Research Article

Severity Estimation of Plant Leaf Diseases Using Segmentation Method

Chyntia Jaby Entuni*, Tengku Mohd Afendi Zulcaffle, and Kuryati Kipli
Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, Malaysia

Fatih Kurugollu
College of Engineering and Technology, University of Derby, United Kingdom

* Corresponding author. E-mail: tiajaby@gmail.com DOI: 10.14416/j.asep.2020.11.004
Received: 1 May 2020; Revised: 17 September 2020; Accepted: 21 September 2020; Published online: 9 November 2020

Abstract
Plants have assumed a significant role in the history of humankind, for the most part as a source of nourishment for human and animals. However, plants typically powerless to different sort of diseases such as leaf blight, gray spot and rust. It will cause a great loss to farmers and ranchers. Therefore, an appropriate method to estimate the severity of diseases in plant leaf is needed to overcome the problem. This paper presents the fusions of the Fuzzy C-Means segmentation method with four different colour spaces namely RGB, HSV, L*a*b and YCbCr to estimate plant leaf disease severity. The percentage of performance of proposed algorithms are recorded and compared with the previous method which are K-Means and Otsu's thresholding. The best severity estimation algorithm and colour space used to estimate the diseases severity of plant leaf is the combination of Fuzzy C-Means and YCbCr color space. The average performance of Fuzzy C-Means is 91.08% while the average performance of YCbCr is 83.74%. Combination of Fuzzy C-Means and YCbCr produce 96.81% accuracy. This algorithm is more effective than other algorithms in terms of not only better segmentation performance but also low time complexity that is 34.75s in average with 0.2697s standard deviation.

Keywords: Fuzzy C-Means, K-Means, Otsu's, Plant leaf disease detection, Corn

1 Introduction

Plant disease severity is a significant parameter to quantify disease level and hence can be utilized to foresee yield and prescribe treatment [1]. The precise finding of plant leaf disease severity will diminish yield losses. Customarily, plant disease severity is scored with a visual inspection of plant tissue prepared by specialists [2]. For the most part in remote towns, a rancher looks for agro-exhortation on the diseases by portraying obvious indications to a specialist via telephone [3]. In this way, the turnaround time can be enormous as specialists are not accessible in huge numbers. The other way is to physically convey the specimens to the experts, which is likewise tedious. However, with the increasing number of digital cameras and the advances in computer vision, the automated disease diagnosis models are profoundly applied by precision agribusiness, high-throughput plant phenotype, smart green house, and so forth [4].

For example, Tekam *et al*. [5] accomplished an 80% precision to distinguish cotton leaf diseases. The algorithm used to segment cotton leaves was an artificial neural network (ANN). One of the disadvantages of using ANN according to Mohamed *et al*. [6] is that ANN requires large amounts of data to train and this

Please cite this article in press as: C. J. Entuni, T. M. A. Zulcaffle, F. Kurugollu, and K. Kipli, "Severity estimation of plant leaf diseases using segmentation method," *Applied Science and Engineering Progress*, 2020, DOI: 10.14416/j.asep.2020.11.004

was computationally intensive. Zhu *et al*. [7] utilized iterative thresholding and morphological algorithms to locate the disease spots in a corn leaf. An 80% detection accuracy was achieved with 30 images of corn leaves. However, there are a problem with the morphological method.

The problem with the morphological method is that it does not work well if too many edges are present or not fit for flat valleys [8]. Other than that, Revathi *et al*. [9] extract disease spots in cotton leaf using edge detection segmentation algorithm. However, this algorithm is not working if the edges are not defined perfectly and it is less immune to noise [8]. Hence, more effective algorithms are needed for the estimation of plant leaf disease severity.

The study conducted by Singh *et al*. [10] utilized K-Means segmentation and Back Propagation Neural Network (BPNN) to detect diseases on a leaf of *Phaseolus vulgaris* and *Camellia sinensis*. In their study, the conversion of image colour space was from RGB to HIS. However, HIS has a deficiency of numerically unstable at low saturation because of its nonlinear transformation [11]. A few scientists were additionally working on the correlation of various colour spaces execution in recognizing objects. For example, Jeyalaksshmi *et al*. [12] worked on HSV and RGB image colour spaces. They expressed that RGB is non-helpful for items determination and acknowledgement of colours. Other than that, HSV colours may be characterized effectively by human discernment but the hue point was to only value deviations. Hence, the information required is less available.

Ziya *et al*. [13] determine the Cercospora leaf spot disease level in sugar beets leaf using K-Means segmentation and matching the coordinating of the results with visual assessment done by an expert with disease severity scale. The results obtained by K-Means and expert's observation are close but only 12 images of leaves were used in this study. Other than that, Mokhled [14] utilised Fuzzy C-Means with polygon auto-cropping segmentation for detection and severity rating of olive leaf spot disease. A precision rate of 86% was achieved after being compared with manual scoring and picture analysis. The result obtained was less satisfying because of the small database which contained only 100 images of olive leaves. According to Kong *et al*. [15], the limitation of data size will lead to insignificant results. Thus, the larger the dataset, the better the results obtained. Consequently, database extension is necessary in order to reach more accuracy.

Hence, this paper presents a new proposed algorithm to estimate disease severity in a plant leaf. Fusion of Fuzzy C-Means and YCbCr colour space are applied which has never been used before in the field of plant pathology. This new algorithm performs better than other algorithms.

Fuzzy C-Means achieves high quality clustering with repetitive iterations. The iteration minimizes an objective function that represents the distance from any given data point to a cluster centre weighted by the membership of that data point in the cluster.

