

A Model for Detecting Fungal Diseases in Cotton Cultivation using Segmentation and Machine Learning Approaches

Odukoya O.H.¹, Aina S.², Dégbéssé F.W.³

Computer Science and Engineering Department, Obafemi Awolowo University, Ile-Ife, Nigeria^{1,2,3}

Abstract— This research detailed a model for detecting fungal diseases via techniques for processing images of cotton leaves. The work allowed to develop a model based on the set of preprocessed data, to formulate the developed model, to simulate and evaluate the model. It is about detecting fungal diseases in cotton cultivation. The image data records were collected in an online data repository consisting of images of cotton leaves infected with fungal diseases and normal leaf images. In addition, other images of infected and uninfected cotton leaves were collected in cotton production fields in the Ségbana region in Benin Republic. The model was formulated based on watershed segmentation technique by applying Edge Detection algorithm and K-Means Clustering; and Support Vector Machine (SVM) for classification. The simulation was done using MATLAB with Image Processing Toolbox 9.4. The results gave an accuracy of 99.05%, specificity 90%, misclassification rate 0.95%, recall rate 99.5% and precision 99.5%. In addition, with less computational effort and in less than a minute, the best results were obtained, showing the efficiency of the image processing technique for the detection and classification of infected and uninfected leaves. It was concluded that this approach was applied to detect fungal diseases on cotton leaves to promote the production and harvest of good quality cotton and valuable cotton products.

Keywords—Fungal diseases; watershed segmentation; SVM; K-means; Edge Detection algorithm

I. INTRODUCTION

Cotton is a soft fibrous substance, usually white, composed of the hairs encompassing the seeds of various tropical plants with free branches (genus *Gossypium*) of the mauve family [1][2]. The ordinary cotton name comes from the Arabic "quon" and generally refers to plants that produce spinnable (plush) fibers on their integument [3].

Cotton cultivation improved the living conditions of rural and urban populations. It contributed to the economic development of the CFA franc zone countries in Africa. Production of cotton has quadrupled in the last two decades, making the region the world's second largest cotton exporter, accounting for 15% of global exports. As a result, generating the most extensive receipts in various regional countries being the main cash crop for government revenues. Cotton cultivation has impacted the poverty rate positively by employing above two million rural households [4].

Due to fungal disease problems in cotton growing, agricultural industries face economic losses and a qualitative and quantitative decline in annual production. Nowadays,

modern cotton industries are trying to ensure the quality and safety of products to provide healthy cotton to consumers.

Africa is an agricultural zone where most of the population is based on cotton farming. Farmers have a wide selection of diversity to pick appropriate cotton crops. However, the growth of these plants for greatest yield and quality products is very technical. It can be enhanced with the support of technology. The management of cotton crops needs careful observation, particularly to treat diseases that will affect production and, consequently, the life after harvest.

A disease is an alteration of one or more physiological processes caused by irritation from some factors or agents resulting (pests) in loss of coordination in plants. Plant disease(s) is a hurtful modification from the traditional functioning of the process related to physiology in a particular plant. Plant diseases generally can be classified into three major parts: Classification based on the localization area, classification based on the occurrence, classification based on the causal agent [5].

Plant disease problems have arisen as a result of a considerable drop in the quality and quantity of agricultural output. Plant disease losses in Georgia (United States) in 2007 were estimated to be over \$539.74 million. Approximately \$185 million of this total was spent on disease control, with the remainder being the result of disease damage [6]. Several pests, including bacteria, aphids, insects, fungi, and others, cause serious plant diseases, but fungi are responsible for the majority of global losses [7].

FUNGI are microorganisms that exist in the planet and are responsible for nearly 75% of all diseases. They are threadlike organisms known as HYPHAE. Anthracnose, Ascochyta blight, Black root rot, Boll rot, Charcoal rot, Leaf spot, Escobilla, Fusarium wilt, Lint contamination Rust, Phymatotrichum root rot, Powdery mildew, Sclerotium stem and root rot or southern blight, Stem canker, Seedling disease complex, and Verticillium wilt are some of the diseases that can affect your plants.

Observations by experts with the naked eye are the most available method for detecting and identifying diseases in practice. However, this requires the constant supervision of professionals, which may be unaffordable on large farms. Furthermore, in certain poor nations, farmers must travel considerable distances to reach specialists, making consultants prohibitively expensive and time-consuming.

The automatic diagnosis of plant diseases is a crucial area of research. It could be useful for large-scale surveillance cultures, allowing for the automatic detection of symptoms of the disease as soon as they emerge on the plant's leaves. As a result, seeking a rapid, automatic, less expensive, and more accurate method will be a distinct advantage for the quality and development of agricultural products and derivatives, particularly when concentrating on a certain plant and disease category. This paper aims to develop a model that can detect Fungal Diseases for cotton crops only based on quality in image (of cotton leaf) processing.

In summary, the paper's contributions can be summarized as follows:

- 1) Fungal diseases Image Acquisition was done using digital camera or any other device and it was loaded and saved using Matrox Imaging Library (MIL) software and semistructured interviews and questionnaires were also administered to gather the requirements of the model.
- 2) A classification model of fungal diseases detection was formulated based on the watershed segmentation technique using K-Means Clustering and Edge Detection algorithm and Support Vector Machine (SVM) was used for classification.
- 3) Simulation of the classification model of fungal diseases detection using Matlab's Image Processing Toolbox 9.4.
- 4) Assessing the performance of the model using accuracy, efficiency, and specificity as parameters.

