Jackfruit Phytophthora Palmivora (Butler) Disease Recognizer Using Naïve Bayes Classifier

Jonah Flor V. Oraño,[1,*] Jomari Joseph A. Barrera,[2] and Elmer A. Maravillas,[3]

Abstract

***Abstract* — This paper presents a technique to detect the presence of *P. palmivora* disease in jackfruit trunk using Naïve Bayes classifier. In this study, 200 sample images of jackfruit trunk were used, which were divided into two sets: for training and for testing. Each set contains 50 images for healthy and 50 images for disease infected. The input images were subjected to image pre-processing such as cropping, scaling, and brightness and contrast adjustment. Then, the images were segmented into two regions using color masking. Texture features such as angular second moment (uniformity) and sum of squares (variance) were also extracted from the images. Next, Naïve Bayes classifier was used to classify whether the jackfruit is infected with the disease or not. Finally, the performance of the classifier was evaluated by computing the overall accuracy of the system. Based on the result, the classifier achieved 94% accuracy in detecting the disease incidence. Moreover, this rate can be further improved by adding texture features and by applying other classification algorithms.**

***Keywords: Phytophthora palmivora*, Naïve Bayes classification, texture features, image processing**

I. Introduction

Jackfruit, which is scientifically known as *Artocarpus heterophyllius Lam* and locally known as "nangka" or "langka," is one of the most widely grown fruit crops in the Philippines, which produces the largest edible fruit that may weigh as much as 50 kg [1]. It is an emerging industry in the country especially in Eastern Visayas, where it is marketed as the "flagship fruit." It is a multipurpose tropical fruit tree which can be the source of food, timber, fodder, dyes, latex, and medicinal and other value-added products [2]. In 2013, jackfruit was planted in a total area of 14,526 hectares (ha) with a total production of 46,080 metric tons (mt) [3]. A follow-up report in [4] shows that production of jackfruit amounted to about 44,605 mt in 2014, 43,666 mt in 2015, and 42,021 mt in 2016. It indicates that the country's production continued to decline.

This decline can be attributed to a variety of reasons such as natural calamities, pests, and boring insects; however, damage caused by pathogens plays a significant role in crop reduction in both quality and quantity. To name one, *Phytophthora palmivora (Butler)*, which is known as one of the most destructive genera of plant pathogens in temperate and tropical regions [5], was identified as the major cause of jackfruit production decline in southern Philippines, which affects 85% of jackfruit orchards in Leyte and Samar [6]. The occurrence of this disease hampered its production and threatened the livelihood of local farmers.

Jackfruit infected with *P. palmivora* shows symptoms including trunk cankers (Fig. 1), chlorosis and wilting of the foliage, root lesions, and tree death. As described by the authors in [7], stem cankers appeared firstly as wet lesions on the bark surfaces, often close to the insertion of large branches, but more frequently at trunk bases. A reddish-brown resin oozed from cracks in the bark. The wood tissues under the lesions showed cream to reddish brown discoloration. The infected areas enlarged, girdling the stems and causing severe decline of the trees.

Fig. 1. Jackfruit infected with *P. palmivora* disease. a. Exterior tree trunk showing canker lesions. b. Exterior surface removed to show reddish color disease

[1] Jonah Flor V. Oraño, Visayas State University, Visca Baybay City, Leyte, Philippines (e-mail: jforano@gmail.com)
[2] Jomari Joseph A. Barrera, Visayas State University, Visca Baybay City, Leyte, Philippines
[3] Elmer A. Maravillas CIT-University, N. Bacalso Ave., Cebu City, Philippines