It gives the best result for an overlapped dataset of plant leaf images. Hence, Fuzzy C-Means gives more accurate segmentation than K-Means and Otsu's thresholding as presented in the results and discussion section of this paper. Application of YCbCr also makes this algorithm more efficient than the other existing algorithms that commonly utilize only RGB. It is proven by the result obtained that YCbCr has a high detection rate of diseases spot compared to RGB, HSV and L*a*b. In addition, Fuzzy C-Means and YCbCr have low time complexity (40.57s), making it more efficient than other algorithms. In this study, a comparison of proposed and previously used segmentation algorithms that had been utilized to segment diseased plant leaves in terms of their accuracy and time complexity was carried out and presented in result and discussion section.

2 Methods

The tools used in this study is a laptop computer, Acer Aspire E5-471 with MATLAB R2018b. The proposed methodology intends to recognize the most appropriate disease severity estimation algorithm for a plant leaf. For the experimental purpose, corn leaves from PlantVillage Image dataset are utilized. The content of PlantVillage Image dataset is summarized in Table 1.

Disease severity estimation of corn leaves was partitioned into the accompanying advances: 1) Image Colour Spaces 2) Image Segmentations. Figure 1 presents the process block diagram of severity estimation of corn leaf diseases.

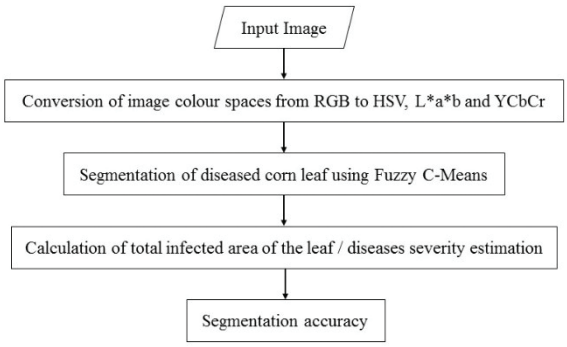


Figure 1: Process block diagram for segmentation of the corn leaf images.

Table 1: Samples of PlantVillage Image dataset

Type of Disease	 Blight	 Gray Spot	 Rust	 Healthy
Total Images	984	508	1192	105
Size Images	227×227	227×227	227×227	227×227

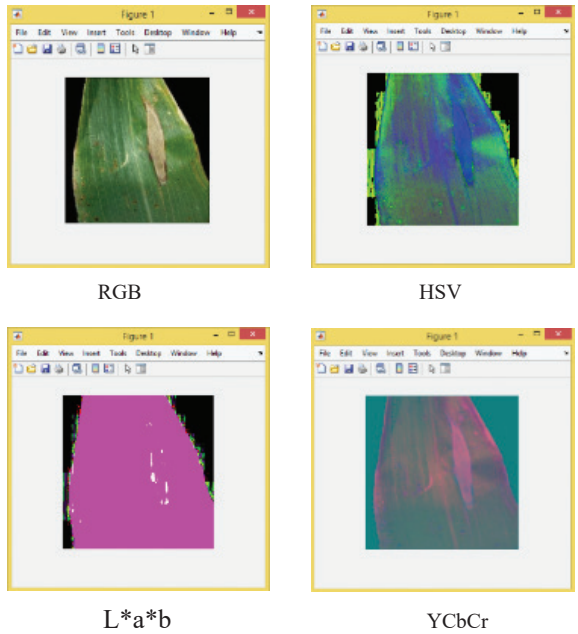


Figure 2: Image of a corn leaf converted to different colour spaces.

2.1 *Image colour spaces*

All the corn leaf images from PlantVillage Image dataset are in JPEG format. Leaf diseases in corn leaf are distinctive in severities. As such, disease spot is different in term of amount of infected area as shown in Table 1. Before segmentation, these images are colour convert from RGB image to HSV, L*a*b and YCbCr. In plants, leaf vein is totally different in intensity and disease spot is different in colour when contrasted with plant leaf itself [16]. Other than that, a few pictures have a shadow in it. In this manner, vein and shadow will appear in the binary image with the uninfected area during the segmentation process. To limit the effect of the presence of vein and shadow, RGB was being reworked before segmentation. Simply from that point onward, the images were converted to HSV, L*a*b and YCbCr as shown in Figure 2.

In this study, different colour spaces were used to improve the effectiveness and adequacy of existing image processing methods. Other than that, different colour spaces have different effects on image segmentation of corn leaf images. Each colour spaces possessed dissimilar colour channels and brightness which affect the image segmentation process. There are four different colour spaces being investigated which are RGB, HSV, L*a*b and YCbCr as shown in Figure 2.

2.1.1 HSV

HSV (hue, saturation, value) are optional representations of the RGB colour model. Hue and saturation channels carry colour information, while value carries information about intensity (essentially brightness) [17]. Equation (1) below shows how to convert RGB to HSV colour space to obtain the value for H, value S and value of V.

$$R' = R/255 \qquad G' = G/255 \qquad B' = B/255$$

$$H = \begin{cases} 0^\circ, & \Delta = 0 \\ 60^\circ \times \left(\dfrac{G'-B'}{\Delta} mod 6\right), & C_{max} = R' \\ 60^\circ \times \left(\dfrac{B'-R'}{\Delta} + 2\right), & C_{max} = G' \\ 60^\circ \times \left(\dfrac{R'-G'}{\Delta} + 4\right), & C_{max} = B' \end{cases}$$

$$C_{max} = \max(R', G', B') \qquad S = \begin{cases} 0, & C_{max} = 0 \\ \dfrac{\Delta}{C_{max}}, & C_{max} \neq 0 \end{cases}$$

$$C_{min} = \min(R', G', B') \qquad V = C_{max}$$

$$\Delta = C_{max} - C_{min} \tag{1}$$

where R', G' and B' is gamma-compressed of red, green and blue components, C_{max} is the maximum value in the RGB scale and C_{min} is the minimum value in the RGB scale. While, H is hue, S is saturation and V represents value. Hue and saturation channels carry information about colour while for value, it carries information about intensity (essentially brightness).