This paper is organized as follows. Section II illustrates relevant literature on image processing based on plant diseases and specifically fungal diseases on cotton cultivation. Section III discusses the model design approach and methodologies adopted in the course of the research. Section IV discusses an overview of the results of the simulation of the model designed and analysis of the results emanating from the simulation. Section V recaps the research work and conclusion and future work is suggested.

II. RELATED WORKS

There are numerous literature studies on plant disease detection and image processing technologies. According to Anju [8], the watershed transformation was a good tool for image segmentation in mathematical morphology. Marker-controlled segmentation was used for watershed transformation-based segmentation. A powerful tool for image segmentation, the Prewitt Edge Detection Operator was used to demonstrate another image segmentation technique that included image enhancing and noise removal strategies. The new method was evaluated and compared to the current one. However, this work has to be improved, particularly in terms of evaluating the method's performance.

Rani and Mahip [9] published an article on the use of machine learning to detect various plant diseases using various image processing techniques. Today, the technical processing of images becomes an essential technique for diagnosing different plant diseases during cultivation. Any part of the gold harvesting area can be affected by the disease. This paper focused on detecting various diseases of cotton

cultivation and classifying them. There is as much technical classification as k-Nearest Neighbor Classifier, genetic algorithm, k-means classifier, probabilistic Neural Network, , vector support artificial analysis of the machine and main components neural network, fuzzy logic. Choosing a classification was a tedious task because the quality of the result may be different for various input data. This document provided an overview of various technical classifications used to classify plant leaf diseases.

Thikjarathi and Abirami [10] proposed an Application of Image Processing in Diagnosing Guava Leaf Diseases. In this study, scaled leaf images with improved contrasts are subjected to region expanding segmentation, color transformation (YCbCr, CIELAB), and Scale Invariant Feature Transform (SIFT). Support Vector Machines (SVM) and K-nearest Neighbor (k-NN) classifiers were investigated for disease-wise classification accuracy. While both SVM and k-NN perform well, the former has a little accuracy advantage. However, in the future, this methodology for evaluating classifiers for large datasets can be improved.

Zhang et al. [11] used a natural situation to test an artificial image segmentation algorithm for damaged cotton leaves. The authors employed a segmented monotone decreasing edge composite function, energy function guidance information, the Heaviside function, and the t penalty function to create their model $\phi(x)$. The results indicated that a model of a cotton leaf edge profile curve could be constructed from a model of a cotton leaf coated in bare soil, straw mulching, and plastic film mulching, and that even with non-uniformity of light, the optimal edge of the ROI could be achieved. In a complex background, the model can classify cotton leaves with uneven illumination, shadow, and weed backdrop, and it is preferable to realize the perfect extraction of the blade's edge. However, this task is restricted by the length of time it takes to complete the Segmentation (problem of efficiency). Pre-treatment, feature extraction, and picture categorization are among the other image processing processes that are bypassed.

"Detection and Identification of Rice Leaf Diseases Using Multiclass SVM and Particle Swarm Optimization Technique," by [12], for example, "By combining K-means, SVM, and multiple classifications, the author presented a new approach for detecting and identifying rice leaf illnesses." The gray level co-occurrence matrix (GLCM) was utilized to extract features. An SVM classifier was used to categorize the diseases, and Particle Swarm Optimization was used to improve detection accuracy (PSO). According to testing results, the proposed methodology was 97.91 percent accurate in terms of disease identification. Furthermore, the Feed Forward (FFNN) and SVM neuron networks accounted for 77.96 percent, 85.64 percent, and 90.56 percent, respectively, of the closest neighbour (KNN) networks.

Kumari et al. [13] suggested an automatic disease detection system for three fungal diseases in cotton crops: Alternaria leaf spot fungal disease (ALSFD), Rust Folar Fungal Disease (RFFD), and Grey Mildew Cotton Disease (GMCD). For disease segmentation on cotton leaves, the k-means clustering approach was utilized, and the results were

transferred for disease categorization to Artificial Neural Network (ANN) and Support Vector Machine (SVM) classifiers. On a Cotton plant leaf, Khalmar and Khan [14] worked on automatic early leaf spot disease segmentation. Image processing and machine learning are used in this project. The image is captured with a digital camera, and the contaminated region is segmented using a k-means clustering technique after pre-processing. The study only discovered bacterial leaf infections and did not categorize the leaves of cotton plants.

Carderia et al. [15] proposed that Deep Learning Techniques be used to identify Cotton leaf defects. This study proposes a deep learning-based method for cotton leaf screening in order to monitor the health of the leaves. In this work segmentation was not performed on the leaves, features of the leaves were identified and used for classification.

Statement of Problem Existing techniques used in detecting diseases in cotton cultivation suffer several setbacks

such as: a confusion in detection of two different diseases but having the same apparent symptoms, stages of disease detection are often incomplete and lack performance. As such, this leads to improper management of the cotton crops and causes low productivity. Therefore, there is a need for a classification model for diseases detection that will focus on specific class of the diseases (fungal diseases) to improve cotton cultivation; To fill this gap, the proposed study will provide a model that can accurately and automatically provide the right detection for fungal diseases in cotton cultivation. The outcome of this research will ensure the quality of cotton and provide healthy cotton to the consumers.

III. PROPOSED MODEL

This section describes the developed model. This proposed model architecture depicted in Fig. 1 is an improvement of Zhang et al. [11] automation of Segmentation of infected cotton leaves under a natural environment.