If timely handling and proper management are not taken into consideration, this disease poses a serious threat to jackfruit yield and long-term viability of plantation. In this context, early and advanced disease detection is of utmost importance. Currently, jackfruit growers and domain experts identify the occurrence of the disease through naked-eye observation and laboratory tests. On the other hand, these could be time consuming and laborious especially in monitoring big jackfruit orchards. With the advent of technology, this process can be automated through image processing and machine learning techniques. These techniques have been applied to various agricultural applications such as to detect the incidence and severity of plant diseases, to determine plant varieties, and to identify the quality grading level of fruits and vegetables. For instance, in the study of [8], features such as color, morphology, and color coherent vector (CCV) were extracted and support vector machine (SVM) classification was used to determine incidence of pomegranate disease, in which the authors obtained an accuracy rate of 82%. Aside from SVM, other classification algorithms were proven to be effective, notably the Naïve Bayes [9] classifier wherein it outperforms the conventional classifiers as indicated in the studies of [10] and [11] in terms of classification accuracy.

There are other numerous studies pertaining to automated crop diseases identification [12] [13]; however, detection of *P. palmivora* disease occurrence in jackfruit does not exist yet. This could be beneficial in monitoring big farms and in the absence of domain experts, hence the conduct of this study.

II. Methodology

Figure 2 illustrates that the system architecture applied in this study consists of two main components: image processing and Naïve Bayes classification. In the following, details about each component were presented.

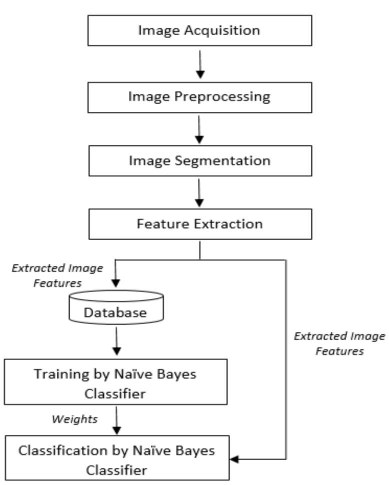

Fig. 2. System architecture of jackfruit P. palmivora disease recognizer.

A. Image Acquisition and Image Pre-Processing

The images of jackfruit trunk (healthy and infected) were captured using a 12-megapixel digital camera last July 24, 2017, and September 5, 2017, in two jackfruit farms located at Mahaplag, Leyte. Cropping was applied to images to remove unwanted details/objects and to emphasize the region of interest. Further pre-processing such as scaling into 600 × 600 and brightness and contrast adjustments were also performed.

B. Image Segmentation and Feature Extraction

Features such as color and texture were extracted from the images. For the color features, images were segmented into two regions using color masking based on the formula shown in Eq. (1). Threshold values that indicate the color characteristics of the infected region are reflected in Figure 3.

$$f(C_i) = \begin{cases} R_1, \ if \ \measuredangle_i\left(hue(C_i)\right) < 10^{\circ} \mid brightness(C_i) < 0.0625 \\ R_2, \ otherwise \end{cases} \tag{1}$$

where

C_i is color

R_1 is infected region

R_2 is not infected region

Hue is an angle between 0° and 360°. Equations (2) and (3) show the computations for hue angle and brightness [14].

$$hue(C_i) = \begin{cases} 0 & \text{if } \max = \min \\ (60^{\circ} \times \frac{g-b}{\max - \min} + 0^{\circ}) \bmod 360^{\circ}, & \text{if } \max = r \\ 60^{\circ} \times \frac{b-r}{\max - \min} + 120^{\circ}, & \text{if } \max = g \\ 60^{\circ} \times \frac{r-g}{\max - \min} + 240^{\circ}, & \text{if } \max = b \end{cases} \tag{2}$$

$$brightness(C_i) = \frac{1}{2}(\max + min) \tag{3}$$

The angle of incidence (∠) is used to compute the distance between hue of red and the hue of C_i.

$$\measuredangle_i(\theta) = \min(360^{\circ} - \theta, \theta) \tag{4}$$

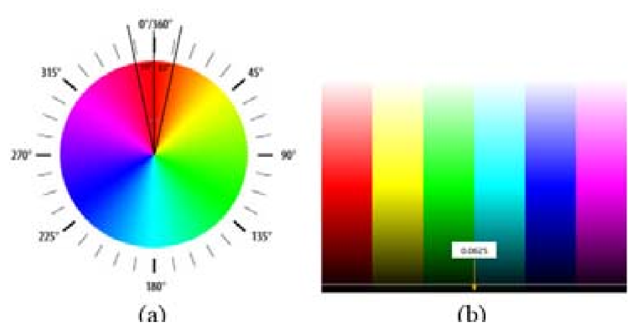

Fig. 3. Color space for (a) hue and (b) brightness.