*2.1.2 L*a*b*

L*a*b* expresses colour as three values: L^* for the lightness from black (0) to white (100), a^* from green (−) to red (+), and b^* from blue (−) to yellow (+) [18]. Equation (2) shows how to convert RGB to L*a*b* and calculate the value of L^*, a^* and b^*.

$$R' = R/255$$
$$G' = G/255$$
$$B' = B/255$$
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} sR' \\ sG' \\ sB' \end{bmatrix}$$
$$L^* = 116(Y/Y_n)\tfrac{1}{2} - 16,$$
$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right)$$
$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \quad (2)$$

where $X_n = 95.0489$, $Y_n = 100$, $Z_n = 108.8840$ and $Y = 100$. XYZ represents tristimulus values, R', G' and B' represents gamma-compressed of red, green and blue components and sR', sG' and sB' represents lowest common denominator in RGB colour spaces. While L^* is Luminance, a^* is Red-green axis layer and b^* is blue-yellow axis layer.

2.1.3 YCbCr

In YCbCr, the Y' is the brightness (luma), Cb is blue minus luma (B-Y) and Cr is red minus luma (R-Y). YCbCr colour space can successfully isolates luminance from chrominance compared to other colour spaces like RGB [19]. YCbCr demonstrates the best execution contrasted with RGB because it can isolate luminance from chrominance more effectively than RGB colour space [20].

Equation (3) below shows how to identify the value of Y', C_b and C_r after converting the image from RGB to YCbCr.

$$R' = R/255$$
$$G' = G/255$$
$$B' = B/255$$
$$Y' = 16 + (65.481 \cdot R' + 128.553 \cdot G' + 24.966 \cdot B')$$
$$C_b = 128 + (-37.397 \cdot R' - 74.203 \cdot G' + 112.0 \cdot B')$$
$$C_r = 128 + (112.0 \cdot R' - 93.786 \cdot G' - 18.214 \cdot B') \quad (3)$$

where R', G' and B' represents gamma-compressed of red, green and blue components while Y' represents luma or brightness. C_b is blue minus luma and C_r is red minus luma.

2.2 *Image segmentation*

Segmentation means partitioning of an image into various part of same features with similarity. The features can be colour, texture morphology and pixel value. Segmentation can be done using various methods. In this study, Fuzzy C-Means, K-Means and Otsu's thresholding-based segmentations are utilized to perform the segmentation of the uninfected area of corn leaf from the infected area.

2.2.1 Fuzzy C-Means

Fuzzy C-Means is a method of clustering which enables one piece of data to belong to two or more clusters. It selects the number of cluster centres indiscriminately and all data points have been automatically given cluster centre fuzzy membership. Cluster centre constantly experiences certain revise iteration method. An iterative procedure was done to limit all the data points to each cluster centre distance and the weighted degree of membership for optimal target or ideal objective [21]. Other than that, an iterative process is employed to manage the outliers and improve the overall estimation performance [22].

In this study, Fuzzy C-Means partitions a corn leaf image into two clusters. One is a cluster of

segmented infected area and the other one is cluster of segmented uninfected area. Clustering in Fuzzy C-Means is firstly done with choosing primary essential centroids c_i and processing the degree of membership of all feature vectors in every one of the clusters following the Equation (4).

$$u_{ij} = \frac{\left[\frac{1}{d^2(x_j, c_k)}\right]^{\frac{1}{(q-1)}}}{\sum_{k=1}^{K}\left[\frac{1}{d^2(x_j, c_k)}\right]^{\frac{1}{(q-1)}}} \tag{4}$$

where u_{ij} is membership of ith data to jth cluster centre, K is number of data points, c_k is centroid of data k, x_j is data point of j, k is point k, q is parameter that determines the influence of the weights and d is distance of centroid to the data point. From that point forward, figuring the new centroid before updating the memberships as per the condition of u_{ij} above. The new centroid is computed using Equation (5).

$$c^{\wedge}{}_i = \frac{\sum_{j=1}^{M}(u_{ij})^q x_j}{\sum_{j=1}^{M}(u_{ij})^q} \tag{5}$$

where c_i is the centroid of cluster i, M is fuzziness index, u_{ij} is membership of ith data to jth cluster centre, q is the parameter that determines the influence of the weights and x_j is point j.

Favourable position of Fuzzy C-Means is that it gives the adaptability to express that data points can belong to more than one cluster [23]. It is done with the minimization function presented as in the Equation (6).

$$J_M = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^M \left\| x_j - c_j \right\|^2 \tag{6}$$

where J_M is minimization function, N is number of data points and c is the cluster centre. While, M is fuzziness index, u_{ij} is membership of ith data to j^{th} cluster centre, x_i is point i and j^{th} cluster centre is presented as c_j.

While segmenting the infected area of a corn leaf, data are bound to each cluster by means of a Membership Function. Membership Function used to represent the fuzzy behaviour of this algorithm. Then, a proper matrix named U whose factors are somewhere in the range of 0 and 1 is worked to represent the degree of membership between data and centres of clusters.