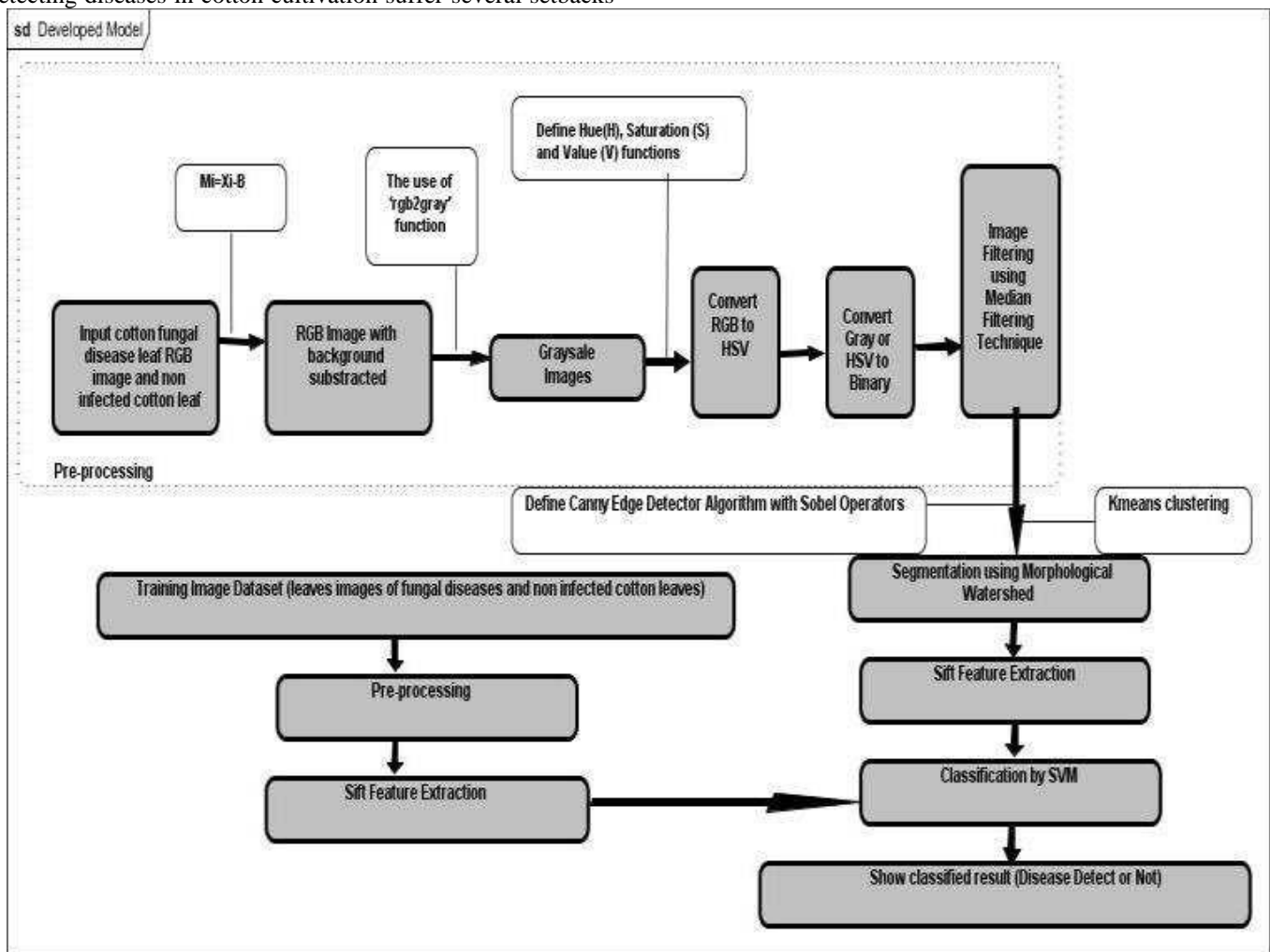


Fig. 1. Developed model.

A. Steps for Pre-processing

The preprocessing steps are Input picture; Subtraction of background; Converting RGB image to grayscale and HSV

image; Converting grayscale image to binary image; Filtering. Fig. 2 shows the different stages of image preprocessing.

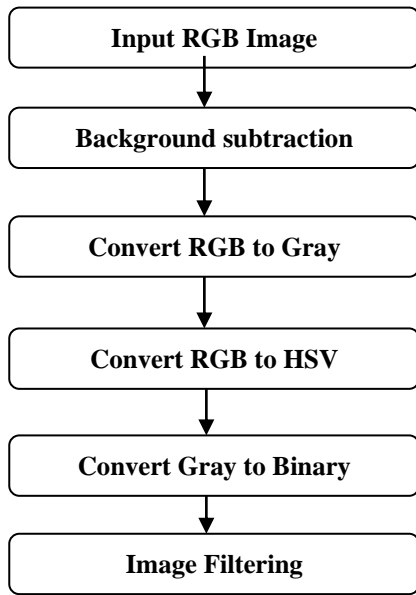


Fig. 2. Architecture flow.

B. Development of a Formula for RGB to Grayscale

The formula of RGB to grayscale is generally expressed with a small variation of the coefficients of R, G, and B according to many papers including:

- i. $I=0.2989R+0.5870G+0.1140B$ (1)
according to Padmarathi and Thangadurai [16]
- ii. $I=0.299R+0.5870G+0.1140B$ (2)
according to Saravanan [17]
- iii. $I= 0.3R+0.59G+0.11B$ (3)
according to Kanan and Cottrell [18]
- iv. $I=0.21R+0.71G+0.07B$ (4)
according to Samuel et al. [19]

From i, ii, iii and iv luminance,

$$I=(r\%R) +(g\%G) +(b\%B) \quad (5)$$

Where

- (a) Red has contributed r%, Green has contributed g% and Blue has contributed b%;
- (b) $r\%+g\%+b\%=100\%$ (6)

C. Binary Image

A Binary Picture is a numeric picture. It is also considered as bi-level or two level because each pixel of this image can require a single bit 0 or 1. Generally, 0 shows the black color, and 1 shows the white color, which has two assigned pixel values. In the numeric treatment of a picture, a binary picture is used like masks or as the result of some frequent operations such as Segmentation, thresholding, and dithering, etc. Fig. 3 illustrates an example of transforming an RGB image to a grayscale image and a binary image. Fig. 4 also shows another example of transforming an RBG image to grayscale, binary and HSV image.

