers [3]. Plant Disturbing Organism (PDO) factors such as pests, diseases, and weeds are causing a decline in rice yields [4]. PDO is one of the risk factors in plant cultivation that can lead to a loss of agricultural productivity [5]. Tungro disease (tungro virus), leaf sheath blight, grass dwarf, bacterial leaf blight, stem spot, and hollow dwarf are all important diseases of rice plants in Indonesia [6].

Farmers will be impacted by diseased rice plants, and the resulting decrease in yields will have a negative impact on rice agricultural production. Spots of a specific color and pattern will appear in some parts of the infected rice plant as symptoms of this disease. The characteristics that are very visible on rice leaves have a greater area than other parts of the plant structure, they can be used for early identification of rice plant diseases [7]. For the identification of leaf diseases in rice plants, color features can be used as features/feature extraction [8]. From a infection to another, the patterns on rice leaves looked to have different shapes and colors [9]. Color, texture, and spots are some of the key factors used in the analysis of rice leaves to identify the disease type [10]. This can be controlled and handled with by maintaining the rice plant's health.

Image processing and computer vision have high potential and important role in agricultural technologies [11]. Image processing techniques may be a solution since they can measure the area of diseased leaves and identify the color difference between leaves that are attacked by pests and leaves that are not damaged by pests [12]. There are components in rice plants, such as leaf color, that can be applied as an input for digital image processing analysis [13].

Previous research have shown using map image data, Genetic Algorithm (GA) and Fuzzy C-Means (FCM) perform better than conventional FCM in clustering and validity clusters [14]. Other research has been done by Biju and Mythii are segmented microarray pictures with DNA microarray images using Genetic Algorithm

Accepted: 02-03-2022 | Received in revised: 14-04-2022 | Published: 20-04-2022

based Fuzzy C-Means [15]. In terms of grouping foreground and background pixel signals, the suggested GA-FCM algorithm is more efficient than FCM and K-Means.

Image processing in rice farming is the focus of this research. The goal of this research is to build a clustering of rice plant diseases and use genetic algorithm to optimize FCM results. Clustering of rice plant diseases using FCM and GA-FCM on rice leaf pictures is an analysis that should make it easier and more effective to discover information about the types of rice plant diseases that are attacked.

## 2. Research Methods

### 2.1 Research Dataset

The following primary datasets were used in this research, picture data of diseased rice leaves taken using a Canon EOS 700D camera. The rice plants are 65 to 75 days old which includes the reproductive phase, which is defined by the appearance of rice seeds. On November 20, 2020, the first location was in Teaching Farm Sawah Baru, Department of Agronomy and Horticulture, Faculty of Agriculture, IPB University, which totaled 60 picture objects. The image of rice leaf in Teaching Farm Sawah Baru of IPB is shown in Figure 1. The second location of rice field farm in the Cianjur area Bogor had a total of 60 picture objects was taken on December 3, 2020. The image of rice leaf in the Cianjur rice fields Bogor is shown in Figure 2.

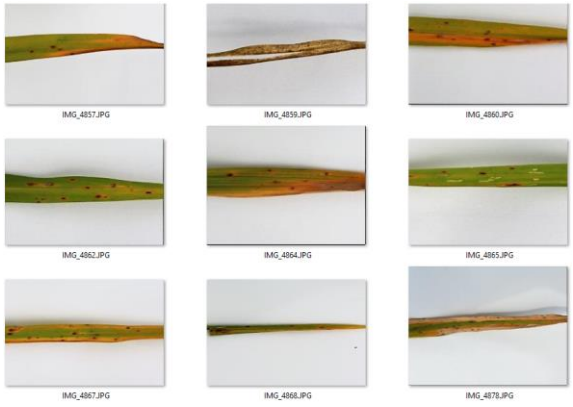

Figure 1. The Image of Rice Leaf on the Teaching Farm Sawah Baru of IPB

Rice leaf diseases secondary dataset, which was also used in this research, was obtained from UCI Machine Learning. Three types of diseases are identified with picture objects for each disease type in the leaf image collection, which has a total of 120 image objects: Bacterial Leaf Blight, Brown Spot and Leaf Smut. The image of rice leaf diseases dataset is shown in Figure 3.