2.2.2 K-Means

K-Means algorithm is a partitioning method that segments data into k totally unrelated clusters and returns the index of the cluster to which it has appointed every perception [24]. K-Means clustering works on genuine perceptions and makes a solitary level of clusters. The distinctions mean that K-Means clustering is often more suitable than hierarchical clustering for large amount of data [25]. K-Means treats each observation in your data as an object having a location in space [26]. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. In this study, K-Means partitions corn leaf images into two different unique clusters of which the segmentation is carried out depending on pixel value with various brightness of each cluster. The k centres change their location gradually until no more changes are done. At last, the algorithm targets at limiting an objective function known as the squared error function given by Equation (7).

$$J(V) = \sum_{i=1}^{c}\sum_{j=1}^{c_i}\left(\left\| x_i - v_j \right\|\right)^2 \tag{7}$$

where $J(V)$ is squared error function, c is number of cluster centre, c_i is the number of data points in the ith cluster, $\|x_i - v_j\|$ is Euclidean distance between x_i and v_j, x_i is point i and v_j is j^{th} cluster centre. K-Means algorithm at first arbitrarily select 'c' cluster centres, calculate the distance between each data point and cluster centres, assign the data point to the cluster centre whose distances from the cluster centre is minimum of all the cluster centres and recalculate new cluster centre utilizing Equation (8).

$$v_i = \left(\frac{1}{c_i}\right)\sum_{j=1}^{c_i} x_i \tag{8}$$

where v_i is ith cluster centre, c_i is the number of data points in the ith cluster and x_i is point i. From that point onward, the distance between each data, point and newly acquired cluster centres were recalculated

and lastly, the process ends to obtain the clusters. One of the focal points of using K-Means segmentation is if there are a tremendous measure of factors, K-means clustering is quicker than other clustering methods [27].

2.2.3 Otsu's Thesholding

Otsu is an automatic threshold selection region-based segmentation method [28]. In Otsu's segmentation, a corn leaf image is isolated into two classes that are foreground and background depending on pixel values. Toward the part of the segmentation, the background image that is the uninfected area is segmented from the foreground image, which is the infected area.

Otsu's segmentation includes iterating through all the conceivable threshold values and calculating a measure of spread for the pixel levels on each side of the threshold, for instance, the pixels that either fall in foreground or background [29].

Then, the threshold was employed at level k. At the point when the category means (u_0, u_1) and sophistication variances (σ_B, σ_W) area unit is being determined; a threshold k is searched. This is to maximize one among the article functions (l,k,n) as in Equation (9).

$$l = \frac{\sigma_B^2}{\sigma_W^2}; k = \frac{\sigma_\tau^2}{\sigma_W^2}; n \frac{\sigma_B^2}{\sigma_\tau^2} \tag{9}$$

where l, k and n are article function and σ_B and σ_W represents sophistication variances.

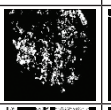
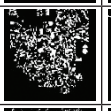
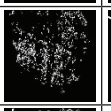
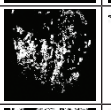
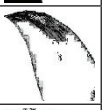
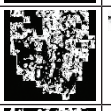
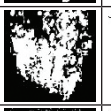
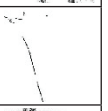
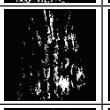
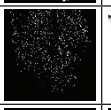
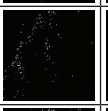
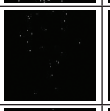
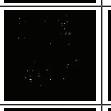
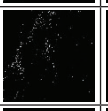
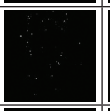
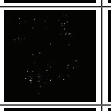
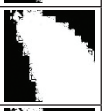
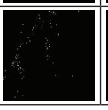
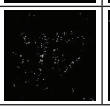
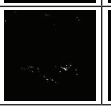
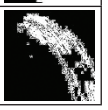
In this study, corn leaves are segmented into one binary thresholded images. The thresholded image is acquired after the algorithm automatically segment using the threshold value, k and it depends on the observed distribution of pixel values.

3 Results and Discussion

3.1 *Total infected area*

Table 2 shows the binary images of RGB, HSV, L*a*b and YCbCr segmented leaves using the proposed method of Fuzzy C-Means and the previous method of K-Means and Otsu's algorithms. The infected areas of the leaf are presented as white and the uninfected areas are presented as black. For healthy leaf images, the healthy part is presented as white and background of the images presented as black. The total infected area is equal to the total pixel value of the white area in the binary image. It is shown in Table 2.

Table 2: Result of segmented diseased corn leaf images using the fusion of different segmentation algorithms with different colour spaces presented in binary image presentation

Algorithm	Disease / Colour Spaces	Blight	Gray Spot	Rust	Healthy
Fuzzy C-Means (Proposed Method)	RGB				
	HSV				
	L*a*b				
	YCbCr				
K-Means (Previous Method)	RGB [30]				
	HSV [31]				
	L*a*b [32]				
	YCbCr [33]				
Otsu's (Previous Method)	RGB [34]				
	HSV [35]				
	L*a*b [36]				
	YCbCr [37]				

Initially, for healthy leaf images, the healthy part is supposed to be presented as white and background of the images presented as black. However, as shown in Table 2, only Fuzzy C-Means algorithm was able to segment healthy corn leaf perfectly with the fusion of different colour spaces. Therefore, Fuzzy C-Means can be considered as an ideal segmentation algorithm and can be a point of reference to identify the performance of other segmentation algorithms with different colour spaces to detect plant leaf diseases severity.

3.2 *Performance of methods*

Before the performance of each method identified, the percentage error of each method is calculated. Percentage of error is identified by comparing the image similarity between manually segmented image and image segmented using Fuzzy C-Means, K-Means and Otsu's thresholding method. A manually segmented image in this study is considered as the ideal and model of excellence of segmented image.

In this study, manual segmentation is the first thing done by us based on Colour Threshold module in MATLAB R2018b application and used to remove parts of the image that fall within a specified colour range. This module can be used to detect objects of consistent colour values. The interface displays the Red, Green and Blue histograms. To perform the segmentation is by grabbing the histogram sliders and moving them across the spectrum of values to remove the unwanted part that is green.

