ISSN : 2460 – 7223

Study of Detecting Corn Plant Leaf Disease with Fuzzy C-Means and RNN

1st Enrico Findley
Master of Information Technology Department
Institute of Informatics and Business (IIB) Darmajaya
Bandar Lampung, Indonesia
enrico96.2121211004@mail.darmajaya.ac.id

2nd Suhendro Yusuf Irianto
Master of Information Technology Department
Institute of Informatics and Business (IIB) Darmajaya
Bandar Lampung, Indonesia
suhendro@darmajaya.ac.id

***Abstract*—Corn is a top commodity after rice at supporting food self-sufficiency in Indonesia. However, because of corn leaf spot disease caused by plant pests the quality and quantity of corn is greatly decreasing. The problem with detecting spot-on corn leaf is required high accuracy and the plantation is huge. Therefore, this research study to determine corn leaf spot disease using Fuzzy C-Means and RNN methods. The research process in this study is first preprocessing to CIE-L*A*B* color space, next step is doing corn plant leaf disease detection with Fuzzy C-Means and RNN methods, the third step is to reconstruct the image from Fuzzy C-Means and RNN method result in grayscale level, and the last step is to evaluate the Fuzzy C-Means and RNN algorithm. In this paper only Fuzzy C-Means segmentation and training the RNN model are implemented. The result of the experiment is first from training RNN model with 80% training data and 20% testing data. The data trained for 20 epochs with 38 minutes and 1.8 seconds in total execution time and resulting with 0.9403 for accuracy and 0.9572 for validation accuracy. Next is the Fuzzy C-Means segmentation result, the Fuzzy C-Means execution time is 94 minutes and 16 seconds. For future research the RNN can be trained with much higher epoch and for Fuzzy C-Means can be combined with classification algorithm. We hope that this study can contribute to detecting leaf spot disease for corn plant at faster rate.**

Keywords—Fuzzy C-Means, Image Segmentation, RNN, Corn Disease Detection

I. INTRODUCTION

Corn is a top commodity after rice at supporting food self-sufficiency in Indonesia. Lampung province its self is a one of central production for corn with 8.6% contribution to national production with total production can reach 2.58 million ton. Corn harvest in Lampung reaching 486.313 hectare with 5.3 ton per hectare productivity [1].

Although Lampung is producing high volume of corn, but corn consumption in this province is also high, especially for animal feed. Corn demand for animal feed in Lampung is 1.8 million ton per year which means corn production in Lampung cannot fulfill the demand for corn. Therefore, Lampung province government is always strived to increase corn production through enlargement of plantation area, and increasing cultivation intensification. Lampung government have target for national corn production in 2017 which is 20% national corn production will be originated from Lampung [2].

One factor that become problem for increasing corn plant production is an attack from plant pest organism. According to Food Crop Research and Development Center (2016) and [3] there is six main pests for corn plant which is Ostrinia furnacalis, Helicoverpa armigera, Locusta migratoria, Atherigonia, Agrotis Ipsilon, and Spodoptera litura [3].

At the end of year 2016, according to Sumardiyono there is a report from farmer stating that corn plantation in Lampung attacked by planthoppers [3]. Observation result confirmed the farmer report that there is planthoppers attacked corn plant. Different from other planthoppers, this planthoppers have white wax layer at the surface of ventral abdomen. Because of its unique characteristics this planthoppers are called white bellied planthopper (*Stenocranus pacificus kirkaldy*) [4].

According to informal information obtained from farmer state that white bellied planthoppers attack do some damage to corn. Corn that attacked by planthoppers its production is reduced to 30%, if not attacked the normal production of corn can reach 7 ton per hectare and if attacked by planthopper its production will only reach 4-5 ton per hectare [5].

For increasing the corn production, the corn farmer is already doing many things to eliminate pests. In Indonesia, there is some report about corn production result is decreasing because of corn leaf spot disease which is 5-50%. If corn leaf spot disease attacked the corn plant before its female flower appeared, then the decreasing in corn production result will decreasing to 50% [5].