Furthermore, for the texture features, two of the 14 descriptors defined by Haralick [15] from GLCM were extracted. These include angular second moment (Eq. 5) and sum of squares: variance (Eq. 6). GLCM produces features which describe well the relationship of adjacency among pixels in a texture image.

$$\textit{Angular Second Moment (ASM)} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)^2 \tag{5}$$

$$\textit{Sum of Squares: Variance} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} (i-\mu)^2\ p(i,j) \tag{6}$$

The extracted features values were then categorized based on the defined range of values for a specific category level shown in Table 1 to Table 3. These features from all images in the training set were stored in the database to be used in the classification process.

TABLE 1
CATEGORY OF VALUES FOR THE MASKED REGIONS

Category	Values
Very low	>0 to <=10
Low	>10 to <=20
Moderate	>20 to <= 35
High	>35 to <= 55
Very high	>55

TABLE 2
CATEGORY OF VALUES FOR ASM

Category	Values
Very low	<= 0.00105200
Low	>0.00105200 to <=0.00203048
Moderate	>0.00203048 to <=0.00300897
High	>0.00300897 to <=0.00398745
Very high	>0.00398745

TABLE 3
CATEGORY OF VALUES FOR VARIANCE

Category	Values
Very low	<=58.05314194
Low	>58.05314194 to <=92.45157227
Moderate	>92.45157227 to <=126.85000259
High	>126.85000259 to <=161.24843291
Very high	>161.24843291

C. *Training and Classification Phase*

During the training phase, a data set of 100 images was used, which is composed of 50 healthy trunks and 50 infected trunks.

Class values for Naïve Bayes classifier were defined, which include: not infected and infected. The probabilities for each attribute/feature conditional on the class value were computed. Then, product rule was applied to obtain a joint conditional probability for the attributes while Bayes' rule in Eq. (7) was used to derive the conditional probabilities for each class, wherein the class with the highest probability was considered as the outcome of the prediction [16].

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \tag{7}$$

where

$P(c|x)$—posterior probability
$P(x|c)$ —likelihood
$P(c)$ —class prior probability
$P(x)$—predictor prior probability

III. EXPERIMENTAL RESULTS

The graphical interface for the training phase is shown in Figure 4. It illustrates that once the image is loaded into the system, the extracted feature values and its equivalent category level are displayed. The user th en needs to indicate the expert's classification on the image before clicking the save button. These values will be saved into the database and will be used for the training process.

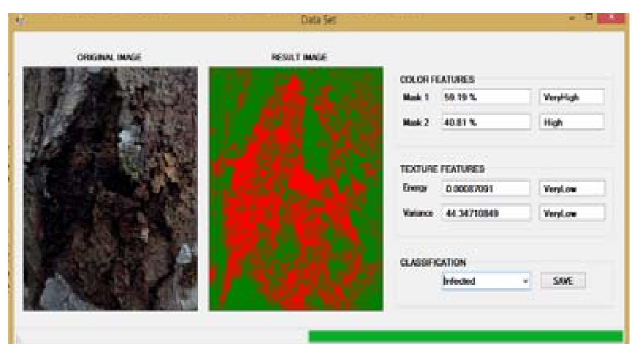

Fig. 4. Features extraction result of the system.

The difference between healthy jackfruit trunks and those with *P. palmivora* disease infection is indicated in Table 4.

TABLE 4
Sample Images of Infected and Not Infected Jackfruit Trunks

INFECTED	NOT INFECTED
	
	
	

Table 5 illustrates the training set for the system after extracting features from all 100 images. It contains columns for extracted numerical values, its equivalent category, and its expected classification values.