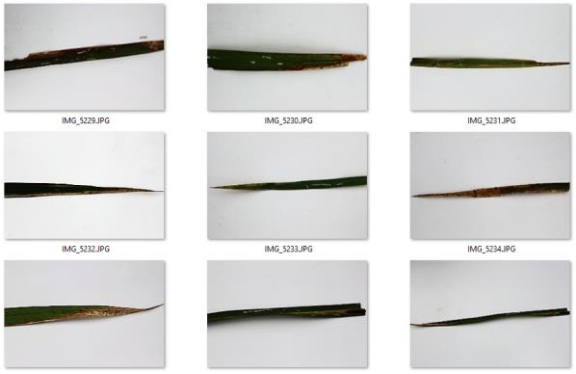

Figure 2. Image of Rice Leaf in Cianjur Rice Fields Bogor

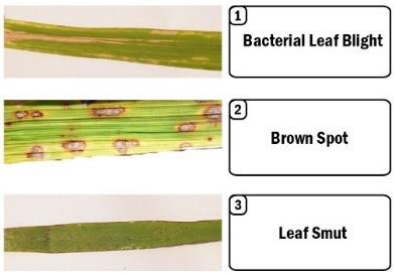


Figure 3. Rice Leaf Image in UCI Repository Rice Leaf Diseases Dataset

### 2.2 Research Stages

This research consists of several stages including: input of rice leaf image data, image preprocess, feature extraction using Gray Level Co-occurrence Matrix (GLCM), data dimension reduction using Principal Component Analysis (PCA), clustering using Fuzzy C-Means (FCM) and Genetic Algorithm - Fuzzy C-Means (GA-FCM), output model results clustering, identification of rice plant diseases and evaluation. Figure 4 shows flowchart of the research stages.

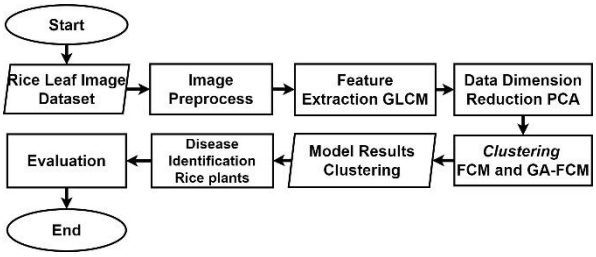


Figure 4. Research Stages Flowchart

### 1. Image Preprocess

The image is cropped to 750 220 pixels during the preprocessing stages to obtain a precise disease object by focusing on the visible blotchy lesions that will be used as a point of interest. Figure 5 shows a preprocessed image of the primary dataset.

DOI: https://doi.org/10.29207/resti.v6i2.3912

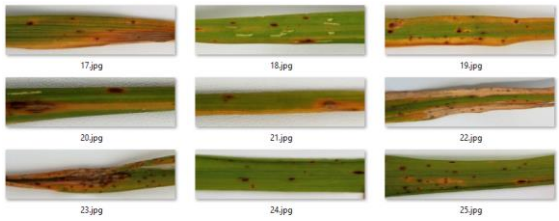


Figure 5. Image Preprocess on Primary Dataset

2. Feature Extraction

Feature extraction is the process of extracting information or features from an image in order to distinguish it from others. Texture is the most basic component of an image that can be applied as a clustering base. The size and parameters of the view, as well as the environment and lighting situation, determine the texture's image [16]. The texture analysis in this research uses Gray Level Co-occurrence Matrix (GLCM) to extract information from grayscale images in the form of energy, entropy, contrast, and homogeneity [17].