Corn leaf spot detection is not efficient because the corn plantation area is huge and need high accuracy. Therefore, automatic detection for corn leaf spot disease is needed. With automatic detection, more corn leaf data can be observed with high accuracy and in a shirt time. One technique that can be performed to automatic detection according to color and shape using image processing is image segmentation. Image segmentation is already successfully proven that this technique can differentiate object among its background in image, because of that reason it is suitable for detecting corn leaf spot disease [6].

Therefore, in this research will be doing corn leaf spot disease with title detection of corn plant pest and disease with Fuzzy C-Means and RNN method.

II. METHODS

In this research detecting corn plant disease from corn leaf image will be done with Fuzzy C-Means and RNN method. These 2 methods are usually used for text processing but in this part some research that have been done implementing Fuzzy C-Means and RNN in image processing field are reviewed and to reaching more understanding about Fuzzy C-Means and RNN algorithm.

A. Preprocessing

Preprocessing is the first step taken after getting from image acquisition. This process makes the system easier to recognize the objects. So, in this process the first thing is to resize the image for faster computing process and converting the image which is in RGB color space to CIE-L*A*B* color space and the A* and B* values become color markers in the A*B* space. Before converting RGB to CIE-L*A*B* we have to convert RGB to XYZ first, we be using matrix provided by [7]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0{,}412453 & 0{,}357580 & 0{,}180423 \\ 0{,}212671 & 0{,}715160 & 0{,}072169 \\ 0{,}019334 & 0{,}119193 & 0{,}950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

From that matrix we multiply it with RGB matrix and we get the XYZ matrix. Now we have to convert it to CIE-L*A*B*, according to [7] the formulation to convert XYZ to CIE-L*A*B* is:

$$L^* = 116\left(\frac{Y}{Y_0}\right)^{1/3} - 16, \text{ where } \frac{Y}{Y_0} > 0.008856 \quad (2)$$

$$L^* = 903.3\left(\frac{Y}{Y_0}\right)^{1/3} - 16, \text{for } 0.0 \le \frac{Y}{Y_0} \le 0.008856 \quad (3)$$

$$a^* = 500\left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right)\right] \quad (4)$$

$$b^* = 200\left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right)\right] \quad (5)$$

For,

$$f(w) = w^{1/3}, \text{ where w} > 0.008856 \quad (6)$$

$$f(w) = 7.787(w), \text{ where } 0.0 \le \text{w} \le 0.008856 \quad (7)$$

B. Fuzzy C-Means Clustering

Past research related to this research mostly about using Fuzzy C-Means and RNN in image processing. There are some previous works using Fuzzy C-Means to do image processing such as [7] which stating the problem with the Fuzzy C-Means algorithm is sensitive to image that have noise so that research proposed detecting the corn plant leaf disease using Fuzzy C-Means combined with Morphology technique and the result is the image is segmented properly with above 99% accuracy to differentiate between sick and healthy corn leaf. There also research [8] which comparing K-Means, Classification, and Fuzzy C-Means algorithm in detecting plant leaf disease detection to find which algorithm is best for detecting plant leaf disease, the proposed method is from Fuzzy C-Means detected area in leaf image, then feature from that image is extracted, and the last step is classification of the leaf image, same step is used for the K-Means algorithm, this research having result from mid-90 to top 90% accuracy for Fuzzy C-Means algorithm. Research [9] using the Fuzzy C-Means algorithm and Statistical Region Merging algorithm to do image segmentation and to eliminate the common negative result in image segmentation such as brightness and image noise, the proposed method is segmenting the image with spatial function and preprocessing with the Statistical Region Merging method and resulting in good accuracy. Research [10] is doing cucumber leaf disease detection with combining the Fuzzy C-Means and Neighborhood Grayscale Information algorithms, the idea of this research is because of the different characteristic of greenhouse plant and plant in outdoor plantation having the algorithm used for detecting greenhouse plant perform badly so the proposed method is first the data preprocessing then morphological filtering method is applied and the next step is watershed segmentation is applied until the solidity value is reached maximum value, the result of this research are the leaf feature is properly extracted despite the background of the leaf object in the image, the noise filtering capability is increased, and the performance of the Fuzzy C-Means algorithm is increased.