Apart from 100 jackfruit trunk images used during the training phase, a different set of 100 images was used for the testing phase. These images were loaded into the system and were then converted into features form. For each image, the predicted class label obtained by Naïve Bayes classifier was saved into the database table alongside with its values (Table 6).

The system can now be used to determine whether the image of the jackfruit trunk shows occurrence of the disease or not. Figure 5 illustrates the decision of the classifier based on the features extracted from the given image.

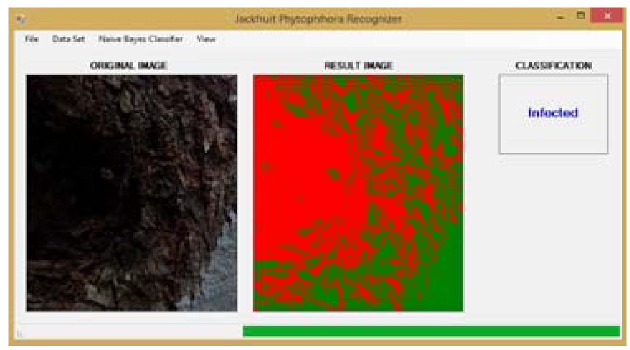

Fig.5. Recognition result of the system.

Table 5
Training Dataset

Sample No.	Mask 1	Mask 2	ASM	Variance	Mask 1 Category	Mask 2 Category	ASM Category	Variance Category	Expected Result
1	34.76	65.24	0.00051	55.11715	Moderate	Very High	Very Low	Very Low	Infected
2	48.62	51.38	0.00252	23.65471	High	High	Moderate	Very Low	Infected
3	36.17	63.83	0.00033	46.04551	High	Very High	Very Low	Very Low	Infected
4	22.99	70.01	0.00108	30.03209	Moderate	Very High	Low	Very Low	Infected
5	59.31	40.69	0.00497	23.97449	Very High	High	Very High	Very Low	Infected
6	18.80	81.20	0.00027	54.99160	Low	Very High	Very Low	Very Low	Infected
7	4.22	95.78	0.00011	140.87128	Very Low	Very High	Very Low	High	Not Infected
8	3.15	96.85	0.00018	124.68015	Very Low	Very High	Very Low	Moderate	Not Infected
9	10.31	89.69	0.00012	134.12725	Low	Very High	Very Low	High	Not Infected
10	7.22	92.78	0.00016	169.53530	Very Low	Very High	Very Low	Very High	Not Infected
11	12.50	87.50	0.0009	178.21163	Low	Very High	Very Low	Very High	Not Infected
:	:	:	:	:	:	:	:	:	:
100	0.16	99.84	0.00025	78.19798	Very Low	Very High	Very Low	Low	Not Infected

TABLE 6
TESTING DATASET

Sample No.	Mask 1	Mask 2	ASM	Variance	Mask 1 Category	Mask 2 Category	ASM Category	Variance Category	Expected Result	Actual Result
1	43.57	56.43	0.00031	36.31061	High	Very High	Very Low	Very Low	Infected	Infected
2	33.09	66.91	0.00008	31.70688	Moderate	Very High	Very Low	Very Low	Infected	Infected
3	30.21	69.79	0.00043	71.98274	Moderate	Very High	Very Low	Low	Infected	Infected
4	19.57	80.43	0.00012	103.2699	Low	Very High	Very Low	Moderate	Infected	Not Infected
5	16.54	83.46	0.00007	105.0948	Low	Very High	Very Low	Moderate	Infected	Not Infected
6	16.04	83.96	0.00021	67.90312	Low	Very High	Very Low	Low	Infected	Infected
7	13.71	86.29	0.00030	84.07898	Low	Very High	Very Low	Low	Infected	Infected
8	12.3	87.70	0.00018	160.4463	Low	Very High	Very Low	High	Not Infected	Not Infected
9	10.15	89.85	0.00023	96.82886	Low	Very High	Very Low	Moderate	Infected	Not Infected
10	9.82	90.18	0.00012	166.3141	Very Low	Very High	Very Low	Very High	Not Infected	Not Infected
11	7.28	92.72	0.00016	118.4576	Very Low	Very High	Very Low	Moderate	Not Infected	Not Infected
:	:	:	:	:	:	:	:	:		:
100	7.02	92.98	0.00015	107.681	Very Low	Very High	Very Low	Moderate	Not Infected	Not Infected