3. Data Dimension Reduction

Principle Component Analysis (PCA) or principal component analysis is a linear combination of the previous variables that is used to create new variables. PCA can also be used to extract features from an image when the image's total dimensions are more than the total sample data [18]. Another advantage of reducing the data dimensions is that it makes the model easier to understand because it is constructed with less features and it also makes the data easier to visualize [19]. Data clustering can use the value is created by the primary component of the two main new variables (principal component).

4. Fuzzy C-Means

Fuzzy C-Means (FCM) is a data grouping method in which the degree of membership determines the position of each data in a cluster. FCM is simple in nature, making it simple to apply and data [20]. The first FCM's stages determine the cluster center, that will be marked with the average cluster location. Each data has a degree of membership in each cluster, which is determined by repeatedly updating the cluster center and membership value of each data, resulting in a cluster center in the correct position. The loop is based on the objective function minimization [21].

5. Genetic Algorithm

Genetic Algorithm (GA) is a searching that adapts natural biological evolution with the aim of determining high-quality chromosomes or individuals in an environment called a population. The selection of individuals from a population will be evaluated based on the value of the fitness function. Shaped chromosomes string are candidates for each operation process known as generation [22]. The basic arrangement in GA includes chromosome representation, calculating fitness for evaluation, new individuals obtained by the crossover, mutations to increase existing variations in the population and selection to create new populations.

6. Clustering FCM and GA-FCM

Clustering also known as cluster analysis is the process of determining the correct group or class from an image based on the results of feature extraction. In FCM, the value of the cluster center's position is generated at random, results in a local optimum, which is then determined by the value generated at random, the genetic algorithm's role being to optimize the value of the initial cluster center [23]. Genetic algorithm as an optimization method can be implemented in clustering based objective function. In the genetic algorithm for fuzzy clustering, fitness function is obtained from objective function the minimized in each generation, the chromosomes will be evaluated based on the value fitness functions. In search of cluster center matrix the cluster center with operators such as selection, crossover and mutation using fitness functions.

7. Identification of Rice Plant Diseases

The type of disease is labeled at this stage using the results of the primary dataset clustering based on the three types of diseases in the secondary dataset of Rice Leaf Diseases.

8. Evaluation

At this stage, an evaluation of the results of clustering with GA-FCM is carried out using the Silhoutte Coefficient, Random Index (RI) and Adjusted Random Index (ARI) to be used to analyze the quality of the clustering in an algorithm so that an evaluation process is needed. Silhoutte coefficient aims to determine the quality cluster how well an object is positioned in a cluster. Based on the results silhoutte coefficient will have a negative weight when the distance of the data to clusters is smaller than the distance of the data to the cluster to data and will have a positive weight if the distance data to clusters is greater than the distance data to the cluster itself. ARI aims to measure the quality of the global solution by comparing the label on the results of the clustering to the predicted label. The ARI method is an extension of the RI. ARI values have a range between -1 to 1 ([-1, 1]), while for rand index has a range from 0 to 1 ([0, 1]). ARI value is the better quality of a cluster. The range of ARI values greater than the rand index can be used as a better measure of clustering [24].

## 3. Results and Discussions

The following is a summary of the results and a discussion of the stages research method used.

DOI: https://doi.org/10.29207/resti.v6i2.3912

### 3.1 GLCM Feature Extraction Results and PCA Data Reduction Results

The results of feature extraction using GLCM on the primary dataset. Obtained 4 attributes, namely: homogeneity, contrast, energy, and entropy which will then be processed for data dimension reduction using the PCA method. The results of GLCM feature extraction on the primary dataset are as shown in Table 1.

Table 1. Primary Dataset GLCM Feature Extraction Results

| Number | Homogeneity | Contrast | Energy | Entropy |
|---|---|---|---|---|
| 0 | 0.288082 | 32.330093 | 0.042218 | 4.131268 |
| 1 | 0.32581 | 86.114583 | 0.050693 | 4.135463 |
| 2 | 0.251673 | 29.09213 | 0.04822 | 3.998545 |

The results of PCA for reduction data dimensions in GLCM feature extraction are two new variables PCA 1 and PCA 2 which are then used for clustering formation as shown in Table 2.