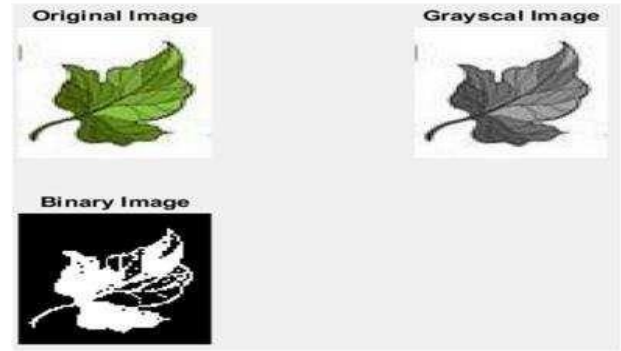


Fig. 3. Illustrates an example of converting an RGB image to a grayscale image and a binary image.

D. Characteristic of Hue, Saturation, and Value(HSV)

$$H = \begin{cases} 0 & \text{if } \max = \min \\ \frac{(60^\circ \times g - b + 360^\circ)}{\max - \min} \text{ mod } 360^\circ & \text{if } \max = r \\ \frac{(60^\circ \times b - r + 120^\circ)}{\max - \min} & \text{if } \max = g \\ \frac{(60^\circ \times r - g + 240^\circ)}{\max - \min} & \text{if } \max = b \end{cases} \quad (7)$$

$$S = \begin{cases} 0 & \text{if } \max = 0 \\ \frac{\max - \min}{\max} = 1 - \frac{\min}{\max} & \text{otherwise} \end{cases} \quad (8)$$

$$V = \text{Max} = \frac{1(r+g+b)}{3} \quad (9)$$

Where r, g, and b stands for red, green and blue normalized in value [0, 1] [20].

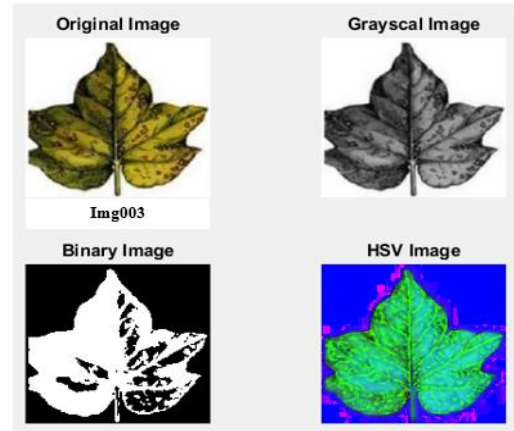


Fig. 4. illustrates an example of converting an RGB image to a grayscale image a binary image, and an HSV image.

E. Image Filtering

In image pre-processing, filtering is an important step to use to reduce image noise and improve the visual quality of an image. Basically, image filtering is useful in various applications such as smoothing, sharpening, noise suppression, and edge detection. There are several types of image filtering techniques, such as Laplacian filtering, low-pass filtering, high-pass filtering, and so on. However, in this study, the median filtering technique was used. The image

filtering technique is used in an image pre-processing step to reduce noise to improve the result of subsequent processing such as an edge detection example of an image. The median filter is called a nonlinear digital filtering process in image processing. The median filtering technique is used in the image to suppress the noise of the image. Fig. 5 shows an example of obtaining a filtered image from an original image.

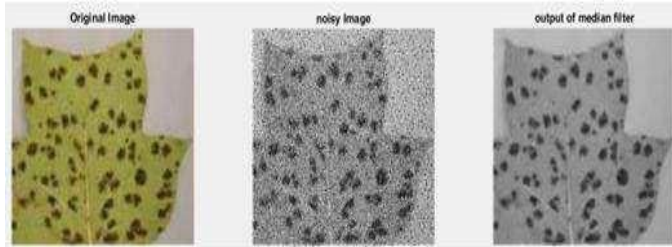


Fig. 5. Obtaining a filtered image from an original image.

K-means Clustering: The K-mean is frequently used to define the natural grouping of pixels in a photograph. It is a simple and quick method that attracts a significant number of consumers. For the purpose of creating vector spaces, a clustering approach is useful. The objects are grouped together at the centroids. With luminosity layer 'L*', chromaticity-layer 'a*' indicating where the color falls along the red-green axis, and chromaticity-layer 'b*' representing where the color falls along the blue-yellow axis, K-Means clustering uses the L*a*b color format. The 'a*' and 'b*' layers contain all color information. The generalized pseudocode of the traditional k-means algorithm and the traditional k-means method according to Oyelade et al. [21] are shown in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1: Generalized pseudocode of the traditional k-means

1. Accept the number of clusters to group data into and the clusters as input values.
 2. Initialize the first k clusters
 - 2.1 Take the first k instances or
 - 2.2 Take Random Sampling of k elements
 3. Calculate the arithmetic mean of each cluster
 4. K-means assigns each record in the dataset to only one of the initial clusters.
 - 4.1 Each record is assigned to the nearest cluster using a measure of Distance (e.g Euclidean distance).
 5. K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.
-