Fuzzy C-Means is a clustering algorithm that use fuzzy clustering model so that the data able to become member in all cluster that created with different membership degree from 0 to 1. After converting the RGB to CIE-L*A*B* color space we set the number of clustering for Fuzzy C-Means algorithm and partition the data into fixed number of clusters. According to [7] and [11] the Fuzzy C-Means clustering algorithm is first define the data input such as step (1), then determine the required variable for Fuzzy C-Means Clustering such as step (2), then define initial partition matrix with the eq. (8), after that calculate the center of the cluster with eq. (9), next is calculating the objective function with eq. (10), and calculate the changes in partition matrix with eq. (11), finally check the stop condition given in step (7). The complete algorithm of Fuzzy C-Means algorithm is given below.

1) Data input X which is a matrix with n x m dimension (n is total of sample data and m is attribute of each data), X_{ij} = data-i (i = 1,2, ..., n), attribute-j (j=1,2, ..., m).

2) Determine c as total of clusters, w as weighting component, *IterMax* as maximum iteration, ε as smallest error expected, P_0=0 as the initial objective function, and t=1 as initial iteration.

3) Define μ_{ik} as variable to contain random number with i = 1,2, ..., n and k = 1,2, ..., c as elements of the U initial partition matrix. Next calculate the number of each column (attribute) with equation:
$$Q_j = \sum_{k=1}^{c} \mu_{ik} \text{ with } j = 1,2, \dots, m \quad (8)$$

4) Calculate the center of the cluster-k which is V_{kj} with k = 1,2, ..., c and j = 1,2, ..., m with equation:
$$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} * X_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^{w}} \quad (9)$$

5) In the iteration-t calculate the objective function with equation:
$$P_t = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right](\mu_{ik})^{w}\right) \quad (10)$$

6) Calculate changes to partition matrix with equation:
$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (11)$$

7) Check the stop condition:
   A. If $(|P_t - P_{t-1}| < \varepsilon)$ or (t > *IterMax*), then stop

B. If it doesn't meet the condition of point (A), then $t = t + 1$, then repeat the step 4

C. Artificial Neural Network (ANN)

Before learning about Recurrent Neural Network, learning about ANN is a must. ANN is a computer system that inspired by human nervous system. It is based on complexity of interconnection and architecture of the biological system. An ANN model is data-driven mathematical model that can solve machine-learning problem. Advantages of using ANN is that the user can identify complex relationship of data and outputting good result without inputting direct knowledge[12].

The most common ANN have 3 layers, that is input layer, hidden layer, and output layer. The complexity of ANN architecture is depended on how many is the number of hidden layer and as well as number of neurons in each hidden layer. ANN have some problem such as vanishing gradient problem, so from that problem RNN algorithm is introduced to solve that problem[12]. Fig.1 is illustration of ANN.