Table 2. Results of Primary Dataset PCA Dimension Reduction

| Number | PCA 1 | PCA 2 |
|---|---|---|
| 0 | -104.65651162102685 | 0.022242572337850625 |
| 1 | -50.87203814894065 | 0.043003362072732296 |
| 2 | -107.8944973975265 | 0.14441053228450168 |

### 3.2. Results of C-Means Fuzzy Clustering Model

The primary dataset used in the FCM calculation is 120 objects, with the following parameters: number of clusters = 3, fuzzier = 2, smallest error = 0.005, and maximum iteration = 10. The number of clusters in this research is based on the number of 3 types of disease labels in the secondary dataset. 3 clusters is the number of clusters given as a parameter. The choice of the fuzzier = 2 parameter is based on Klawonn and Hoppner's research, which determined that a value of 2 for the fuzzier parameter is the optimal value within the fuzzy value limit [25]. The results of clusters in the primary dataset are 3 clusters such as cluster 0 with a total of 40 objects, cluster 1 with a total of 72 objects, cluster 2 with a total of 8 objects, as shown in Figure 6. The results of clusters in the secondary dataset are 3 clusters, namely: cluster 0 with a total 3 objects, cluster 1 with a total of 14 objects, and cluster 2 with a total of 103 objects are shown in Figure 7.

### 3.3. Genetic Algorithm Results

The total number of generations is used as a parameter in the experimental GA to obtain the best fitness value. This experiment was done with a target generation of 4000 in order to increase the results of the generation from 1 to 3000. Furthermore, the best fitness value in the 3000 generation that has had a stall generation is 9.88059480760816e-11 in the 3905 generation. The plot of the fitness value against the number of generations is shown in Figure 8.

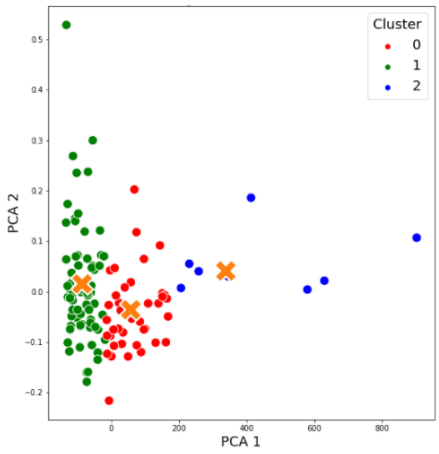


Figure 6. Primary dataset FCM Results Plot

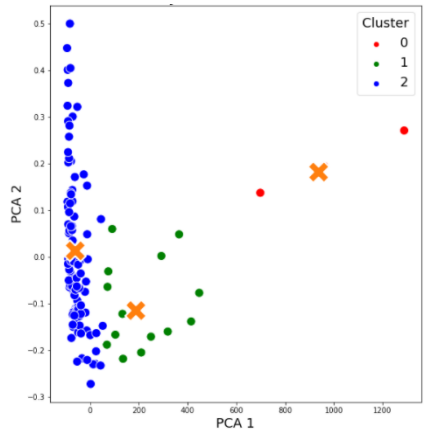


Figure 7. Secondary dataset FCM Results Plot

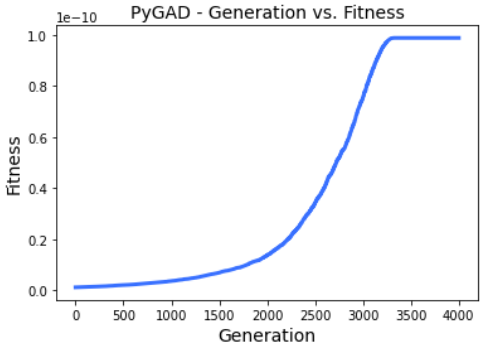


Figure 8. Plot of Fitness Value on Number of Generations

### 3.4. Results of Identification of Rice Plant Diseases

Based on the results of the secondary dataset clustering which has 3 types of disease labels, label 0 is brown spot disease, label 1 is leaf smut disease, and label 2 is bacterial leaf blight. Figure 9 shows the plot of the identification of rice diseases using FCM on primary dataset and secondary dataset.