Segmented images of Fuzzy C-Means, K-Means and Otsu's algorithms are used to calculate the total infected leaf area (A_I) and manually segmented images are used to determine the total leaf area (A_T). Total infected leaf area (A_I) and total leaf area (A_T) are used to discover the percentage of infection by using the Equation (10).

$$\textit{Percentage of infection}\ (\%) = \frac{\textit{total infected area}\ (A_I)}{\textit{total leaf area}\ (A_T)} \times 100 \quad (10)$$

Figure 3 shows an example of total leaf area (A_T) and Figure 4 shows the total infected leaf area (A_I) on plant leaf represented as white pixels. Using the manual segmented images and Equation (10), the percentage of infection of the ground truth-values are computed.

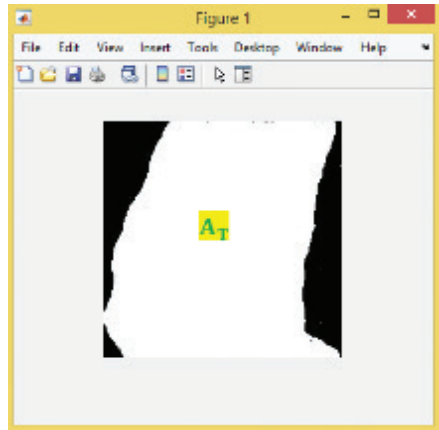


Figure 3: Total leaf area (A_T).

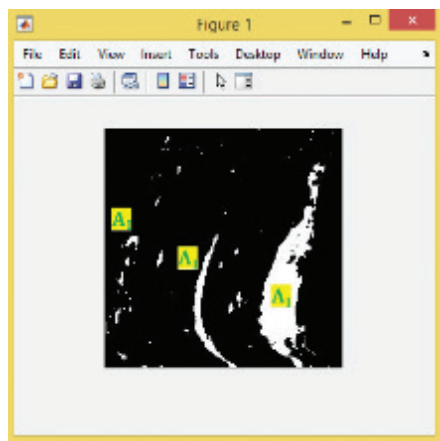


Figure 4: Total infected area (A_I).

After the percentage of infection is identified, the percentage error is determined. The method that gives the least percentage of error is considered as the most relevant method and performs better than other methods with a high percentage of performance. The performance of each method is presented in Figure 5. Percentage error is calculated as in Equation (11).

$$\textit{Percentage error}\ (\%) = [\textit{total leaf area}\ (A_T) - \textit{total infected leaf area}\ (A_I) \times 100 \quad (11)$$

Figure 5 shows the performance of Fuzzy C-Means, K-Means and Otsu's method on all three types of diseases, blight, gray spot and rust. It also shows the average performance of each method on each disease.

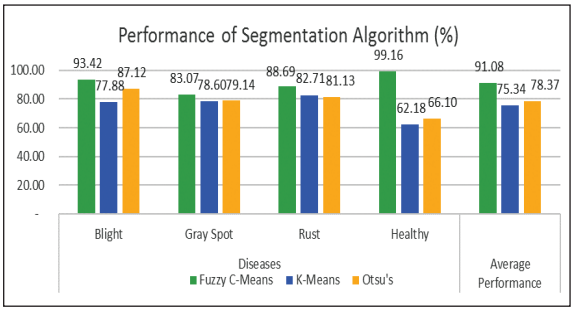


Figure 5: Performance of different segmentation algorithms.

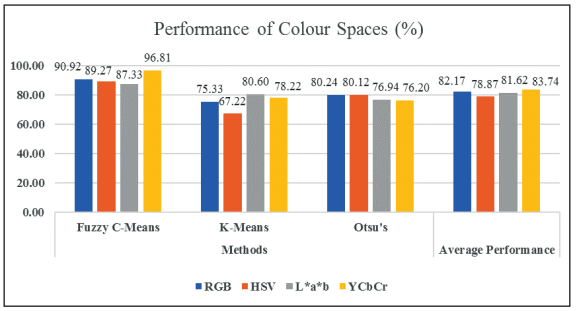


Figure 6: Performance of the segmentation algorithms on different colour spaces.

In overall, Fuzzy C-Means shows the best performance compared to other methods with an average performance of 91.08%. While in terms of time taken, Fuzzy C-Means takes considerably longer computational time (45.92s) than K-Means (32.49s) and Otsu's (32.24s). In order to find the time complexity of segmentation algorithms, the algorithms are run 10 times and their maximum, minimum, standard deviations and averages are recorded. Time taken is computed based on how long it takes for each algorithm to segment the infected area from the uninfected area of diseased leaf. The specification of the computer used is the same as in Section 3.4. Despite having longer computational time, Fuzzy C-Means is still considered to be better on criteria as lower time complexity than other existing methods to estimate the severity of plant leaf diseases.

Fuzzy C-Means takes more computational time than the other diseases severity estimation algorithms. This is because the Fuzzy C-Means performs algorithm iteratively. However, it can be improved by a decision tree approach with it by mining the data in accurate and sequential manner and secondly by creating the noise-free log file. This is because in some of the study done in the related field such as [30], [31], decision tree method is able to improve the performance of algorithms.

3.3 *Accuracy of algorithms*

Different algorithms produced different performance in different colour spaces. Figure 6 shows the graph of the performance of the segmentation algorithms on different colour spaces.

The graph is produced from Table 3 which is the accuracy obtained by fusing the algorithms. From Figure 6, YCbCr gives the best segmentation performance on segmentation algorithms with an average performance of 83.74%. YCbCr also the fastest to segment diseased leaf with the computational time of 99.26 s compared to L*a*b (100.28 s), the HSV (101.33 s) and the RGB (141.71 s). The uses of YCbCr to separate luminance and chrominance is more effective than RGB, HSV, L*a*b.

RGB, HSV, L*a*b are confronting issues with things that have the same infected area's pixel such as shining and reflected things in surrounding. Therefore, YCbCr gives more competitive result of segmentation than the other colour spaces.