Algorithm 2: Traditional k-means algorithm

INPUT: Dataset of n data points x_1, x_2, \dots, x_n

OUTPUT: $m_j (j = 1, 2, \dots, k)$ = Cluster centroids

MSE (Mean Square Error) = Large Number

Select Initial Cluster centroid m_j ;

While (MSE < OldMSE) do

OldMSE = MSE;

MSE1 = 0;

For (j = 1 to k)

$m_j = 0; n_j = 0;$

endfor

For (j = 1 to k)

Compute Squared Euclidean Distance $d^2(x_i, m_j)$;

endfor

For (j = 1 to k)

Find the closest centroid m_j to x_i ;

$m_j = m_j + x_i; n_j = n_j + 1;$

$MSE1 = MSE1 + d^2(x_i, m_j)$;

endfor

For (j = 1 to k)

$m_j = m_j + x_i;$

$n_j = n_j + 1;$

$MSE1 = MSE1 + d^2(x_i, m_j)$;

$n_j = \max(n_j, 1)$;

$m_j = m_j / n_j;$

endfor

MSE = MSE1;

F. Edge Detection

The identification of edge is employed in image analysis to determine the region's boundaries. In human vision, and probably in many other biological vision systems as well, edges and contours play a significant role. It is not just the sides that catch the eye; a few essential lines can frequently be used to describe or reconstruct an entire figure. Edges are considerable local changes of intensity in a picture. The reasons for intensity include events of geometry (orientation of surface (boundary) discontinuities, discontinuities of depth, discontinuities of color and texture) and non-geometric events (changes of illumination, specularities, inter-reflections of shadows).

There are many techniques for the recognition of edges on a picture. In this study, an appropriate estimate of the first derivative determined by two operators of Sobel will be used. As the derivatives heighten noise; the smoothing aspect is an interesting feature of operators of Sobel. Firstly, the derivatives are executed with the use of gradient amplitude. For a function $f(x, y)$, the gradient f at the coordinate (x, y) is determined as the two-dimensional column vector. Fig. 6 shows the edge detection with the use of Sobel operators in segmentation.

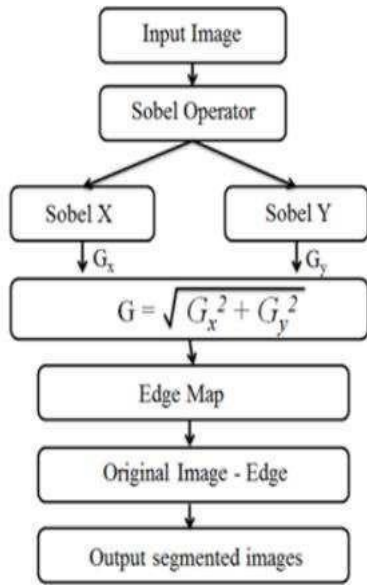


Fig. 6. The edge detection using Sobel Operators in Segmentation.

$$\Delta f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \partial f / \partial x \\ \partial f / \partial y \end{bmatrix} \quad (10)$$

$$\begin{aligned} \Delta f = \text{mag}(\Delta f) &= [G_x^2 + G_y^2]^{1/2} \\ &= \left[\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right]^{1/2} \\ &\cong [G_x + G_y] \quad (11) \end{aligned}$$

Where

G_x is the gradient along the x direction

G_y is the gradient along the y direction

Sandiya and Patial [22].

G. Image Segmentation

The process of segmenting image is breaking down a numeric image to form many segments. Segmentation is a technique for giving meaning to a picture or making an analysis more intelligible. The Segmentation of the picture is used to localize the limits and objects as curves, lines, etc. Literally, in image segmentation operation, all pixels in a picture have been assigned a label where the same label allots the same visual features. The outcomes of image segmentation are multiple segments that a set of contours extracted from the picture or collectively cover the whole image. In a region, each pixel is a computed property or similar with respect to some feature, such as color, texture, or intensity.

H. Morphological Watershed Segmentded Images

The key to using the watershed transform for segmentation is to change your image into one with catchment basins that match to the objects you wish to recognize. Infected and non-infected leaves are shown in Fig. 7 as a Morphological Watershed Segmented Image.

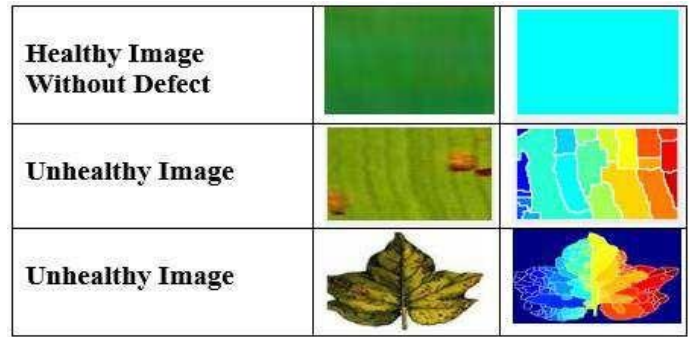


Fig. 7. Infected versus non-infected leaves in a morphological watershed image.

I. Watershed Approach

1) The Algorithm

The possibilities of the algorithm are as follows: assume that a hole is drilled at least in each regional neighborhood and that all geology is submerged from below, allowing the water to climb at a constant rate through the perforations. Flooded pixels are those that are under the water level at any given time. As the water level increases, the size of the flooded areas increases. Water will reach a level where two sites immersed in separate catchment basins will mix in the long term. When this occurs, the algorithm constructs a one-pixel thick barrier between the two sections. This overflow continues till the entire image is divided into numerous different basins separated by watershed edge lines or ridgelines.