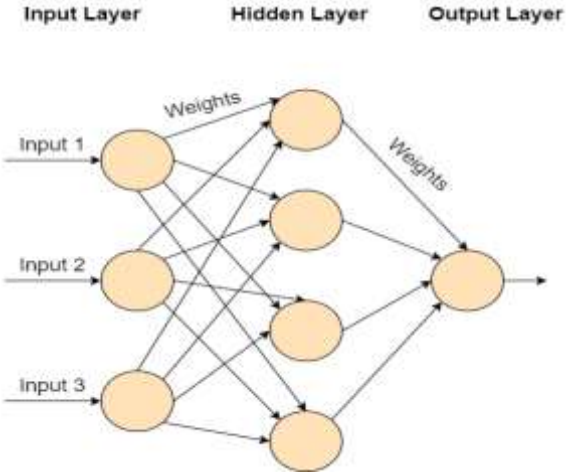

Fig. 1. Illustration of ANN workflow

D. Recurrent Neural Network (RNN)

In this research the detection of corn plant disease is also done with deep learning algorithm that is RNN. RNN is widely used for text processing unlike Convolutional Network (CNN) that used in image processing, but there is already some previous research that has been done implementing RNN for image processing purposes such as [13], the problem studied in this research are that even with many method that can detected the disease of plant but the main problem is the dataset, so this research proposed the method for detecting plant disease using deep learning method in this case using the RNN variant that is LSTM to identify the plant disease and comparing it to Random Forest and K-Nearest Neighbor algorithm, the result is LSTM in second place with 95.79% accuracy, Random Forest at third place with 94.95% accuracy, and K-Nearest Neighbor at first place with 98.19% accuracy. Though the LSTM come in second for accuracy in that research but it is been proved that the algorithm can be used in image processing, in this case to identify plant disease. The next research is [14], the main problem researched in this study is that CNN is widely used for image processing but still have some weakness such as detecting irrelevant background and recognize it as main object, so the proposed method is implementing Attention-Based RNN to classify plant disease with 4 dataset and comparing it to the widely used deep learning-based image classification algorithm that is Convolutional Network (CNN), and resulting in RNN having the best overall accuracy in 3 dataset and losing in accuracy to CNN in 1 dataset but overall, the RNN algorithm which is implemented in that research prove that RNN can be used for classify plant disease. There also [15] which have research problem that the four chambered view left ventricle in cardiac MRI image is hard to be segmented, so the proposed method is implementing RNN-LSTM in medical field which is using Dense RNN-LSTM algorithm to do image segmentation on cardiac MRI image and the result is the Dense RNN-LSTM algorithm have the accuracy of 92.13% and indicated that it can be used for cardiac disease diagnosis and cardiac mechanism analysis

Recurrent Neural Network (RNN) is a class of neural network where the connection between nodes of a layer forms a directed graph along a sequence of variables. RNN have some variation in order to improve the main algorithm, the variation is Gated Recurrent Units (GRU) and LSTM (Long-Short-Term Memory)[14].

Typical RNN is consist of 1 input layer, 1 or more hidden layer, and output layer and have a structure that similar to chain which have recurrent module for memory that save important information from the process before this process. The workflow of RNN is described in Fig.2 where h_t is output from time step t and X_t is input from time step t [12].

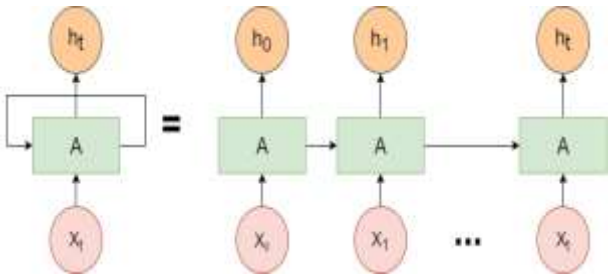

Fig. 2. Illustration of RNN Workflow

RNN-LSTM is an improved version of RNN, according to [15] the RNN-LSTM algorithm can be described such as Fig.3.

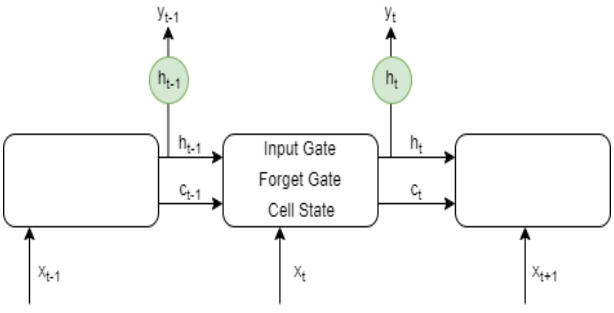

Fig. 3. Illustration of RNN-LSTM Workflow

According to [16] LSTM have memory cell and gate inputs (input gate, forget gate, cell gate, and output gate). In forget gate, each data that flows in will be thrown out or will be saved with the sigmoid activation function (if sigmoid activation function has 1 as value, then data will be saved but if sigmoid activation function have 0 as value, then data will

be thrown out), below is the eq. (12) that is sigmoid activation function used at forget gate:

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right) \quad (12)$$

Next gate is input gate, in this gate there is 2 process that must be done, the first process is select the value which will be renewed with sigmoid activation function and the second step is calculate function tanh that will create new vector value and save the value at memory cell. Below is the eq. (13) and eq. (14) used in input gate:

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (13)$$

$$\hat{c}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (14)$$

Then, there is cell gate that will do replacing process for memory cell value before the process to the new value that calculated from combining the value in forget gate and input gate with the following eq. (15):

$$c_t = f_t * c_{t-1} + i_t * \hat{c}_t \quad (15)$$

Last gate is output gate, this gate will be doing selection of memory cell value that will be selected as output value with the calculation of sigmoid activation function, then that value from sigmoid activation function is multiply by tanh activation function which has selected memory cell value as input, the it will produce the output value. Below is the eq. (16) and eq. (17) used at output gate:

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (16)$$

$$h_t = o_t \tanh(c_t) \quad (17)$$

E. Evaluation

Lastly, evaluation of the algorithm is using accuracy eq. (18) provided by [7]:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (18)$$

Information:

TP = the corn leaf image disease detection from experts and the corn leaf image disease detection from the program are equally appropriate

TN = the corn leaf image disease detection from the corresponding program while the corn leaf image disease detection from experts is not appropriate

FP = the corn leaf image disease detection from the program and the corn leaf image disease detection from experts are equally inappropriate

FN = the corn leaf image disease detection from the program is not suitable while the corn leaf image disease detection from experts is appropriate

III. Result And Discussion

A. Dataset

This research using the dataset from Kaggle (URL: https://www.kaggle.com/datasets/nafishamoin/new-bangladeshi-crop-disease) in form of image with format .jpg. The dataset has 4 folder which represent type of plant that is corn, potato, rice, and wheat. Since in this research the main focus is about detecting corn plant leaf disease so only the corn plant folder is selected. In the corn folder is having 4 subfolder that represent the class of the corn leaf (common rust 1192 images, gray leaf spot 513 images, northern leaf blight 985 images, and healthy 1162 images) with all corn plant image total size is 49.8 MB. Therefore, this research will analysis the image by extracting the color feature from the image then converting the color feature to CIE-L*A*B* color space then doing clustering process with Fuzzy C-Means algorithm and the image is reconstructed in grayscale image, finally the Fuzzy C-Means algorithm is evaluated to see the performance of this algorithm. Example of the data in the dataset given in Fig.4 for healthy corn plant leaf, Fig.5 for corn plant leaf that have common rust disease, Fig.6 for corn plant leaf that have gray leaf spot disease, and Fig.7 for corn that have northern leaf blight disease.

Fig. 4. Example of a healthy corn leaf

Fig. 5. Example of a common rust disease on corn leaf

Fig. 6. Example of a gray leaf spot disease on corn leaf

B. *Research Stages*

The first step of this research is preprocessing the corn leaf dataset such as resizing and converting the color space from RGB to CIE-L*A*B* then the next step is the data that already preprocessed is given to the Fuzzy C-Means and RNN algorithm to detect the disease of the corn leaf or corn leaf is healthy, after that process the evaluation of method is applied to calculate accuracy and algorithm execution time. The flow of this research can be seen at Fig.8

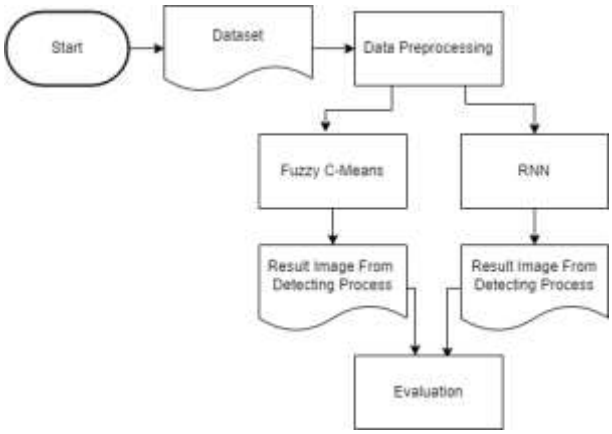

Fig. 8. Flowchart Research Stages

C. *Experiment Result*

In this research will only experimenting the training of the deep learning RNN model and doing segmentation with Fuzzy C-Means The first result is from training the RNN model with 80% training data and 20% validation data with 224x224 resized image. The data trained for 20 epochs with 38 minutes and 1.8 seconds in total execution time and resulting with 0.9403 for accuracy and 0.9572 for validation accuracy, the detail off all iteration accuracy can be seen at Fig. 9, the blue line is the training accuracy and the orange line is validation accuracy.