Other than that, there are several reasons why RGB, HSV and L*a*b gave low performances on the segmentation algorithms compared to YCbCr. RGB space is not suitable for segmentation or colour processing, because of the high correlation between the components R, G and B. Besides that, in a binary image, hue parameter in HSV colour space is undefined, due to these colours are considered as singularities within this colour space due to no specific chromaticity. While L*a*b having problems with nonlinear transformations and singularity problem, Hence, they are less suitable to be used to estimate disease severity in plant leaf compared to YCbCr.

Table 3 shows the overall result of using the proposed method of Fuzzy C-Means and the previous method of K-Means and Otsu's algorithms to segment diseased plant leaves. The accuracy of algorithms is achieved by comparing the image segmented manually with image segmented using a fusion of algorithms. This is done in MATLAB programming by identifying the differences of pixels between them for comparison.

The manually segmented image which is done using Colour Threshold module in MATLAB R2018b application.

It is considered as the ideal and model of excellence of segmented image and the pixel that does not match with it is consider as an error which can be calculated using Equation (11). Then, the accuracy of algorithms is calculated using Equation (12). The average accuracy of the algorithm is a result of the accuracy of single fusion of algorithms.

$$\textit{Accuracy of algorithm}\ (\%) = 100\% - \textit{Percentage error}\ (\%) \quad (12)$$

Table 3: Accuracy of fusion of algorithms

Algorithms	Colour Spaces	Blight (%)	Gray Spot (%)	Rust (%)	Healthy (%)	Average Accuracy (%)
Fuzzy C-Means (Proposed Method)	RGB	94.18	84.46	88.01	97.02	90.92
	HSV	91.72	79.62	85.92	99.83	89.27
	L*a*b	89.88	76.57	82.98	99.89	87.33
	YCbCr	97.89	91.63	97.83	99.89	**96.81**
K-Means (Previous Methods)	RGB [30]	70.72	83.44	84.70	62.47	75.33
	HSV [31]	68.51	67.56	69.75	63.07	67.22
	L*a*b [32]	84.42	88.79	87.64	61.54	80.60
	YCbCr [33]	87.87	74.62	88.74	61.65	78.22
Otsu's (Previous Methods)	RGB [34]	76.53	88.19	90.09	66.17	80.25
	HSV [35]	93.53	79.43	80.07	67.43	80.12
	L*a*b [36]	88.50	73.74	77.01	68.50	76.94
	YCbCr [37]	89.92	75.22	77.35	62.31	76.20

Therefore, the accuracy of the proposed algorithm of fusion of Fuzzy C-Means and YCbCr colour space produced a better result than other previously used algorithms to segment diseased plant leaf.

Besides that, recent CNN-based semantic segmentation is also commonly used in image segmentation. There are two types of CNN-based semantic segmentation which are the rectangular region of interest and pixel-based segmentation. Hence, in this study, CNN-based semantic segmentation is investigated, and the result is presented in Figures 7 and 8.

In semantic segmentation, the object is segmented based on the region of interest and pixel-based segmentation. However, from the result obtained as in Figures 7 and 8, the CNN-based semantic segmentation is not suitable to be used in this study. This is due to many tiny infected parts that make the labelling task tedious and impractical.

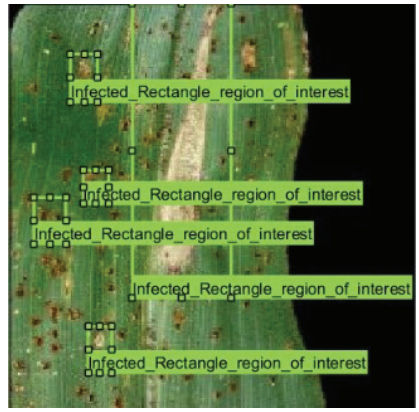


Figure 7: Image segmented by rectangular region of interest.CNN-based semantic segmentation.

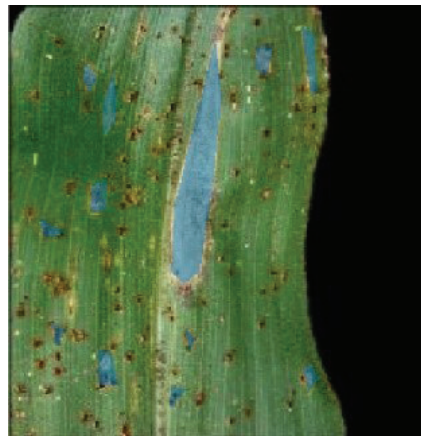


Figure 8: Image segmented by pixel-based CNN-based semantic segmentation.

Therefore, the fusion of the proposed algorithm of Fuzzy C-Means and YCbCr is proved to be better than the recent image segmentation algorithm that is CNN-based semantic segmentation.

Fusion of Fuzzy C-Means and YCbCr algorithm shows the best result with 96.81% average accuracy followed by the fusion of Fuzzy C-Means with RGB algorithm that is 90.92% and fusion of Fuzzy C-Means with HSV algorithm that is 89.27% respectively. While the fusion of K-Means and HSV algorithm shows the lowest average accuracy, which is 67.22%.

Fusion of Fuzzy C-Means and YCbCr algorithm also have a higher accuracy compared to the existing algorithms that have been used in this study. The existing method are a fusion of K-Means segmentation

and RGB which produced an average accuracy of 75.33%, the fusion of K-Means segmentation and HSV with 67.22% average accuracy and fusion of Otsu's segmentation and RGB with 80.25% average accuracy.

These three fusions of segmentation algorithms with different colour spaces were used in the segmentation stage in the study done by Aravind *et al*. [32], Mohanapriya *et al*. [33] and Sibiya *et al*. [34] to extract the symptomatic patterns on the leaf of maize from the same dataset which is PlantVillage image dataset. The outcome of their work has been reviewed, analysed and finally being compared with the proposed method in this study.