A. Discussions

Table 7 shows the classifier created from the training set where NI means not infected and I means infected. The probabilities are computed as the number of samples of a class having the value (category) for a feature divided by the frequency of a class. In this case, the probability of not infected p(NI) is 0.50 and the probability of infected p(I) is 0.50 as well.

TABLE 7
PROBABILITY OF FEATURE CONDITIONAL ON THE CLASS

Category	Features							
	Mask 1		Mask 2		ASM		Variance	
	NI	I	NI	I	NI	I	NI	I
Very low	.47	.00	.00	.00	.50	.43	.00	.44
Low	.03	.20	.00	.00	.00	.05	.01	.06
Moderate	.00	.22	.00	.00	.00	.01	.11	.00
High	.00	.07	.00	.02	.00	.00	.34	.00
Very high	.00	.01	.50	.48	.00	.01	.04	.00

Table 8 below contains the values for all the extracted features from a sample image to be classified as not infected or infected.

TABLE 8
SAMPLE EXTRACTED VALUES TO BE CLASSIFIED AS INFECTED OR NOT INFECTED

Feature	Value
Mask 1	Low
Mask 2	Very high
ASM	Very low
Variance	Low

For the classification as not infected, the posterior is given by Eq. (8).

$$posterior(NotInfected) = \frac{p(Mask\ 1|NotInfected)\, p(Mask\ 2|NotInfected)\, p(ASM|NotInfected)\, p(Variance|NotInfected)\, p(NotInfected}{predictor} \quad (8)$$

While for the classification as infected, the posterior is given by Eq. (9).

$$posterior(Infected) = \frac{p(Mask\ 1|Infected)\, p(Mask\ 2|Infected)\, p(ASM|Infected)\, p(Variance|Infected)\, p(Infected)}{predictor} \quad (9)$$

where

$$\begin{aligned} predictor = {} & p(Mask\ 1|NotInfected)\, p(Mask\ 2|NotInfected) p(ASM|NotInfected) \\ & p(Variance|NotInfected)\, p(NotInfected) + \\ & p(Mask\ 1|Infected)\, p(Mask\ 2|Infected)\, p(ASM|Infected) \\ & p(Variance|Infected)\, p(Infected) \end{aligned}$$

However, given the sample, the predictor is a constant and thus scales both posteriors equally. In this case, it does not affect the classification and can be ignored.

Therefore, considering the posterior numerator of Not Infected, which was calculated as

$$posterior(NotInfected) \\ = .03 \times .50 \times .50 \times .01 \times .50 = .00004$$

and the posterior numerator of Infected, which was calculated as

$$posterior(Infected) \\ = .20 \times .48 \times .43 \times .06 \times .50 = .00012$$

it can be predicted that the sample is infected.

For each image, the predicted class label obtained by Naïve Bayes classifier was then compared to the actual class label specified by the domain expert. The overall results are displayed in the confusion matrix (Table 9) to emphasize how many images from the total of each class are accurately predicted.

Then detection accuracy for each class was compute as shown in Eq. (10).

$$\frac{\text{sum of correctly predicted class label}}{\text{total number of predictions}} \tag{10}$$

While the overall accuracy of the system was calculated using Eq. (11).

$$\frac{\text{sum of correct classification}}{\text{total number of classification}} \tag{10}$$

TABLE 9
CONFUSION MATRIX

Class	Not Infected	Infected	Detection Accuracy
Not Infected	50	0	100.00%
Infected	6	44	88.00%
Overall Accuracy			94.00%

The result shows that the detection accuracy of infected class is lower (88%) compared to the not infected class (100%). However, it is notable that the system was able to achieve a high overall accuracy rate of 94%.