J. Feature Extraction: Sift-Scale Invariant Feature Transform

Any object can have various features and important details that can be accessed to produce a description. This layout is used to identify an element in an image that contains a lot of other things. The SIFT method converts a picture into a large set of local feature vectors in order to generate image characteristics. The image's size, translation, or rotation have no effect on any of these feature vectors. The SIFT algorithm uses a four-stage filtering approach to help in the extraction of these properties:

- 1) Detection of Extremes in Scale and Space
- 2) Keypoint Location
- 3) Assignment of Orientation
- 4) Descriptor of a Keypoint

K. Development of Formula for Difference of Gaussians

Calculation $D(x, y, \sigma)$

$$D(x, y, \sigma) \text{ if } L(x, y, \sigma) = I(x, y) * G(x, y, \sigma)$$

$$\text{Then } L(x, y, k\sigma) = I(x, y) * G(x, y, k\sigma)$$

$$\text{Therefore } D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

$$= I(x, y) * G(x, y, k\sigma) - I(x, y) * G(x, y, \sigma)$$

$$= I(x, y) * [G(x, y, k\sigma) - G(x, y, \sigma)]$$

$$\text{If } G(x, y, \sigma) = \frac{1}{2\pi e^2} e^{-\frac{(x^2+y^2)}{2e^2}} \text{ then}$$

$$D(x, y, \sigma) = I(x, y) * [G(x, y, k\sigma) - G(x, y, \sigma)]$$

$$= I(x, y) * \left[\frac{1}{2\pi k e^2} e^{-\frac{(x^2+y^2)}{2k e^2}} - \frac{1}{2\pi e^2} e^{-\frac{(x^2+y^2)}{2e^2}} \right]$$

$$= \frac{1(xy)}{2\pi e^2} * \left[\frac{1}{k^2} e^{-\frac{(x^2+y^2)}{2e^2}} \right]^{1/k^2} - e^{-\frac{(x^2+y^2)}{2e^2}} \right]$$

But with $A = \frac{I(xy)}{2\pi e^2}$ and $B = e^{-\frac{(x^2+y^2)}{2e^2}}$

$$D(x, y, \sigma) = A \times B \left[\frac{(B)^{1/k-B}}{k^2} \right] \quad (12)$$

Where:

The difference of Gaussians calculation is $D(x, y, \sigma)$.

* Denotes the convolution operator.

$I(x, y)$ is the input picture.

Then the scale space is calculated by the function:

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma)$$

with $G(x, y, \sigma) = \frac{1}{2\pi e^2} e^{-\frac{(x^2+y^2)}{2e^2}}$

L. Support Vector Machine (SVM) Classification

SVM is a technique for data analysis and recognition. It is used to do things like regression analysis and categorization. The SVM algorithm takes a set of input data and predicts classes for each individual input. SVM is a linear model-based trainer that uses extracted features to build a hyper plane that translates a piece of data into the separation between the hyperplane and the nearest training points.

IV. RESULTS

a) Contrast

The intensity difference between a pixel as well as its adjacent for the entire image is returned by this function. Range=[0(size(GLCM, 1) - 1)2]. There's no difference in a continuous image.

$$Contrast = \sum_{i,j=0}^{N-1} (i, j)^2 \quad (13)$$

b) Energy

The GLCM Range = [0 1] returns the total of squared elements. A continual image necessitates the use of energy.

$$Energy = \sum_{i,j=0}^{N-1} C(i, j)^2 \quad (14)$$

c) Homogeneity

Returns a value that indicates how near an element's GLCM distribution is to the GLCM diagonal. [0 1] is the range. For a diagonal GLCM, homogeneity is 1

$$Homogeneity = \frac{\sum_{i,j=0}^{N-1} C(i, j)}{(1+(i-j))^2} \quad (15)$$

According to the results of Table I, the normal leaves of cotton have the following features: the contrast of different leaves is between 0 to 0.09; homogeneity is between 0.98 to 1.0; energy is between 0.98 to 1.0. According to the results of Table II, the disease leaves of cotton have the following features: the contrast of different leaves is between 0.1400 to 0.3; homogeneity is between 0.2000 to 0.97; energy is

between 0.3 to 0.97. The above result analysis concludes that diseased leaf image contrast is more than the normal leaf. On the other hand, diseased leaf image homogeneity and energy are less than the normal leaf. Hence, we conclude that images are defective.

TABLE I. NORMAL LEAF IMAGE ANALYSIS

Normal	Leaf1	Leaf2	Leaf3
CONTRAST	0.0898	0.0501	0.0
ENERGY	0.9879	0.9897	1.0
HOMOGENITY	0.9807	0.9759	1.0

TABLE II. DISEASED LEAF IMAGE ANALYSIS

Diseased	Leaf1	Leaf2	Leaf3
CONTRAST	0.2644	0.1628	0.1591
ENERGY	0.4889	0.5737	0.6834
HOMOGENITY	0.9725	0.9617	0.9343

A. Classification Results and Somecomparisons with the Existing Model

As the first manifestations and symptoms of the fungal disease appear on the leaves of the plant, detection techniques and classification are developed. The characteristics of color, texture, and form are widely used in detecting and classifying plant infections. Following segmentation, the texture, pigment, and form properties of the diseased areas are removed or separated and used as input to the SVM classifier.