These algorithms contribute to good classification accuracy in their study. Therefore, as the result of the fusion of proposed algorithm of fusion of Fuzzy C-Means and YCbCr are better, it could contribute to a much better classification accuracy in the extended classification stage of plant leaf diseases detection.

3.4 *Time complexity of algorithms*

Meanwhile, time taken of the proposed method of using Fuzzy C-Means and previous methods of using K-Means and Otsu's algorithms to segment diseased plant leaves are recorded in Table 4.

Table 4: Time complexity of algorithms

Algorithms	Colour Spaces	Maxi-mum (s)	Mini-mum (s)	Standard Deviation (s)	Average (s)
Fuzzy C-Means (Proposed Method)	RGB	78.91	76.01	1.1556	77.15
	HSV	36.92	35.72	0.3110	36.14
	L*a*b	35.98	35.36	0.2122	35.64
	YCbCr	35.29	34.38	0.2697	**34.75**
K-Means (Previous Methods)	RGB [30]	33.76	31.95	0.5435	32.27
	HSV [31]	33.39	32.37	0.3224	32.62
	L*a*b [32]	33.34	32.26	0.3232	32.47
	YCbCr [33]	33.25	32.42	0.2534	32.59
Otsu's (Previous Methods)	RGB [34]	33.39	32.03	0.3991	32.30
	HSV [35]	33.25	32.32	0.3100	32.58
	L*a*b [36]	32.78	31.75	0.3514	32.18
	YCbCr [37]	33.69	31.17	0.7796	**31.92**

The analysis is carried out using MATLAB R2018b software with the following computer specifications:

- Computer System: Laptop Computer, Acer Aspire E5-471
- Microprocessor: Intel® Core™ i3-4030U CPU
- Microprocessor Clock Speed: 1.90GHz
- Random Access Memory (RAM): 6.00GB
- Operating System: Windows 8.1 Single Language with Bing

The computational cost analysis of the algorithms is based on the evaluation made by Tengku *et al*. [35]. The algorithms are run 10 times and their maximum, minimum, standard deviations and averages are recorded. Time taken is computed based on how long it takes for each algorithm with different colour spaces to segment the infected area from the uninfected area of the diseased leaf. Segmentation is done based on colour extraction. It is shown that besides having the best average accuracy, the fusion of Fuzzy C-Means and YCbCr algorithm shows low average time complexity that is 34.75s as indicated in blue colour in Table 4. The fusion of Fuzzy C-Means and RGB algorithm somehow shows the highest time complexity that is 77.15s with the maximum time taken of 78.91s. Algorithm with lowest time complexity which is 31.92s is a fusion of Otsu's and YCbCr. Compare to the previous method, the time taken for the fusion of Fuzzy C-Means and YCbCr to segment the diseased leaves was longer. However, this is basically because, in this study, the size of the dataset is bigger than the dataset used in the previous method. It can be proved from the study done by Aravind *et al*. [32] where only 61 images were used and Mohanapriya *et al*. [33], 500 images were used, and the result produced shows that segmenting 500 images take a shorter time than segmenting 61 images.

4 Conclusions

From the result, it can be concluded that the best overall severity estimation algorithm and colour space used to estimate the severity of the disease of plant leaf is the fusion of Fuzzy C-Means and YCbCr colour space. Fusion of Fuzzy C-Means and YCbCr algorithm produce 96.81% of accuracy. YCbCr colour space has a greater detection rate as to compare to RGB, HSV, L*a*b because YCbCr can separate luminance from chrominance more effectively. Fusion of Fuzzy C-Means and YCbCr algorithm has an ideal overall computational time complexity, which is 34.75 s average and standard deviation of 0.2697 s to segment 2789 images of diseased and healthy corn leaf images. Hence, the fusion of Fuzzy C-Means and

YCbCr algorithm could establish a highly reliable system to estimate diseases severity of plant leaf. This algorithm is also better than the existing diseases severity estimation algorithm with the application of YCbCr that give more competitive result than the other colour spaces.

References

[1] G. Wang, Y. Sun, and J. Wang, "Automatic image-based plant disease severity estimation using deep learning," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1–8, 2017.

[2] C. H. Bock, G. H. Poole, P. E. Parker, and T. R. Gottwald, "Plant disease severity estimated visually, by igital photography and image analysis, and by hyperspectral imaging," *Critical Reviews in Plant Sciences*, vol. 29, no. 2, pp. 59–107, 2010.

[3] A. Parikh, M. S. Raval, C. Parmar, and S. Chaudhary, "Disease detection and severity estimation in cotton plant from unconstrained images," in *IEEE International Conference on Data Science and Advanced Analytics*, pp. 594–601, 2016.

[4] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, pp. 1–10, 2016.

[5] E. R. Tekam and P. J. Pimple, "A survey disease detection mechanism for cotton leaf: Training & precaution," *The International Journal of Innovative Research in Science, Engineering and Technology*, vol. 6, pp. 205–210, 2017.

[6] A. S. M. El-bhrawy and N. E. K. Mohamed, "Artificial neural networks in data mining," *International Journal of Scientific & Engineering Research*, vol. 7, no. 11, pp. 158–161, 2016.

[7] J. Zhu, A. Wu, and P. Li, "Corn leaf diseases diagnostic techniques," *Communications in Computer and Information Science*, vol. 288, pp. 334–341, 2012.

[8] N. Tokas, S. Karkra, and M. K. Pandey, "Comparison of digital image segmentation techniques-A research review," *International Journal of Computer Trends and Technology*, vol. 5, no. 5, pp. 215–220, 2016.

[9] P. Revathi and M. Hemalatha, "Advance computing enrichment evaluation of cotton leaf spot disease detection using image edge detection," in *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*, 2012, vol. 1, pp. 1–5.