IV. CONCLUSION AND RECOMMENDATION

The overall accuracy rate of the system, which is 94%, indicates that the application of Naïve Bayes classifier on the extracted color and texture features can significantly support an accurate detection of the *P. palmivora* disease. The recognition accuracy rate can be further improved by using more high-resolution images, applying other image processing techniques, adding more features, and applying other classification algorithms such as artificial neural network, C4.5 classifier, SVM, etc. It would also be better that the system be able to classify the severity of the disease infection for a basis of appropriate disease management strategies.

REFERENCES

[1] "Jakfruit in the Philippines, Part 1," March 1985. [Online]. Available: http://rfcarchives.org.au.

[2] C. R. Elevitch and H. I. Manner, "Artocarpus heterophyllus (jackfruit)," April 2006. [Online]. Available: www.traditionaltree.org.

[3] "Philippines: Jackfruit planted area and production 2008-2013," [Online]. Available: https://www.statista.com/.

[4] "CountrySTAT Philippines," 10 November 2017. [Online]. Available: http://countrystat.psa.gov.ph. [Accessed 18 April 2018].

[5] A. Drenth and D. Guest, "Diversity and Management of Phytophthora," *ACIAR Monograph series no. 114. Australian Centre for International Agricultural Research, Canberra,* 2004.

[6] L. Borines, V. Palermo, G. Guadalquiver, C. Dwyer, A. Drenth, R. Daniel and D. Guest, "Jackfruit decline caused by Phytophthora palmivora (Butler)," *Australasian Plant Pathology,* vol. 42, pp. 123-129, 2013.

[7] M. Van Tri, N. Van Hoa, N. Minh Chau, A. Pane, R. Faedda, A. De Patrizio, L. Schena, C. H. Olsson, S. A. Wright, M. Ramstedt and S. O. Cacciola, "Decline of jackfruit (Artocarpus heterophyllus) incited by Phytophthora," *Phytopathologia Mediterranea,* pp. 9-14, 2015.

[8] M. Bhange and H. Hingoliwala, "Smart Farming: Pomegranate Disease Detection Using Image Processing," *Procedia Computer Science,* pp. 280-288, 2015.

[9] M. M. Roomi and S. Saranya, "Bayesian Classification of Fabrics Using Binary Cooccurrence Matrix," *International Journal of Information Sciences and Techniques (IJIST),* vol. 2, no. 1, pp. 1-9, 2012.

[10] D.-C. Park, "Image Classification Using Naive Bayes Classifier," *International Journal of Computer Science and Electronics Engineering (IJCSEE),* vol. 4, no. 3, pp. 135-139, 2016.

[11] M. Mursalin, M. M. Hossain, M. K. Noman and M. S. Azam, "Performance Analysis among Different Classifier Including Naive Bayes, Support Vector Machine and C4.5 for Automatic Weeds Classification," *Global Journal of Computer Science and Technology Graphics & Vision,* vol. 13, no. 3, 2013.

[12] A. Fahrurozi, S. Madenda, Ernastuti and D. Kerami, "Wood Classification Based on Edge Detections and Texture Features Selection," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 6, no. 5, p. 2016, 2167-2175.

[13] S. Arivazhagan, R. N. Shebiah, S. Ananthi and S. V. Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Agric Eng Int: CIGR Journal,* vol. 15, no. 1, pp. 211-217, 2013.

[14] M. K. Agoston, Computer Graphics and Geometric Modeling: Implementation and Algorithms, Springer, 2005.

[15] R. Haralick, K. Shanmugam and L. Dinstein, "Textural Features for Image Classification," *IEEE Transaction on Systems, Man and Cybernatics,* vol. 3, no. 6, pp. 610-621, 1973.

[16] M. K. Leung, "Naive Bayesian Classifier," 28 November 2007. [Online]. Available: cis.poly.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf. [Accessed 2018].