DOI: https://doi.org/10.29207/resti.v6i2.3912

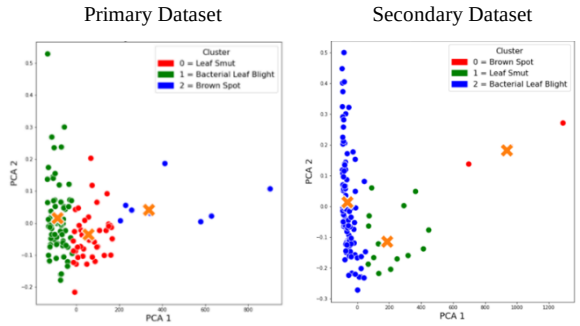


Figure 9. Plot of Rice Disease Identification Results Using FCM on Primary Dataset and Secondary Dataset

The results of the FCM cluster will be optimized using GA-FCM to determine the cluster center in the next stage. The fitness value is calculated by summing the distance between each sample and the cluster center in GA-FCM. After that, the GA-FCM plot on the primary dataset produced three labels. Cluster 0 is a leaf smut with 26 objects, cluster 1 is a bacterial leaf blight with 90 objects, and cluster 2 is a brown spot type with 4 objects. Figure 10 shows the result of optimizing the identification of rice plant diseases using GA-FCM.

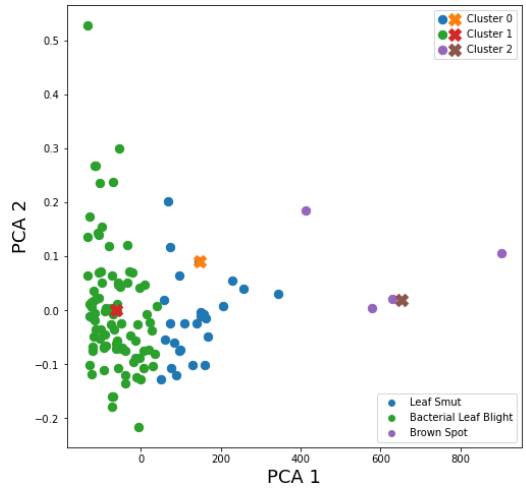


Figure 10. Optimization Results of Rice Plant Disease Identification Using GA-FCM on Primary Dataset

### 3.5. Evaluation result

Furthermore, measurements were made using the silhouette coefficient, random index score and adjusted random index score from the cluster using FCM and GA-FCM. If we look at the silhouette coefficient using GA-FCM is better because it has a value of 0.655 while using FCM with a value of only 0.601. Silhoutte coefficient is close to 1, then the feature structure used is good, so the clustering are correct. Next on random index is better with a value of 0.60294117 while FCM with a value of 0.51036414. Adjusted random index on GA-FCM is also better with a value of 0.23475880 while FCM with a value of 0.00086458. Value from random index and adjusted random index is obtained from the match of the resulting label with the target or predicted label. The larger of random index value and the adjusted random index value is the better performance of a clustering method. So the use of GA-FCM as a method that can optimize the results of clustering. Comparison of the silhouette coefficient, random index, and adjusted random index on the results of clustering FCM and GA-FCM as shown in Table 3.