[10] M. K. Singh and S. Chetia, "Detection and classification of plant leaf diseases in image processing using detection and classification of plant leaf diseases in image processing using MATLAB," *International Journal of Life Sciences Research*, vol 5, pp. 120–124, 2018.

[11] A. M. Vyas, B. Talati, and S. Naik, "Colour feature extraction techniques of fruits: A survey," *International Journal of Computer Applications*, vol. 83, no. 15, pp. 15–22, 2013.

[12] N. A. Ibraheem, M. M. Hasan, R. Z. Khan, and P. K. Mishra "Understanding color models: A Review," *ARPN Journal of Science and Technology*, vol. 2, no. 3, pp. 265–275, 2012.

[13] Z. Altas, M. M. Ozguven, and Y. Yanar, "Determination of sugar beet leaf spot disease level (Cercospora Beticola Sacc.) with image processing technique by using drone," *Current Investigations in Agriculture and Current Research*, vol. 5, no. 3, pp. 669–678, 2018.

[14] M. S. Al-Tarawneh, "An empirical investigation of olive leave spot disease using auto-cropping segmentation and fuzzy C-Means classification," *World Applied Sciences Journal*, vol. 23, no. 9, pp. 1207–1211, 2013.

[15] Z. Kong, T. Li, J. Luo, and S. Xu, "Automatic tissue image segmentation based on image processing and deep learning," *Journal of Healthcare Engineering*, vol. 10481, pp. 1–10, 2019.

[16] P. Chaudhary, A. K. Chaudhari, and S. Godara, "Color transform based approach for disease spot detection on plant leaf," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 6, pp. 4–9, 2012.

[17] Q. Zhang and S. Kamata, "A novel color space based on RGB color barycenter," in *The 41st IEEE International Conference on Acoustics, Speech Signal Processing*, 2016, pp. 1601–1605.

[18] P. B. Pathare and U. L. Opara, "Colour measurement and analysis in fresh and processed foods: A review," *Food and Bioprocess Technology*, vol. 6, no. 1, pp. 36–60, 2015.

[19] M. Ugale, A. Nunes, L. Dias, and S. Pereira,

"Smoke and fire detection in videos," *International Journal of Computational Science and Engineering*, vol. 3, no. 3, pp. 132–144, 2015.

[20] N. Izzati, N. Anis, M. Razali, and H. Achmad, "Fire recognition using RGB and YCbCr color space," *ARPN Journal of Science and Technology*, vol. 10, no. 21, pp. 9786–9790, 2015.

[21] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.

[22] D. Majumdar, A. Ghosh, D. K. Cole, A. Chakraborty, and D. D. Majumder, "Application of fuzzy c-means clustering method to classify wheat leaf images based on the presence of rust disease application of fuzzy C-Means clustering method to classify wheat leaf images based on the presence of rust disease," in *the 3rd International Conference on Frontiers in Intelligent Computing: Theory and Applications*, 2014, vol. 327, pp. 277–284.

[23] A. Kumari, S. Meenakshi, and S. Abinaya, "Plant leaf disease detection using fuzzy C-Means clustering algorithm," *International Journal of Engineering Research and Development*, vol. 6, no. 3, pp. 157–163, 2018.

[24] S. K. Sajan and M. G. Alex, "Comparing the performance of texture and cluster based segmentation algorithms in svm classifier for mammogram analysis," *International Journal of Emerging Technologies in Engineering Research*, vol. 5, no. 4, pp. 239–244, 2017.

[25] S. P. Patel and A. K. Dewangan, "Automatic detection of plant leaf disease using K-Means clustering and segmentation," *International Journal of Engineering Sciences & Research Technology*, vol. 6, no. 7, pp. 774–779, 2017.

[26] N. Bharathi and K. C. Santosh, "Implementation of K-Means clustering approach for the identification and edge detection of cotton leaves image processing technique," *International Journal of Engineering and Manufacturing Science*, vol. 8, no. 1, pp. 135–144, 2018.

[27] S. Kaur, S. Pandey, and S. Goel, "Semi-automatic leaf disease detection and classification system for soybean culture," *IET Image Processing*, vol. 12, no. 6, pp. 1038–1048, 2018.

[28] H. J. Vala and P. A. Baxi, "A review on otsu image segmentation algorithm," *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 2, no. 2, pp. 387–389, 2013.

[29] D. Alsaeed, A. Bouridane, and A. El-zaart, "A novel fast otsu digital image segmentation method," *International Arab Journal of Information Technology*, vol. 13, no. 4, pp. 427–434, 2016.

[30] Y. Y. Song and Y. Lu, "Decision tree methods: Applications for classification and prediction," *Shanghai Arch Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.

[31] H. Qabbaah, G. Sammour, and K. Vanhoof, "Decision tree analysis to improve e-mail marketing campaigns," *International Journal Information Theories and Applications*, vol. 26, no. 1, pp. 3–36, 2019.

[32] K. R. Aravind, P. Raja, K. V Mukesh, R. Aniirudh, R. Ashiwin, and C. Szczepanski, "Disease classification in maize crop using bag of features and multiclass support vector," in *2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, vol. 5, no. 6, pp. 1191–1196.

[33] C. Mohanapriya and P. R. Tamilselvi, "Maize leaf disease severity analysis using integrated color filtering and threshold masking," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 5863–5871, 2019.

[34] M. Sibiya and M. Sumbwanyambe, "An algorithm for severity estimation of plant leaf diseases by the use of colour threshold image segmentation and fuzzy logic inference: A proposed algorithm to update a 'Leaf Doctor' application," *AgriEngineering*, vol. 1, pp. 205–219, 2019.

[35] T. M. A. Zulcaffle, F. Kurugollu, D. Crookes, A. Bouridane, and M. Farid "Frontal view gait recognition with fusion of depth features from a time of flight camera," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1067–1082, 2019.