According to Table III relating to the achievement of the preprocessing and image processing steps, only the image segmentation step is performed with the existing model while the preprocessing, segmentation, extraction, and classification of images are done with the developed model.

Table IV provides information on the number of infected and uninfected leaves, size, and types of images in the SVM Learning database.

TABLE III. ACHIEVEMENT OF IMAGE PREPROCESSING AND PROCESSING STEPS

Leaf Type	Number of Images	Size(pixels)	Image Type
Healthy	10	193x228	JPG
Disease	500	193x228	JPG

TABLE IV. SVM LEARNING DATABASE

S/N	Steps	Existing Model [22]	Developed Model
1	Stages in image preparation	-	+
2	Categorization of image	+	+
3	Feature Extraction	-	+
4	Classification	-	+

Achieved (+).

Not achieved (-).

B. Distribution of Images

A set of 500 images of infected cotton leaves was used to validate the detection approach of the disease. In this sample of 500 images, 300 (60%) were used for the training of the system and 200(40%) for the tests. Table V illustrates the distribution of images.

TABLE V. IMAGE DISTRIBUTION

Disease	Images	Training Sample	Testing Sample
			Diseases Detection
Fungal Disease	500	300	200
Percentage	100%	60%	40%

C. Evaluation of the Model for Fungal Diseases Detection in Cotton Cultivation

The evaluation of the proposed approach was achieved based on precision and recall, and the images are grouped in accordance with the generated results. Based on the system’s results, these images were categorized as TP (rule matched and disease present), FP (rule matched and no disease present), TN(no rule matched and no disease present), and FN (no rule matched and disease present) based on results generated by the system.

- 1) *TP (true positive)*: The infected images are predicted to have a fungal disease.
 - 2) *TN (true negative)*: Images not infected by fungal disease and are predicted to be negative of infection.
 - 3) *FP (false positives)*: No fungal diseases present in the image but predicted to be present by the system.
 - 4) *FN (false negatives)*: Presence of fungal disease in the image but the system predicted its absence.
 - 5) *GT (ground truth)*: It displays the number of comparisons made during the testing process.
 - 6) *TC (total cases)*: The total number of comparisons made during the testing process.
 - 7) *RM (result of method)*: This shows the overall number of false-positive and true positive system predictions.
 - 8) *Accuracy*: The performance of the system as a percentage of correct prediction.
 - 9) *Misclassification*: The percentage of times the system has predicted incorrectly.
 - 10) *Recall*: Represents the percentage of affirmative cases discovered by the system represented.
 - 11) *Precision*: It shows the percentage of the positive predictions made by the system.
- These results are presented in Tables VI and VII.

TABLE VI. THE CONFUSION MATRIX BASED ON FUNGAL DISEASES

Fungal Diseases	TP	TN	FP	FN
200	199	9	1	1

TABLE VII. SUMMARY OF THE RESULTS

Cases	Predicted: Negative	Predicted: Positive
Negative Cases	TN:9	FP:1
Positive Cases	FN:1	TP:199

TABLE VIII. EVALUATION OF THE PROPOSED METHODOLOGY BASED ON CLASSIFICATION RESULTS

Feature	Formula	Resulted Value	Accuracy (%)
GT	Actual Positive Images	200	-
TC	TP+TN+FP+FN	210	-
RM	FP+TP	200	-
Accuracy	(TN+TP)/TC	0.9905	99.05
Misclassification	(FP+FN)/TC	0.0095	0.95
Recall	TP/GT	0.995	99.5
Precision	TP/RM	0.995	99.5
Specificity	TN/(TN+FP)	0.90	90%

The evaluation results in Table VIII showed that the proposed technique gives 99.05% accuracy with a 0.95% misclassification rate. Furthermore, the recall rate is 99.5%, Specificity is 90% and precision is 99.5%. The system's efficiency is justified by reducing processing time and expertise costs. For example, under a 3.0 GHz Pentium IV PC with 1 GB of integrated RAM, the average processing time between preprocessing the image and the classification of the improved method was 49.6 seconds. The time required to obtain a result has been considerably reduced. This allows farmers to save time (a few seconds for a result) and resources (the cost of the expertise is considerably reduced, and it is not necessary to make an appointment with an expert in the hope of defining the quality of cotton).

Table IX represents the performance comparison between different cotton leaf disease identification and classification Systems.

TABLE IX. PERFORMANCE COMPARISON BETWEEN DIFFERENT COTTON LEAF IDENTIFICATION AND CLASSIFICATION MODELS

Reference	Technique	Disease affected Plant	Accuracy
Proposed System	Proposed Method with SVM Classifier	Cotton	99.5%
N. R. Bhimte and V. R. Thool [23]	Image Processing with SVM Classifier	Cotton	98.46%

V. CONCLUSION AND FUTURE WORKS

In this research, cotton plant disease detection of infected leaves and classification of infected and uninfected leaves is done using image-based processing and machine learning approaches to assist farmers during their struggle against disease outbreaks by making the right decision to increase productivity and collect pure cotton. The results proved that the proposed model has the ability to accurately distinguish between infected and uninfected cotton leaves. However, the model is limited to only detecting fungal diseases in cotton leaves. For future research, this work could be extended to consider other categories of cotton plant diseases such as bacterial, seedling, and boll rots and also implement the system as a mobile and web application to eradicate the manual identification of plant defects, which has become a long and costly process.

In general, the use of image processing to detect diseases and classification does not apply to a specific area. For sustainable agriculture, this work can be applied to identify diseases or the quality of vegetables with great accuracy. This work will contribute significantly to the development of agricultural research.