Table 3. Comparison of FCM and GA-FCM Cluster Results

| Method | *Silhoutte coefficient* | *Random Index* | *Adjusted Random Index* |
|---|---|---|---|
| FCM | 0.601 | 0.51036414 | 0.00086458 |
| GA-FCM | 0.655 | 0.60294117 | 0.23475880 |

## 4. Conclusion

Based on the results of the analysis the results clustering rice plant diseases, it can be concluded that the results using FCM obtained cluster 0 (leaf smut) total 40 objects, cluster 1 (bacterial leaf blight) total 72 objects and cluster 2 total 8 objects (brown spots). Furthermore, in the optimization results using GA-FCM obtained cluster 0 (leaf smut) total 26 objects, cluster 1 (bacterial leaf blight) total 90 objects, cluster 2 (brown spot) total 4 objects.

Based on the results of the GA-FCM optimization evaluation resulted increase clustering with a silhoutte coefficient of 65% more appropriate to reach the cluster center compared to FCM with a value only 60%, the addition of GA to FCM can increase accuracy by 5%. Results Random index (RI) on GA-FCM has a value greater than 60% compared to FCM with a value only 51%. Furthermore, the GA-FCM results for the adjusted random index (ARI) is 23% greater than FCM with a value only 0.0008%.

All stages of this research have been could effort to identify the types of rice plant diseases that are still a problem in agriculture and even cause losses in rice yields. Optimization results using GA on FCM have a good increase in accuracy in determining the type of rice disease. So that this research is expected to help farmers and extension workers in taking further action in handling diseases, especially rice plants.

Furthermore, suggestions for this research in the future can be developed on IoT devices such as drones or robots so that rice fields can monitor the health status of their rice plants in real time.

## Acknowledgment

The author would like big thanks to Department of Computer Science at IPB University in terms of funding for scientific publications so that it is very helpful for the author in completing this research.

## Reference

[1] Reflis, M. Nurung, and J. Pratiwi, “Motivasi Petani Dalam Mempertahankan Sistem Tradisional Pada Usahatani Padi

DOI: https://doi.org/10.29207/resti.v6i2.3912

Sawah Di Desa Parbaju Julu Kabupaten Tapanuli Utara Propinsi Sumatera Utara", *J. AGRISEP*, vol. 10, no. 1, pp. 51 – 62, 2011.

[2] Badan Pusat Statistik, "Luas panen dan produksi padi pada tahun 2019 mengalami penurunan dibandingkan tahun 2018 masing-masing sebesar 6,15 dan 7,76 persen", *Bps.go.id*, 2020 [Online]. Avaible:https://www.bps.go.id/pressrelease/2020/02/04/1752/luas-panen-dan-produksi-padi-pada-tahun-2019-mengalami-penurunan-dibandingkan-tahun-2018-masing-masing-sebesar-6-15-dan-7-76-persen.html.

[3] Balai Besar Penelitian Tanaman Padi, Penyakit Blas pada Tanaman Padi dan Cara Pengendaliannya, Subang: Kementerian Pertanian, 2015.

[4] P. Djojosumarto, Teknik Aplikasi Pestisida Pertanian. Yogyakarta: Kanisius, 2000.

[5] Departemen Pertanian, Gerakan Pengendalian Organisme Pengganggu Tumbuhan (OPT) Serealia. Jakarta: Direktorat Jenderal Tanaman Pangan Kementerian Pertanian, 2018.

[6] H. Semangun, Penyakit-Penyakit Tanaman Pangan Di Indonesia, Yogyakarta: Gadjah Mada University Press, 2008.

[7] S. Zahrah, R. Saptono, and E. Suryani, "Identifikasi Gejala Penyakit Padi Menggunakan Operasi Morfologi Citra", *J. Seminar Nasional Ilmu Komputer*, pp. 100–106, 2016.

[8] C. Dewi, E. Anjarwati, and C. Imam, "Implementasi citra digital untuk identifikasi penyakit pada daun padi menggunakan ANFIS", *J. Seminar Nasional Penelitian & Pengabdian pada Masyarakat*, vol. 1, no. 1, pp 114-117, 2018, doi: https://doi.org/10.33019/snppm.v1i0.511.