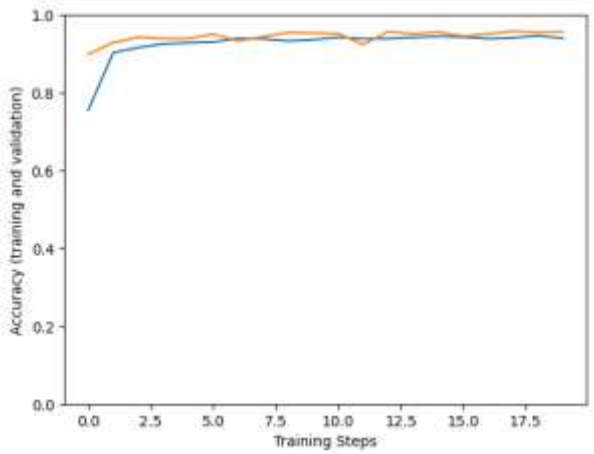

Fig. 9. RNN Training Accuracy

The next step is the segmentation with Fuzzy C-Means clustering algorithm which is resulting in 94 minutes 16 seconds, the sample result of segmentation can be seen in Fig. 10, Fig. 11, Fig. 12, and Fig. 13.

IV. CONCLUSION

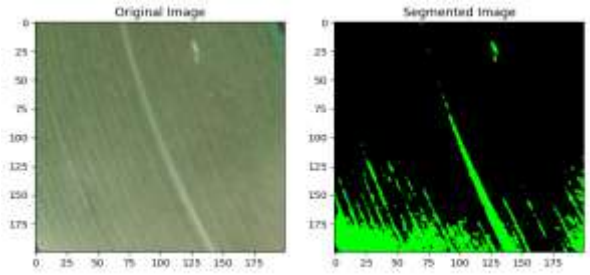

Fig. 10. Healthy Corn Leaf Segmentation Result

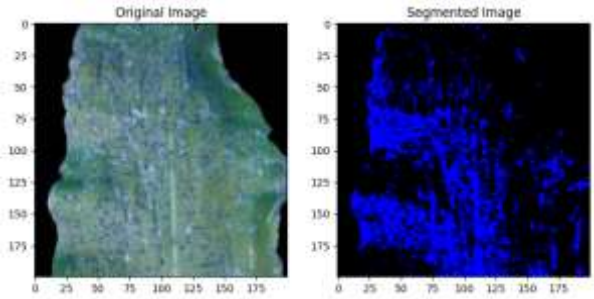

Fig. 11. Common Rust Corn Leaf Segmentation Result

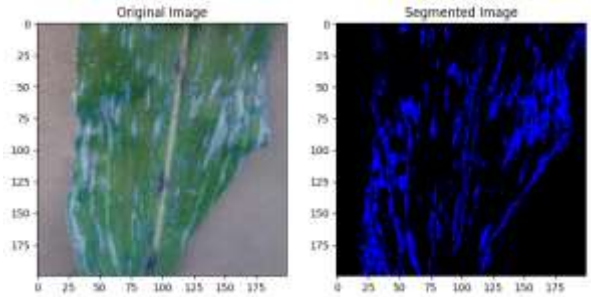

Fig. 12. Gray Leaf Spot Corn Leaf Segmentation Result

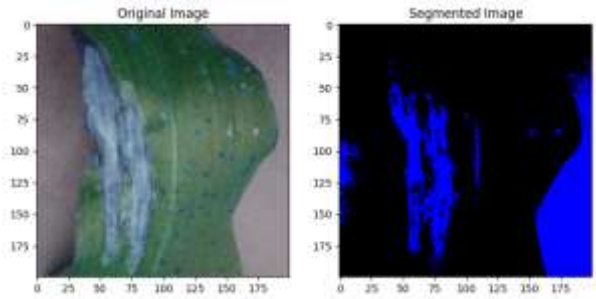

Fig. 13. Northern Leaf Blight Corn Leaf Segmentation Result

IV. CONCLUSION

Therefore, from that all of that previous research we can conclude that Fuzzy C-Means and RNN algorithm can be used for detecting plant leaf disease. In this study the Fuzzy C-Means and RNN algorithm is compared in detecting corn plant leaf disease to find which algorithm is perform better in detecting corn leaf disease. For future research, the dataset can be preprocessed further such as controlling the brightness and

splitting the training and test data with different partition, then for the Fuzzy C-Means method can be combined with classification algorithm for detecting the corn leaf disease also for the RNN model can be tested for prediction to show the model performance in detecting the corn leaf disease and increasing the epochs to see better accuracy.