ACKNOWLEDGMENT

This Research was funded by the TEDFUND Research Fund and Africa Center of Excellence OAK-Park, Obafemi Awolowo University, Ile-Ife, Nigeria.

REFERENCES

- [1] Wendel, J. F., and Albert, V. A. (1992). Phylogenetics of the cotton genus (*Gossypium*): character-state weighed parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany*, 115- 143.
- [2] Seelanan, T., Schnabel, A., & Wendel, J. F. (1997). Congruence and consensus in the Cotton tribe (Malvaceae). *Systematic Botany*, 259-290.
- [3] Lee J (1984). Cotton as a World Crop. *Agronomy. A series of monographs - American Society of Agronomy (USA) Journal*. 24,1-25. ISSN : 0065-4663.
- [4] Badiane, O.G., Dheineswar, G.L. (2002)"Cotton Sector Strategies in West and Central Africa(English). Policy Research working paper;no: WPS 2867. Washington, D.C. ; World Bank Group <http://documents.worldbank.org/curated/en/481531468768541415/Cotton-Sector-Strategies-in-West-and-Central-Africa>.
- [5] Kiran R. and Ujwalla G. (2014). An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN:2278-0661, p-ISSN: 2278-8727.16(1), 10-16.
- [6] Al-Hiary H., Bani-Ahmad S., Reyalat M., Braik M. and ALRahamneh Z. (2011). Fast and Accurate Detection and Classification of Plant Diseases. *International Journal of Compute Applications (0975 – 8887)*.17(1), 31-38.
- [7] Asher M. and Hanson L. (2006). *Fungal and Bacterial Diseases*. pp. 286-315 In Sugar Beet. A.P. Draycott, ed. Blackwell Publishing Ltd.
- [8] Anju,B. (2012)" An Improved Watershed Image Segmentation Technique using Matlab". *International Journal of Scientific and Engineering Research*. Volume3, Issue 6. ISSN:2229-5518.
- [9] Rani P, and Mahip B. (2014). Review Paper on Identification of Plant Diseases Using Image Processing Technique. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 3(11),3667-3669. ISSN: 2278 – 1323.
- [10] Thilagavathi, M, and Abirami, S. (2017)" Application of Image Processing in Diagnosing Guava Leaf Diseases". *International Journal of Scientific Research and Management (IJSRM)*.Volume 5, Issue 07, pp 5927-5933.
- [11] Zhang Jian-hua, Kong Fan-tao, Wu Jian-zhai, Han Shu-qing, Zhai Zhi-fen (2018). Automatic image segmentation method for cotton leaves with disease under natural environment. *Journal of Integrative Agriculture*. 17(8): 1800–1814.
- [12] Prabira, K.S., Nalini, K.B. and Amuya, K.B.(2019) " Detection and Identification of Rice leaf Disease using Multiclass SVM and Particle Swarm Optimization Technique".*International Journal of Innovative Technology and Exploring Engineering(IJITEE)*. Volume 8, Issue 6S2 pg 108-120. ISSN:2278-3075.
- [13] Kumari, C.U., Vignash, N.A., Panigraphy, A.K. ,Ramya, L. and Padma, T.(2019) "Fungal Disease in Cotton Leaf Detection and Classification using Neural Networks and Support Vector Machine, Volume 8, Issue 10. ISSN:2278-3075.
- [14] Khairnar, K. and Khan, S. (2020)" Automatic Early Leaf Spot Disease Segmentation on Cotton Plant Leaf ". *International Journal of Recent Technology and Engineering*. Volume 9, Issue 2, ISSN: 2277-3878.
- [15] Carderia, R.F., Santiago, W.E., Teruel, B. (2021) Identification of Cotton Leaf Lesions Using Deep Learning Techniques. *Sensor*, 21,3169, <https://doi.org/10.3390/s21093169>.
- [16] Padmavathi, K. and Thangadurai, K. (2016)" Implementation of RGB and Grayscale Images in Plant leaves Disease Detection a Comparative Study. *Indian Journal of Science and Technology*. Volume 9, no 6.pg 1-6. DOI:10.17485/jst/2016/v9i6/77739.
- [17] Saravanan Chandran (2010) Color Image to Grayscale Image Conversion. *Second International Conference on Computer Engineering and Applications*.(pp196-199).DOI:10.1109/ICCEA.2010.192.
- [18] Kanan C., Cottrell GW (2012)" Color-to-Gray scale: Does the method matter in Image Recognition?. *PloS ONE*. Volume 7, no 1.pp1-7.e21740. DOI:10.1371/Journal.pone.0029740
- [19] Samuel M., Melo G., Judith K. (2015). A comparative study of grayscale conversion techniques applied to SIFT descriptors. *SBC Journal on Interactive Systems*, 6(2).30-36.
- [20] Jun-Dong Chang, Shyr-Shen Yu, Hong-Hao Chen and Chwei-Shyong Tsai (2010). HSV- based Color Texture Image Classification using Wavelet Transform and Motif Patterns. *Journal of Computers*, 20(4),63-69.
- [21] Oyelade O. J, Oladipupo O. O and Obagbuwa I. C (2010). Application of k-Means Clustering algorithm for prediction of Students' Academic Performance. *(IJCSIS) International Journal of Computer Science and Information Security*,7(1),292-295.ISSN1947-5500.
- [22] Sandiya, V. and Patial, R. (2012) Sobel Edge Detection using Parallel Architecture based on FPGA. *International Journal of Applied Information Systems (IAIS)*.3(4),20-24, ISSN:22490868
- [23] N. R. Bhimte and V. R. Thool (2018) "Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 340–344.