[9] A. Fitriansyah, "Pengolahan Citra Digital Penyakit Tanaman Padi Menggunakan Metode Maksimum Entropy", *J. Semirata FMIPA*, pp. 21–23, 2013.

[10] Zhou, Yingfeng, W. Yaming, and Y. Qing, "Segmentation of Rice Disease Spot Based on Improved BPNN", *IEEE International Conference on Image Analysis and Signal Processing*, pp. 575-578, 2010, doi: 10.1109/IASP.2010.5476050.

[11] N. N. Kurniawati, S. N. H. S. Abdullah, S. Abdullah and S. Abdullah, "Texture analysis for diagnosing paddy disease", *IEEE International Conference on Electrical Engineering and Informatics*, pp. 23-27, 2009, doi: 10.1109/ICEEI.2009.5254824.

[12] S. Dhaygude, and N. Kumbhar, "Agricultural Plant Leaf Disease Detection Using Image Processing", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 1, pp. 599–602, 2013.

[13] E. D. Nurcahya, "Ekstraksi Fitur Pertumbuhan Padi Berdasar Warna Daun Menggunakan Analisa Ruang Warna Hue Saturation Value*", J. Multitek Indonesia: Jurnal Ilmiah* , vol. 13, no. 1, pp. 24–33, 2019, doi: http://dx.doi.org/10.24269/mtkind.v13i1.1515.

[14] N. Widyastuti, and H Amir, "Penggunaan Algoritma Genetika dalam Peningkatan Kinerja Fuzzy Clustering untuk Pengenalan Pola", *J. Berkala Ilmiah MIPA*, vol. 17, no. 2, pp. 1-14, 2007.

[15] V. G. Biju and P. A. Mythili, "Genetic Algorithm based Fuzzy C Mean Clustering Model for Segmenting Microarray Images", *International Journal of Computer Applications*, vol. 52, no. 11, pp. 42–48, 2012.

[16] P. Darma, Pengolahan Citra Digital, Yogyakarta: Andi, 2010.

[17] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973, doi: 10.1109/TSMC.1973.4309314.

[18] M. H. Purnomo and A. Muntasa, Konsep Pengolahan Citra Digital dan Ekstraksi Fitur, Yogyakarta: Graha Ilmu, 2010.

[19] E. Prasetyo, "Reduksi Dimensi Set Data dengan DRC pada Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga", *SNATIF*, pp. 293-300, 2014.

[20] H. Bandemer and S. Gottwald, Fuzzy Sets, Fuzzy Logic, Fuzzy Methods with Applications, England: Wiley, 1996.

[21] M. Gen and R. Cheng, Genetic Algorithms and Engineering Optimization. New York: John Wiley & Sons, 2000.

[22] H. A. Saputro, W. F. Mahmudy and C. Dewi, "Implementasi algoritma genetika untuk optimasi penggunaan lahan pertanian", *J. Mahasiswa PTIIK Universitas Brawijaya*, vol. 5, no. 12, 2015.

[23] P. E. Mas`udia and R. Wardoyo, "Optimasi Cluster Pada Fuzzy C-Means Menggunakan Algoritma Genetika Untuk Menentukan Nilai Akhir", *J. IJCCS*, vol. 6, no. 1, pp. 101-110, 2012, doi: http://dx.doi.org/10.22146/ijccs.2145.

[24] R. R. d. de Vargas and B. R. C. Bedregal, "A Way to Obtain the Quality of a Partition by Adjusted Rand Index", *Workshop-School on Theoretical Computer Science*, pp. 67-71, 2013, doi: 10.1109/WEIT.2013.33.

[25] F. Klawonn and F. Höppner, "What Is Fuzzy about Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzifier", *Advances in Intelligent Data Analysis*, vol. 2810, pp. 254-264, 2003, doi: https://doi.org/10.1007/978-3-540-45231-7_24.

DOI: https://doi.org/10.29207/resti.v6i2.3912