REFERENCES

[1] KEMENTERIAN PERTANIAN REPUBLIK INDONESIA, "Kementerian Pertanian Libatkan PBNU Tanam Jagung Serentak Se Indonesia." https://www.pertanian.go.id/home/?show=news&act=view&id=3357 (accessed May 16, 2022).

[2] I. G. Swibawa, F. X. Susilo, A. M. Hariri, and Solikhin, "TINGKAT SERANGAN DAN POPULASI WERENG PERUT PUTIH:HAMA BARU PERTANAMAN JAGUNG DI LAMPUNG," 2017.

[3] A. M. Adnan, "Teknologi Penanganan Hama Utama Tanaman Jagung," in *Prosiding Seminar Nasional Serealia 2009*, 2009, pp. 454–469.

[4] I. G. Swibawa, F. X. Susilo, A. M. Hariri, and Solikhin, "The Population of White-Bellied Planthoppers and Their Natural Enemies: The New Pest of Corn in Lampung," *Jurnal Hama dan Penyakit Tumbuhan Tropika*, vol. 18, no. 1, pp. 65–74, Nov. 2018, doi: 10.23960/j.hptt.11865-74.

[5] A. H. Talanca, "Deteksi Beberapa Genotipe Jagung Terhadap Penyakit Bercak Daun," in *Prosiding Seminar Nasional Serealia*, 2015, pp. 415–420.

[6] D. Buvaneswari, K. v Sumathi, and K. K. Kavitha, "Segmentation of Mri Images Using Fuzzy C-Means Clustring," *International Journal of Engineering Science Invention*, vol. 7, no. 1, pp. 30–33, 2018.

[7] Nurhidayati, S. Anam, and A. Widodo, "Automatic Detection of Corn Leaf Desease by Using Fuzzy C-Means Algorithm and Morphological Operation," *American Journal of Engineering Research (AJER*, vol. 8, no. 8, pp. 319–327, 2019, [Online]. Available: www.ajer.org

[8] A. Kumari, S. Abinaya, and S. Meenakshi, "Plant Leaf Disease Detection Using Fuzzy C-Means Clustering Algorithm," *International Journal of Engineering Development and Research*, vol. 6, no. 3, pp. 157–163, 2018, [Online]. Available: www.ijedr.org

[9] I. M. B. Adnyana, I. K. G. D. Putra, and I. P. A. Bayupati, "Segmentasi Citra Berbasis Clustering Menggunakan Algoritma Fuzzy C-Means," *Jurnal Teknologi elektro*, vol. 14, no. 1, pp. 16–20, 2015, [Online]. Available: www.weizmann-usa.org

[10] X. Bai, X. Li, Z. Fu, X. Lv, and L. Zhang, "A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images," *Comput Electron Agric*, vol. 136, pp. 157–165, Apr. 2017, doi: 10.1016/j.compag.2017.03.004.

[11] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, "IMPLEMENTASI FUZZY C-MEANS CLUSTERING DALAM PENENTUAN BEASISWA," *Jurnal Ilmu Matematika dan Terapan*, vol. 11, no. 1, pp. 1–11, 2017.

[12] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of Long Short-Term Memory (LSTM) neural network for flood forecasting," *Water (Switzerland)*, vol. 11, no. 7, 2019, doi: 10.3390/w11071387.

[13] G. S, T. B, and M. Sahu, "Long Short-Term Memory Recurrent Neural Networks for Plant disease Identification," *International Journal of Computer Communication and Informatics*, vol. 3, no. 2, pp. 51–62, Oct. 2021, doi: 10.34256/ijcci2125.

[14] S. H. Lee *et al.*, "Attention-Based Recurrent Neural Network for Plant Disease Classification," vol. 11, p. 601250, 2020, doi: 10.3389/fpls.2020.601250ï.

[15] Y. Wang and W. Zhang, "A Dense RNN for Sequential Four-Chamber View Left Ventricle Wall Segmentation and Cardiac State Estimation," *Front Bioeng Biotechnol*, vol. 9, Aug. 2021, doi: 10.3389/fbioe.2021.696227.

[16] M. Wildan Putra Aldi and A. Aditsania, "Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin," in *e-Proceeding of Engineering*, 2018, pp. 3548–3555.